

**Arab American University  
Faculty of Graduate Studies  
Department of Natural, Engineering  
and Technology Sciences  
Master Program in Data Science and Business Analytics**



**“Utilizing Machine Learning for Measuring Nuchal  
Translucency Thickness in First Trimester Ultrasound to  
Detect Chromosomal Abnormalities”**

**Roa Omar Mohammad Baddad  
202216543**

**Supervision Committee:**

**Dr. Amani Yousef Owda**

**Prof. Nashwa Elbendary**

**Dr. Sami Sader**

**This Thesis Was Submitted in Partial Fulfilment  
of the Requirements for the Master Degree  
in Data Science and Business Analytics**

**Palestine, December/2025**

**© Arab American University. All rights reserved.**

**Arab American University**  
**Faculty of Graduate Studies**  
**Department of Natural, Engineering**  
**and Technology Sciences**  
**Master Program in Data Science and Business Analytics**



## **Thesis Approval**

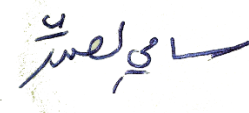
### **Utilizing Machine Learning for Measuring Nuchal Translucency Thickness in First Trimester Ultrasound to Detect Chromosomal Abnormalities**

Roa Omar Mohammad Baddad

202216543

This thesis was defended successfully on 21<sup>th</sup> December, 2025 and approved by:

Thesis Committee Members:

| Name                      | Title                           | Signature                                                                             |
|---------------------------|---------------------------------|---------------------------------------------------------------------------------------|
| 1. Dr. Amani Yousef Owda  | Main Supervisor                 |  |
| 2. Prof. Nashwa Elbendary | Member of Supervision Committee |  |
| 3. Dr. Sami Sader         | Member of Supervision Committee |  |

Palestine, December/2025

## **Declaration**

I declare that, except where explicit reference is made to the contribution of others, this thesis is substantially my own work and has not been submitted for any other degree at the Arab American University or any other institution.

Student Name: Roa Omar Mohammad Baddad

Student ID: 202216543

Signature: Roa Baddad

Date of Submitting the Final Version of the Thesis: January 7<sup>th</sup>, 2026

## **Dedication**

All praise is due to God, by whose grace all good things are perfected, after a long journey of striving and perseverance.

I dedicate this message to everyone who contributed to this achievement, whether directly or indirectly.

To my dearest mother, the lamp of my life, whose love, prayers, and boundless support illuminated every step of this path. Words fall short of expressing my gratitude for your presence in my life.

To my beloved father, my first pillar of strength, the source of my inspiration, a symbol of generosity and sacrifice.

To my dear brothers and sisters, my lifelong support, whose encouragement and love helped bring this achievement to life.

To the one whose presence was energy, whose support was unwavering, and whose faith in me pushed me to continue the journey no matter how long it grew.

To my sister's children, my maternal aunts, my maternal uncles, and every member of my family—you were the strength behind me, the encouragement at every turn, and the smile that nurtured hope throughout my Master's journey and beyond.

To my friends and colleagues, who shared the challenges and the triumphs, and who were a source of support at every stage.

To my esteemed professors, who cultivated in me a love for research and knowledge.

And to beloved Palestine, my homeland that carries my dreams and hopes. To your martyrs, your prisoners, and your wounded, and especially to Gaza Al-Hashim—the emblem of patience and pride—I dedicate this work with all love and loyalty. You remain the greatest dream that inspires me to strive for a brighter tomorrow.

I dedicate this humble work to all of you; it is the fruit of your support, your love, and your sincere prayers.

This achievement is for you and because of you.

Author: Roa Omar Mohammad Baddad

## Acknowledgments

I extend my deepest gratitude and sincere appreciation to everyone who contributed to the completion of this thesis, which would not have been possible without the grace of Allah and the support and encouragement of many individuals and institutions. I also express my profound thanks and appreciation to my esteemed university, the Arab American University.

I am equally grateful to the Faculty of Graduate Studies for their continuous support and dedication to providing an academic environment that fosters excellence and scientific research. My heartfelt thanks also go to the Department of Natural, Engineering, and Technological Sciences for their ongoing support and for creating a stimulating research environment equipped with advanced scientific resources that enabled the successful completion of this work.

I would like to express my sincere gratitude to the committee supervising and evaluating my thesis, represented by my supervisor, Dr. Amani Owda, for her insightful vision, continuous support, and valuable academic guidance. Her dedication and precise supervision played a significant role in shaping my research perspective, refining my academic abilities, and inspiring me to reach this milestone.

My thanks and appreciation also extend to the distinguished committee members, Prof. Nashwa Elbendary and Dr. Sami Sader, for their constructive feedback and valuable insights, which enriched this work with depth and quality.

I would also like to extend my sincere gratitude to Dr. Majdi Owda for his continuous support throughout the entire duration of this thesis.

I extend my sincere thanks to Mr. Hồ Văn Thép, whose provision of the dataset played a crucial role in enabling the advancement of this research.

To everyone who contributed in any way to this work, I offer my sincere thanks and appreciation.

# **Utilizing Machine Learning for Measuring Nuchal Translucency Thickness in First Trimester Ultrasound to Detect Chromosomal Abnormalities**

**Student's Name: Roa Omar Mohammad Baddad**

**Supervision Committee:**

**Dr. Amani Yousef Owda**

**Prof. Nashwa Elbendary**

**Dr. Sami Sader**

## **Abstract**

This study addresses the critical challenge of variability in manual Nuchal Translucency (NT) measurement by developing and evaluating a Deep Learning (DL) pipeline for automated, accurate assessment in first-trimester ultrasound images to detect chromosomal abnormalities. The research was conducted utilizing two ethically approved datasets comprising 2,425 images from pregnancies at 11-13+6 weeks gestation, sourced from Shenzhen People's Hospital, China, and the National Hospital of Obstetrics and Gynecology, Vietnam.

The study's methodology employed a two-stage DL framework. The first stage used a DenseNet121 architecture for automated image quality assessment, which successfully classified images as "Standard" or "Non-Standard" with 94% accuracy, ensuring only diagnostically reliable images were used for analysis. The second stage involved a novel DenseNet-based segmentation model to precisely delineate the NT region for measurement. This model demonstrated superior performance, achieving a Dice coefficient of 0.897 and an overall accuracy of 0.989, significantly outperforming the widely used U-Net architecture.

The results confirm the pipeline's clinical efficacy. Automated NT measurements showed high agreement with expert annotations, with 73.05% of measurements deviating by less than 1mm. The integrated system achieved over 90% sensitivity and specificity for identifying high-risk cases (NT > 3mm), validating its potential as a robust decision-support tool. The study's main recommendations are to integrate DL-driven quality assurance into clinical NT screening to reduce inter-operator variability, to invest in the curation of large, diverse datasets to improve model generalizability, and to foster interdisciplinary collaboration to facilitate the seamless adoption of this technology in prenatal care.

**Keywords:** Nuchal Translucency (NT), Deep Learning (DL), Prenatal Screening, Ultrasound Imaging, Chromosomal Abnormalities

## **Table of Contents**

|                                                 |       |
|-------------------------------------------------|-------|
| Declaration .....                               | I     |
| Dedication .....                                | II    |
| Acknowledgments .....                           | III   |
| Abstract .....                                  | IV    |
| List of Tables .....                            | XI    |
| List of Figures .....                           | XIII  |
| List of Appendices .....                        | XVIII |
| List of Definitions of Abbreviations .....      | XIX   |
| Chapter One. Introduction .....                 | 1     |
| 1.1 Overview .....                              | 1     |
| 1.2 The Significance of the Study .....         | 2     |
| 1.3 Research Problem .....                      | 2     |
| 1.4 Research Objective .....                    | 3     |
| 1.5 Research Questions .....                    | 3     |
| 1.6 Study Hypotheses .....                      | 4     |
| 1.6.3 Diagnostic Performance Hypothesis .....   | 4     |
| 1.6.4 Robustness Hypothesis .....               | 4     |
| 1.6.5 Integration Feasibility Hypothesis .....  | 4     |
| 1.7 Limitations of the Study .....              | 5     |
| 1.7.1 Conceptual Limitations .....              | 5     |
| 1.7.2 Substantive Limitations .....             | 5     |
| 1.7.3 Temporal Limitations .....                | 5     |
| 1.7.4 Spatial Limitations .....                 | 5     |
| 1.7.5 Population and Sample Limitations .....   | 5     |
| 1.7.6 Methodological Limitations .....          | 5     |
| 1.8 Conceptual and Procedural Definitions ..... | 6     |
| 1.8.1 Nuchal Translucency .....                 | 6     |
| 1.8.2 Deep Learning .....                       | 6     |
| 1.8.3 Chromosomal Abnormalities .....           | 6     |
| 1.8.4 Image Quality Assessment .....            | 6     |

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| 1.8.5 NT Measurement Accuracy .....                                         | 6  |
| 1.9 Thesis Structure .....                                                  | 7  |
| 1.10 Conclusion .....                                                       | 7  |
| Chapter Two. Literature Review .....                                        | 8  |
| 2.1 Introduction .....                                                      | 8  |
| 2.2 Stages of Fetal Development .....                                       | 8  |
| 2.3 Screening .....                                                         | 11 |
| 2.3.1 Prenatal Screening .....                                              | 12 |
| 2.3.2 First Trimester Screening .....                                       | 13 |
| 2.3.3 Triple Screening .....                                                | 13 |
| 2.4 First-Trimester Anatomy Assessment .....                                | 14 |
| 2.5 Nuchal Translucency .....                                               | 16 |
| 2.6 NT Measurement Success Rates Across Gestational Ages .....              | 17 |
| 2.7 The Transient Nature of NT Between 10–14 Weeks .....                    | 18 |
| 2.7.1 Reason for Selecting Up to 14 Weeks for NT .....                      | 18 |
| 2.7.2 Reason for Selecting 10 Weeks as the Lower Limit .....                | 19 |
| 2.8 Pathological Correlation Between Increased NT and Euploid Fetuses ..... | 19 |
| 2.9 Etiology of Increased NT .....                                          | 20 |
| 2.9.1 Hemodynamic Disturbances and Cardiac Dysfunction .....                | 20 |
| 2.9.2 Alterations in the Extracellular Matrix .....                         | 20 |
| 2.9.3 Perturbed Lymphangiogenesis and Endothelial Dysfunction .....         | 21 |
| 2.9.4 Additional Contributing Factors .....                                 | 21 |
| 2.10 Factors Assessed at the First-Trimester Nuchal Translucency Scan ..... | 22 |
| 2.10.1 Crown–Rump Length .....                                              | 22 |
| 2.10.2 Gestational Age .....                                                | 23 |
| 2.10.3 Maternal Age .....                                                   | 24 |
| 2.10.4 Co-variables .....                                                   | 25 |
| 2.10.5 Gender .....                                                         | 26 |
| 2.10.6 Maternal Smoking .....                                               | 27 |
| 2.10.7 PAPP-A, Free $\beta$ -hCG, and Birth Weight .....                    | 28 |
| 2.10.8 Ductus Venosus .....                                                 | 28 |
| 2.11 Abnormalities Associated with Increased NT .....                       | 29 |
| 2.11.1 Cardiac Defects .....                                                | 29 |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| 2.11.2 Sporadic Abnormalities .....                                    | 29 |
| 2.11.3 Trisomy 21 (Down Syndrome) .....                                | 30 |
| 2.11.4 Trisomy 18 (Edwards Syndrome) .....                             | 30 |
| 2.11.5 Trisomy 13 (Patau Syndrome) .....                               | 31 |
| 2.11.6 Turner Syndrome .....                                           | 32 |
| 2.12 Risks Associated with Increased NT .....                          | 32 |
| 2.13 Historical View .....                                             | 33 |
| 2.13.1 Pioneering Research on NT Measurement and Aneuploidy            |    |
| Detection.....                                                         | 34 |
| 2.14 Methods of Measuring Nuchal Translucency .....                    | 36 |
| 2.15 Variability in NT Cutoff Values .....                             | 39 |
| 2.16 Case Studies: Increased Fetal Nuchal Translucency .....           | 40 |
| 2.16.1 Case Study 1: Increased NT and Esophageal Atresia.....          | 40 |
| 2.16.2 Case Study 2: Increased NT with Normal Outcome.....             | 41 |
| 2.16.3 Case Study 3: NT Screening Outcomes in a Prospective Study..... | 42 |
| 2.17 Trends in Research Publications on Fetal Health Monitoring .....  | 43 |
| 2.18 Manual Methods for Measuring NT .....                             | 43 |
| 2.18.1 Limitations of Traditional Methods .....                        | 44 |
| 2.19 Semi-Automated Methods for Measuring NT .....                     | 44 |
| 2.20 Machine Learning Methods for Measuring NT .....                   | 50 |
| 2.21 Deep Learning Methods for Measuring NT .....                      | 59 |
| 2.22 Summary of NT Measurement Research .....                          | 73 |
| 2.23 Fundamental Challenges in Implementing ML/DL Systems .....        | 77 |
| 2.23.1 Current Problems and Limitations in NT Models .....             | 79 |
| 2.24 Research Gap .....                                                | 80 |
| 2.25 Conclusion .....                                                  | 80 |
| Chapter Three. Methodology .....                                       | 81 |
| 3.1 Introduction .....                                                 | 81 |
| 3.2 Data Collection .....                                              | 82 |
| 3.2.1 First Dataset (Fetus Framework) .....                            | 82 |
| 3.2.1.1 Size .....                                                     | 83 |
| 3.2.1.2 Data Type, Format, and Views .....                             | 83 |
| 3.2.1.3 Characteristics .....                                          | 84 |

|                                                            |     |
|------------------------------------------------------------|-----|
| 3.2.2 Second Dataset (NHOG) .....                          | 85  |
| 3.2.2.1 Data Collection & Ethics .....                     | 86  |
| 3.2.2.2 Size .....                                         | 87  |
| 3.2.2.3 Data Type, Format, and Views .....                 | 87  |
| 3.2.2.4 Characteristics .....                              | 88  |
| 3.3 Data Preprocessing and Augmentation .....              | 90  |
| 3.3.1 Fetus Framework Pre-processing .....                 | 90  |
| 3.3.2 NHOG Pre-processing .....                            | 93  |
| 3.4 Noise in Ultrasound Images .....                       | 97  |
| 3.4.1 Influence of Noise on Deep Learning .....            | 97  |
| 3.4.2 Noise Type Detection Using CNNs .....                | 97  |
| 3.4.3 Common Types of Noise in Ultrasound Images .....     | 98  |
| 3.4.4 Denoising Method Comparison and Implementation ..... | 99  |
| 3.4.5 Effect of BM3D Denoising .....                       | 100 |
| 3.5 Exploratory Data Analysis .....                        | 101 |
| 3.5.1 Original Dataset Analysis .....                      | 101 |
| 3.5.2 Metadata Features .....                              | 102 |
| 3.5.3 Statistical Analysis of the Datasets .....           | 103 |
| 3.5.3.1 First Dataset: Fetus Framework .....               | 103 |
| 3.5.3.2 Second Dataset: NHOG .....                         | 104 |
| 3.5.4 Data Visualization .....                             | 106 |
| 3.5.5 Qualitative Data Source and Description .....        | 113 |
| 3.5.5.1 Insights from NT Analysis .....                    | 114 |
| 3.5.5.2 Word Frequency Visualization .....                 | 117 |
| 3.6 Proposed Models .....                                  | 118 |
| 3.7 Data Split .....                                       | 119 |
| 3.8 Selected Models .....                                  | 119 |
| 3.8.1 DenseNet121 .....                                    | 120 |
| 3.8.1.1 Architectural Design .....                         | 120 |
| 3.8.2 VGG16 .....                                          | 121 |
| 3.8.2.1 Architectural Design .....                         | 121 |
| 3.8.3 VGG19 .....                                          | 122 |
| 3.8.3.1 Architectural Design .....                         | 122 |

|                                                                        |     |
|------------------------------------------------------------------------|-----|
| 3.8.4 ResNet-18 .....                                                  | 123 |
| 3.8.4.1 Architectural Design.....                                      | 124 |
| 3.8.5 ResNet-50 .....                                                  | 125 |
| 3.8.5.1 Architectural Design .....                                     | 125 |
| 3.8.6 EfficientNetB7 .....                                             | 126 |
| 3.8.6.1 Architectural Design .....                                     | 126 |
| 3.8.7 InceptionV3 .....                                                | 127 |
| 3.8.7.1 Architectural Design .....                                     | 127 |
| 3.8.8 U-Net .....                                                      | 128 |
| 3.8.8.1 Architectural Design .....                                     | 129 |
| 3.9 Segmentation .....                                                 | 134 |
| 3.9.1 ANN-Based Segmentation .....                                     | 134 |
| 3.9.2 Types of Image Segmentation .....                                | 134 |
| 3.10 Model Parameters .....                                            | 134 |
| 3.11 Hyperparameters .....                                             | 135 |
| 3.12 Regularization .....                                              | 135 |
| 3.13 Activation Functions .....                                        | 138 |
| 3.13.1 Purpose and Importance .....                                    | 139 |
| 3.13.2 Types of Activation Functions .....                             | 139 |
| 3.14 Optimizers .....                                                  | 141 |
| 3.14.1 How Optimizers Work .....                                       | 141 |
| 3.14.2 Types of Optimizers .....                                       | 142 |
| 3.15 Epoch .....                                                       | 142 |
| 3.16 Batch Size .....                                                  | 143 |
| 3.17 Iteration .....                                                   | 143 |
| 3.18 Dropout Rate .....                                                | 144 |
| 3.19 Loss Function .....                                               | 145 |
| 3.20 Learning Rate .....                                               | 146 |
| 3.21 Evaluation and Performance Measures .....                         | 147 |
| 3.22 Proposed Approach.....                                            | 150 |
| 3.22.1 Approach I: Standard vs Non-Standard Image Classification ..... | 150 |
| 3.22.2 Approach II: NT Detection, Measurement, and Classification..... | 152 |
| 3.23 Conclusion .....                                                  | 156 |

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| Chapter Four. Results .....                                       | 157 |
| 4.1 Introduction .....                                            | 157 |
| 4.2 Results of the First Dataset Approach (Fetus Framework) ..... | 158 |
| 4.2.1 DenseNet121 .....                                           | 158 |
| 4.2.2 VGG16 .....                                                 | 162 |
| 4.2.3 VGG19 .....                                                 | 165 |
| 4.2.4 ResNet-18 .....                                             | 168 |
| 4.2.5 ResNet-50 .....                                             | 171 |
| 4.2.6 EfficientNetB7 .....                                        | 173 |
| 4.2.7 InceptionV3 .....                                           | 176 |
| 4.3 Results of the Second Dataset Approach (NHOG) .....           | 179 |
| 4.3.1 Segmentation .....                                          | 180 |
| 4.3.1.1 DenseNet Segmentation Results .....                       | 183 |
| 4.3.1.2 U-Net Segmentation Results .....                          | 184 |
| 4.3.2 NT Measurement .....                                        | 186 |
| 4.3.3 Classification .....                                        | 188 |
| 4.3.4 Measurement Reliability .....                               | 188 |
| Chapter Five. Discussion and Recommendations .....                | 190 |
| 5.1 Introduction .....                                            | 190 |
| 5.2 Discussion of the First Dataset Approach .....                | 190 |
| 5.2.1 Interpretation of Results .....                             | 191 |
| 5.2.2 Comparison with Previous Studies .....                      | 191 |
| 5.3 Discussion of the Second Dataset Approach .....               | 192 |
| 5.3.1 Interpretation of Results .....                             | 192 |
| 5.3.2 Comparison with Previous Studies .....                      | 194 |
| 5.4 Recommendations .....                                         | 195 |
| 5.5 Future Work .....                                             | 197 |
| 5.6 Conclusion .....                                              | 198 |
| References .....                                                  | 199 |
| Appendices #1: Research Facilitation Letter .....                 | 217 |
| الملخص باللغة العربية.....                                        | 218 |

## List of Table

| Table # | Title of Table                                                                                                   | Page    |
|---------|------------------------------------------------------------------------------------------------------------------|---------|
| 2.1     | Summary of Fetal Development                                                                                     | 10      |
| 2.2     | Confusion Matrix for Screening Test Outcomes                                                                     | 11      |
| 2.3     | Performance of Different Methods of Screening for Trisomy 21                                                     | 14      |
| 2.4     | Pregnancy Outcomes in Male and Female Fetuses with Enlarged NT                                                   | 27      |
| 2.5     | NT Thickness and the Frequency of Chromosomal Problem                                                            | 33      |
| 2.6     | The Statistic Result of the NT thickness                                                                         | 45      |
| 2.7     | Summarize the References, Technologies, Task, and Performance Metrics for Different Measurement Methods          | 46      |
| 2.8     | Supervised Learning Technique                                                                                    | 52      |
| 2.9     | Performance Comparison of Different Methods in NT Detection                                                      | 73      |
| 2.10    | Summary of NT Measurement Research                                                                               | 74 -77  |
| 3.1     | Class Distribution of the Fetus Framework Dataset                                                                | 85      |
| 3.2     | NHOG Dataset: Labeled Availability                                                                               | 89      |
| 3.3     | Summarize the preprocessing Steps for Frist Dataset                                                              | 92      |
| 3.4     | Post-Augmentation Sample Counts for the NHOG Dataset                                                             | 96      |
| 3.5     | Summarize the preprocessing Steps for Second Dataset                                                             | 96      |
| 3.6     | Denoising Performance: CNN Autoencoder Vs. Traditional Filters                                                   | 99      |
| 3.7     | Metadata Feature for Frist Dataset-Fetus Framework                                                               | 102     |
| 3.8     | Metadata Feature for Second Dataset-NHOG Vietnam Dataset                                                         | 103     |
| 3.9     | Distribution Fetus Framework for Dataset                                                                         | 104     |
| 3.10    | Descriptive Statistics of NT Thickness, Crown-Rump Length (CRL), and Gestational Age in the NHOG Vietnam Dataset | 104-105 |
| 3.11    | Descriptive Statistics of Nuchal Translucency (NT) Thickness in the NHOG Vietnam Dataset                         | 105     |
| 4.1     | Summarize the Models and Their Key Training Setting for Both Dataset                                             | 157     |
| 4.2     | Test Performance of DenseNet121 (Classification) and ROC Summary                                                 | 160     |
| 4.3     | Performance Metrics Across Models                                                                                | 179     |
| 4.4     | Performance Metrics for DenseNet121Segenation Model                                                              | 184     |
| 4.5     | Performance Metrics for U-Net Segmentation Model                                                                 | 185     |

---

|     |                                                                                                                                                                                                             |     |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.6 | Statistics of Measured NT value Differences between Automated Predictions and Ground- Truth Annotations                                                                                                     | 189 |
| 4.7 | Comparison between Ground- Truth Labels and Models Predictions for NT-Based Risk Classification                                                                                                             | 189 |
| 5.1 | Presents A Comparison between the Performance of This Study and Several Notable Works That have Applied Machine Learning and Deep Learning Methods for The Assessment Image Quality for NT Ultrasound Image | 191 |
| 5.2 | Presents A Comparison between the Performance of This Study and Several Notable Works That have Applied Machine Learning and Deep Learning Methods for The Analysis of Nuchal Translucency NT               | 194 |

---

## List of Figures

| Figure # | Title of Figures                                                                                                                                                                         | Page |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1.1      | A prenatal Ultrasound Scan Showing Nuchal Translucency                                                                                                                                   | 1    |
| 1.2      | Thesis Structure                                                                                                                                                                         | 7    |
| 2.1      | Fetal Development                                                                                                                                                                        | 9    |
| 2.2      | Stages of Fetal Development                                                                                                                                                              | 10   |
| 2.3      | Overview of Prenatal Testing Options for Members of Families with OI Risk. NIPT Non- Invasive Prenatal Testing, Ultrasound, CVS – Chorionic Villus Sampling, Cordocentesis Amniocentesis | 13   |
| 2.4      | Triple Screening                                                                                                                                                                         | 14   |
| 2.5      | Normal First-Trimester Anatomic Feature of The Fetal Rain, Face and Neck as Depicted by Two-Dimensional Ultrasound from The Midsagittal Section of The Head                              | 15   |
| 2.6      | Fetal NT Thickness                                                                                                                                                                       | 17   |
| 2.7      | Week of Pregnancy                                                                                                                                                                        | 18   |
| 2.8      | NT Measurement Percentiles by Crown-Rump Length (CRL) in First-Trimester Screening                                                                                                       | 23   |
| 2.9      | Crown-Rump Length by Gestational Age in Week During the Frist Trimester                                                                                                                  | 24   |
| 2.10     | Maternal Age Related for Chromosomal Abnormalities                                                                                                                                       | 24   |
| 2.11     | Estimated Risk for Trisomy’s 21, 18 and 13 (1/Number Given in The Table) in Relation to Maternal Age and Gestation                                                                       | 25   |
| 2.12     | Maternal Age-Related Risk Trisomy 21 at 12 Weeks of Gestation (a Priori) and The Effect of Fetal Nuchal Translucency Thickness (NT)                                                      | 25   |
| 2.13     | Child with Down Syndrome                                                                                                                                                                 | 30   |
| 2.14     | Child with Trisomy 18                                                                                                                                                                    | 31   |
| 2.15     | Child with Trisomy 13                                                                                                                                                                    | 31   |
| 2.16     | Patient with Tuner Syndrome                                                                                                                                                              | 32   |
| 2.17     | History for First -Trimester Screening                                                                                                                                                   | 34   |
| 2.18     | Nuchal Translucency Measurement that Meets Fetal Medicinal Foundation Criteria (t=thalamus, mo=medulla oblongata, p=pons). The Correct Caliper Placement is “on-to-on”                   | 36   |
| 2.19     | Magnification is too Small                                                                                                                                                               | 37   |
| 2.20     | Neck is Hyperextended                                                                                                                                                                    | 37   |
| 2.21     | Neck is too Much Flexed, Not Suitable for Measuring NT                                                                                                                                   | 37   |

|      |                                                                                                                                                     |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.22 | Correct Placement of Calipers                                                                                                                       | 37 |
| 2.23 | Maximum Thickness                                                                                                                                   | 38 |
| 2.24 | Umbilical Cord Surrounds the Fetal Neck That Disrupts NT Measurement                                                                                | 38 |
| 2.25 | Detailed Ultrasound Image                                                                                                                           | 42 |
| 2.26 | Number of Publications about Learning Methods Used in Fetal Health Monitoring, Within Last Decade by Year. Numbers are Obtained from ScienceDirect. | 43 |
| 2.27 | Nuchal Translucency Detection by Ultrasound Scan                                                                                                    | 44 |
| 2.28 | Algorithmic Workflow for NT Segmentation Utilizing NT Region Detection Results                                                                      | 53 |
| 2.29 | Proposed Method Maximum NT Thickness Measurement                                                                                                    | 56 |
| 2.30 | Proposed Method Minimum NT Thickness Measurement                                                                                                    | 56 |
| 2.31 | Comparison Maximum NT Thickness                                                                                                                     | 57 |
| 2.32 | Comparison Minimum NT Thickness                                                                                                                     | 58 |
| 2.33 | Comparative Analysis of Classification Model                                                                                                        | 59 |
| 2.34 | Comparison of NT Thickness Obtained with Manual & Proposed Method                                                                                   | 65 |
| 2.35 | Thickness Measurements                                                                                                                              | 67 |
| 2.36 | Comparison of Traditional DL Methods                                                                                                                | 68 |
| 2.37 | Traditional ML Model                                                                                                                                | 69 |
| 2.38 | Deep CNN Architectures                                                                                                                              | 70 |
| 2.39 | Comparison with The Another Model                                                                                                                   | 72 |
| 3.1  | Overview of Proposed Pipeline                                                                                                                       | 82 |
| 3.2  | Standard NT Image                                                                                                                                   | 84 |
| 3.3  | Non-standard NT Image                                                                                                                               | 84 |
| 3.4  | Sample Image from the NHOG Fetal Ultrasound Dataset                                                                                                 | 88 |
| 3.5  | Original Images and Corresponding ROIs Extraction by The Proposed Method                                                                            | 88 |
| 3.6  | Overview of the Dataset Split                                                                                                                       | 90 |
| 3.7  | Effect of BM3D Denoising on Ultrasound Image Quality                                                                                                | 91 |
| 3.8  | Original Image and Resizing Operation                                                                                                               | 92 |
| 3.9  | Original Image and Normalization Operation                                                                                                          | 92 |
| 3.10 | Ultrasound Image After Augmentation                                                                                                                 | 93 |
| 3.11 | Overview of The Dataset Split                                                                                                                       | 94 |
| 3.12 | Workflow of Annotation Processing and Dataset Synchronization                                                                                       | 94 |

|      |                                                                                                  |     |
|------|--------------------------------------------------------------------------------------------------|-----|
| 3.13 | Ultrasound Image After Augmentation                                                              | 96  |
| 3.14 | Types of Noise in Ultrasound Images                                                              | 99  |
| 3.15 | Effect of BM3D Denoising with Different $\sigma$ Values on Ultrasound Images                     | 100 |
| 3.16 | Visual Guide to The EDA Process                                                                  | 101 |
| 3.17 | Snapshot from The Metadata Sheet for the Fetus Framework Dataset and NHOG Vietnam Dataset        | 102 |
| 3.18 | Distribution of Standard vs. Non-Standard Views in the Fetus Framework Dataset                   | 106 |
| 3.19 | Distribution of Crown–Rump Length (CRL)                                                          | 107 |
| 3.20 | Distribution of Nuchal Translucency (NT) Thickness                                               | 108 |
| 3.20 | Distribution of Gestational Age (Weeks)                                                          | 108 |
| 3.22 | Distribution of NT Thickness by NT Status                                                        | 109 |
| 3.23 | Relationship between Crown Rump Length (CRL) and Nuchal Translucency (NT) Thickness by NT Status | 110 |
| 3.24 | Relationship between GA and NT Thickness by NT Status                                            | 111 |
| 3.25 | Relationship between GA and CRL by NT Status                                                     | 112 |
| 3.26 | Heatmap for Relationships between key Clinical Measurements in The Dataset                       | 113 |
| 3.27 | A Snapshot of the Qualitative Dataset as Collected and Structured                                | 114 |
| 3.28 | Insights from Analysis of Nuchal Translucency                                                    | 117 |
| 3.29 | Word Cloud Showing the Most Frequently Mentioned Terms in NT-Related Reddit Posts                | 117 |
| 3.30 | The proposed workflow for automated NT measurement and classification                            | 118 |
| 3.31 | Dataset Splitting                                                                                | 119 |
| 3.32 | Architectural Design of DenseNet                                                                 | 121 |
| 3.33 | VGG16 Architectural                                                                              | 122 |
| 3.34 | VGG19 Architectural                                                                              | 123 |
| 3.35 | ResNet-18 Architectural                                                                          | 124 |
| 3.36 | ResNet-50 Architectural                                                                          | 126 |
| 3.37 | EfficientNetB7 Architectural                                                                     | 127 |
| 3.38 | Inception V3 Architectural                                                                       | 128 |
| 3.39 | The U-Net Architectural                                                                          | 129 |
| 3.40 | The Contracting Path of the U-Net                                                                | 131 |
| 3.41 | The Expanding Path of the U-Net                                                                  | 133 |

---

|      |                                                                                              |     |
|------|----------------------------------------------------------------------------------------------|-----|
| 3.42 | Model Parameters                                                                             | 134 |
| 3.43 | Model Hyperparameters                                                                        | 135 |
| 3.44 | Regularization                                                                               | 136 |
| 3.45 | Bias-Variance                                                                                | 137 |
| 3.46 | Types of Regularization in Machine Learning                                                  | 138 |
| 3.47 | Sigmoid Activation Function                                                                  | 140 |
| 3.48 | ReLU Activation Function                                                                     | 140 |
| 3.49 | SoftMax Activation Function                                                                  | 141 |
| 3.50 | The Difference between Epochs, Batch Size and Iteration                                      | 143 |
| 3.51 | Dropout on Neural Network Architecture                                                       | 144 |
| 3.52 | Learning Rate                                                                                | 147 |
| 3.53 | Confusion Matrix                                                                             | 148 |
| 3.54 | ROC Curve                                                                                    | 150 |
| 3.55 | First Approach Workflow                                                                      | 152 |
| 3.56 | Second Approach Workflow                                                                     | 156 |
| 4.1  | Training and Validation Performance of The DenseNet121 Model for NT Image-Quality Assessment | 159 |
| 4.2  | Confusion matrix for DenseNet121 in the first approach                                       | 159 |
| 4.3  | AUC for DenseNet121 in the first approach                                                    | 160 |
| 4.4  | Sample Predictions of Standard and Non-Standard Ultrasound Images Using DenseNet 121         | 161 |
| 4.5  | Original Ultrasound Image and Corresponding Grad-CAM Heatmap                                 | 162 |
| 4.6  | Loss and Accuracy vs. Epoch for VGG16                                                        | 163 |
| 4.7  | Confusion Matrix for VGG16                                                                   | 163 |
| 4.8  | AUC for VGG16                                                                                | 164 |
| 4.9  | Distribution of Predicted Probabilities for the Standard Class for VGG16                     | 165 |
| 4.10 | Loss and Accuracy vs. Epoch for VGG19                                                        | 165 |
| 4.11 | Confusion Matrix for VGG19                                                                   | 166 |
| 4.12 | AUC for VGG19                                                                                | 167 |
| 4.13 | Distribution of Predicted Probabilities for the Standard Class for VGG19                     | 168 |
| 4.14 | Loss and Accuracy vs. Epoch for ResNet18                                                     | 169 |
| 4.15 | Confusion Matrix for ResNet18                                                                | 169 |
| 4.16 | AUC for ResNet18                                                                             | 170 |

---

---

|      |                                                                                         |     |
|------|-----------------------------------------------------------------------------------------|-----|
| 4.17 | Distribution of Predicted Probabilities for the Standard Class for ResNet18             | 171 |
| 4.18 | Loss and Accuracy vs. Epoch for ResNet50                                                | 172 |
| 4.19 | Confusion Matrix for ResNet50                                                           | 172 |
| 4.20 | AUC for ResNet50                                                                        | 173 |
| 4.21 | Loss and Accuracy vs. Epoch for EfficientNetB7                                          | 174 |
| 4.22 | Confusion Matrix for EfficientNetB7                                                     | 175 |
| 4.23 | AUC for EfficientNetB7                                                                  | 175 |
| 4.24 | Loss and Accuracy vs. Epoch for InceptionV3                                             | 176 |
| 4.25 | Confusion Matrix for InceptionV3                                                        | 177 |
| 4.26 | AUC for InceptionV3                                                                     | 178 |
| 4.27 | Distribution of Predicted Probabilities for the Standard Class for InceptionV3          | 179 |
| 4.28 | Segmentation Results Using DenseNet: Comparison Between Ground Truth and Predicted Mask | 181 |
| 4.29 | Segmentation Results Using U-Net: Comparison Between Ground Truth and Predicted Mask    | 182 |
| 4.30 | Segmentation Quality—Dice/IoU for DenseNet and U-Net                                    | 186 |
| 4.31 | NT Measurement Workflow                                                                 | 187 |
| 4.32 | Example of Automated NT Measurement for a Normal Case                                   | 188 |

---

## List of Appendices

| Appendix # | Title of Appendix            | Page |
|------------|------------------------------|------|
| 1          | Research Facilitation Letter | 217  |

## List of Definitions of Abbreviations

| Abbreviations | Title                                                     |
|---------------|-----------------------------------------------------------|
| ADASYN        | Adaptive Synthetic Sampling Approach                      |
| ADF           | Anisotropic Diffusion Filtering                           |
| AI            | Artificial Intelligence                                   |
| ANFIS         | Adaptive Neuro-Fuzzy Inference System                     |
| ANN           | Artificial Neural Network                                 |
| ARPS          | Adaptive Risk Prediction System                           |
| ASGDA         | Adaptive Stochastic Gradient Descent Algorithm            |
| AUC           | Area Under the Curve                                      |
| BCE           | Binary Cross-Entropy Loss                                 |
| BM3D          | Block-Matching and 3D                                     |
| BPNN          | Backpropagation Neural Networks                           |
| CCE           | Categorical Cross-Entropy Loss                            |
| CED           | Coherence-Enhancing Diffusion                             |
| CHD           | Congenital Heart Disease                                  |
| CIS           | Circumstance Independent Searching                        |
| CLAHE         | Contrast-Limited Adaptive Histogram Equalization          |
| CNN           | Convolutional Neural Network                              |
| COCO          | Common Objects in Context                                 |
| CRL           | Crown-Rump Length                                         |
| CVIFLR        | Cascaded Voting Isolation Forests and Logistic Regression |
| CVS           | Chorionic Villus Sampling                                 |
| DCCP          | Deep Content and Contrastive Perception                   |
| DL            | Deep Learning                                             |
| DP            | Dynamic Programming                                       |
| DPI           | Dots Per Inch                                             |
| DS            | Down syndrome                                             |
| DVPI          | Ductus Venosus Pulsatility Index                          |
| DWT           | Discrete Wavelet Transform                                |
| EDA           | Exploratory Data Analysis                                 |
| FCM           | Combines Fuzzy Clustering                                 |

---

|          |                                            |
|----------|--------------------------------------------|
| FDA      | Food and Drug Administration               |
| FDR      | False Discovery Rate                       |
| FF       | Fetus Framework                            |
| FFNN     | Feed Forward Neural Network                |
| FMF      | The Fetal Medicine Foundation              |
| FN       | False Negatives                            |
| FNR      | False Negative Rate                        |
| FOR      | False Omission Rate                        |
| FP       | False Positive                             |
| FPR      | False-Positive Rate                        |
| GA       | Gestational Age                            |
| GD       | Gradient Descent                           |
| GDPR     | General Data Protection Regulation         |
| GRNN     | General Regression Neural Networks         |
| GRU      | Gated Recurrent Unit                       |
| GVF      | Gradient Vector Flow                       |
| Grad-CAM | Gradient-weighted Class Activation Mapping |
| HDN      | Hierarchical Detection Network             |
| HMMHA    | Hybrid Maxpool Matrix Histogram Analysis   |
| hCG      | Human Chorionic Gonadotropin               |
| HOG      | Histogram of Oriented Gradients            |
| IF       | Isolation Forests                          |
| IUGR     | Intrauterine Growth Restriction            |
| IoU      | Intersection Over Union                    |
| KDE      | The Kernel Density Estimate                |
| LMP      | Last Menstrual Period                      |
| LR       | Logistic Regression                        |
| LSTM     | Long Short-Term Memory                     |
| MAE      | Mean Absolute Error                        |
| ML       | Machine Learning                           |
| MLA      | Machine Learning Algorithm                 |
| MLP      | Multi-Layer Perceptron                     |

---

---

|        |                                                |
|--------|------------------------------------------------|
| MoM    | Multiple of Median                             |
| MRF    | Markov Random Field                            |
| NB     | Nasal Bone                                     |
| NCGNN  | Automated Nuchal Cord Graphical Neural Network |
| NHG    | National Healthcare Group                      |
| NHOG   | National Hospital of Obstetrics and Gynecology |
| NIPD   | Non-Invasive Prenatal Testing                  |
| NPV    | The Negative Predictive Value                  |
| NT     | Nuchal Translucency                            |
| PAPP-A | Pregnancy associated plasma protein-A          |
| PIA    | Phase Invariant Architecture                   |
| PNN    | Probabilistic Neural Networks                  |
| PPV    | Positive Predictive Value                      |
| PSO    | Particle Swarm Optimization                    |
| PTA    | Posterior Tibial Artery                        |
| PWCA   | Prolonged Wrapper Component Analysis           |
| RGB    | Red Green Blue                                 |
| ROC    | Receiver Operating Characteristic              |
| ROI    | Region of Interest                             |
| RPN    | Region Proposal Network                        |
| RR     | The Relative Risk                              |
| ReLU   | Rectified Linear Unit                          |
| ResNet | Residual Network                               |
| SAAM   | Semi-Automated Adjusted Measurement            |
| SD     | Standard Deviation                             |
| SGA    | Small-for-Gestational-Age                      |
| SGD    | Stochastic Gradient Descent                    |
| SIFT   | Scale-Invariant Feature Transform              |
| SRAD   | Speckle Reducing Anisotropic Diffusion         |
| SVM    | Support Vector Machines                        |
| THI    | Tissue Harmonic Imaging                        |
| TN     | True Negative                                  |

---

---

|      |                               |
|------|-------------------------------|
| TP   | True Positives                |
| TPR  | True Positive Rates           |
| UIT  | Ultrasound Imaging Techniques |
| VGG  | Visual Geometry Group         |
| ViT  | Vision Transformer            |
| VTK  | Visualization ToolKit         |
| WBCE | Weighted Binary Cross-Entropy |

---

## Chapter 1. Introduction

---

### 1.1 Overview

The rapid advancement of medical imaging technologies has revolutionized prenatal care, enabling early detection of fetal abnormalities to improve maternal and fetal outcomes (K. H. Nicolaides, 2011). Among these, Nuchal Translucency (NT) measurement during first-trimester ultrasound has emerged as a critical tool for assessing the risk of chromosomal abnormalities, such as Trisomy 21 (Down syndrome) (K. Nicolaides, 2004). However, manual NT measurement is subject to high variability due to dependency on operator expertise, fluctuations in image quality, and subjective interpretation, leading to inconsistencies in diagnostic accuracy (Salomon, et al., 2013). This study explores the application of Deep Learning (DL) techniques to automate and enhance the precision of NT measurement, addressing a significant gap in current prenatal diagnostic practices. By leveraging advanced computational models, this thesis aims to develop a reliable, scalable solution that reduces human error and improves clinical decision-making (Goodfellow et al., 2016). This study is motivated by the growing incidence of chromosomal disorders and the urgent need for standardized, efficient diagnostic tools to support healthcare providers in diverse clinical settings. The NT measurement process, including the Region of Interest (ROI) and the anatomical landmarks used for accurate thickness estimation, is illustrated in Figure 1.1.

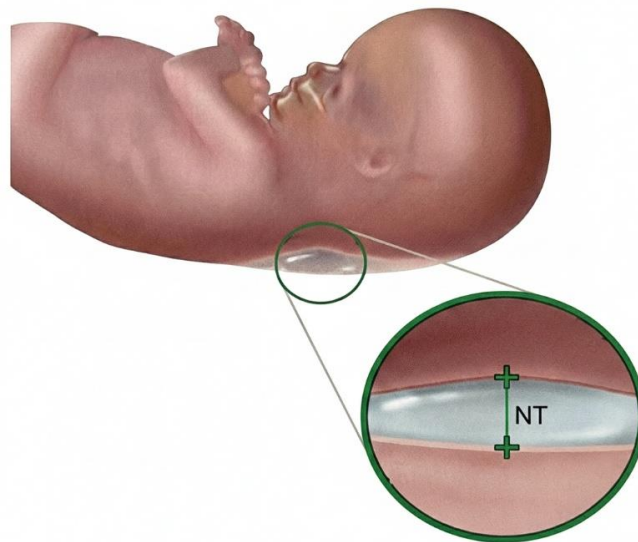


Figure (1.1) Nuchal Translucency (NT) Region

## **1.2 The Significance of the Study**

The significance of this study lies in its potential to transform prenatal diagnostics through the integration of deep learning with ultrasound imaging. The practical significance includes improving the accuracy and consistency of NT measurements, which can enhance early detection of chromosomal abnormalities, thereby facilitating timely interventions and counseling for expectant parents (Kagan et al., 2008). This study benefits healthcare providers, particularly sonographers and obstetricians, by offering an automated tool that reduces the dependency on specialized training and mitigates variability in measurements. The theoretical significance lies in contributing to the growing body of knowledge on Artificial Intelligence (AI) applications in medical imaging, particularly in addressing underexplored challenges in NT measurement automation (LeCun et al., 2015). Locally, regionally, and globally, this study addresses the societal need for accessible, high-quality prenatal care, aligning with global health priorities to reduce congenital anomaly-related morbidity (WHO, 2020).

## **1.3 Research Problem**

The precise measurement of NT in first-trimester ultrasound is indispensable for evaluating chromosomal abnormality risks, yet it is profoundly constrained by multifaceted challenges that undermine diagnostic reliability (Nicolaidis, 2011). A critical shortage of specialists proficient in NT measurement, coupled with widespread deficiencies in sonographer expertise, particularly in resource-limited settings, engenders significant inter-observer variability, resulting in inconsistent and potentially erroneous diagnostic outcomes (Kagan et al., 2008; Salomon et al., 2013). This variability is further exacerbated by the influence of suboptimal fetal positioning, which complicates accurate identification of the NT region, and by poor ultrasound image quality, often attributable to the limitations of low-quality imaging equipment prevalent in under-resourced facilities (Sonek et al., 2007). In addition, the lack of systematic image quality assessment further compromises diagnostic reliability, making it essential to introduce an automated deep learning-based step to classify NT scans into standard and non-standard categories, thereby ensuring that only diagnostically reliable images are utilized for downstream segmentation and measurement (L. Liu et al., 2025). These technical and human factors collectively impede the standardization of NT measurement, heightening the risk of misdiagnoses or

unnecessary invasive procedures with significant implications for maternal-fetal health (Nicolaidis, 2004). This study addresses a pivotal knowledge gap by proposing a deep learning model to enhance measurement precision, reduce reliance on specialist expertise, and ensure robustness across diverse clinical conditions, thereby advancing the accessibility and reliability of prenatal screening (Nicolaidis, 2004).

#### **1.4 Research Objectives**

This study aims to achieve the following objectives, which are aligned with the research problem and methodology:

- To develop a DL model for automated ultrasound image quality assessment, classifying NT images into Standard and Non-Standard prior to measurement.
- To develop a DL model capable of accurately measuring NT thickness in first-trimester ultrasound images.
- To evaluate the performance of the proposed model against manual NT measurements in terms of accuracy, sensitivity, and specificity.
- To assess the model's robustness across diverse ultrasound image qualities and fetal presentations.
- To propose a scalable framework for integrating the DL model into clinical ultrasound systems.

#### **1.5 Research Questions**

This study aims to address the following research questions, each aligned with the study's objectives and intended to guide the methodology and analysis:

- To what extent can a DL model accurately classify NT ultrasound images into Standard and Non-Standard quality categories, in accordance with established clinical guidelines for NT plane acceptability?
- To what extent can a DL model accurately measure NT thickness compared to traditional manual sonographic methods?
- What are the sensitivity and specificity of the proposed DL model in identifying abnormal NT measurements associated with chromosomal abnormalities?

- How does the performance of the DL model vary across different ultrasound image qualities and fetal positioning conditions?
- What are the key challenges and potential solutions for integrating the DL-based NT measurement tool into standard clinical ultrasound workflows?

## **1.6 Study Hypotheses**

Based on the research questions and objectives, the study proposes the following hypotheses to be tested through experimental evaluation:

### **1.6.1 Image Quality Assessment Hypothesis**

The proposed deep learning model will accurately classify NT ultrasound images into Standard and Non-Standard categories with an overall accuracy exceeding 90%, consistent with FMF and ISUOG image acceptability guidelines, thereby enabling reliable pre-screening for measurement suitability.

### **1.6.2 Measurement Accuracy Hypothesis**

The proposed DL model will yield significantly higher accuracy in measuring NT than manual sonographic methods, reducing inter-observer variability by at least 20%.

### **1.6.3 Diagnostic Performance Hypothesis**

The model will achieve sensitivity and specificity exceeding 90% in detecting abnormal NT measurements, matching or surpassing the diagnostic performance of expert sonographers.

### **1.6.4 Robustness Hypothesis**

The model will demonstrate robustness, maintaining consistent performance with less than 10% variance in accuracy across varying ultrasound image qualities and fetal positions.

### **1.6.5 Integration Feasibility Hypothesis**

The integration of the model into existing clinical ultrasound workflows will be feasible, with minimal hardware modifications and limited user training, ensuring seamless integration into current clinical workflows.

## **1.7 Limitations of the Study**

This study presents the key conceptual, substantive, temporal, spatial, population-related, and methodological constraints that define the scope of the study.

### **1.7.1 Conceptual Limitations**

The study is limited to the use of NT measurements as a sole marker for chromosomal abnormality detection, without incorporating additional prenatal screening parameters such as those used in triple screening maternal age, weight, Pregnancy-Associated Plasma Protein-A (PAPP-A), and free Beta-human Chorionic Gonadotropin ( $\beta$ -hCG) which are widely adopted in integrated risk assessment models.

### **1.7.2 Substantive Limitations**

The research is confined to first-trimester ultrasounds performed between 11 and 13 weeks of gestation, the period during which NT measurement is most clinically reliable (Nicolaides, 2004). As such, the findings are not generalizable to earlier or later stages of pregnancy.

### **1.7.3 Temporal Limitations**

The data utilized were collected in the year 2022, which may limit the applicability of the results to future scenarios where ultrasound technology or clinical protocols evolve significantly.

### **1.7.4 Spatial Limitations**

This study relies on ultrasound image datasets obtained from selected hospitals or clinical centers. These may not be fully representative of global practices in prenatal screening, potentially affecting the generalizability of the model to other geographic or clinical settings.

### **1.7.5 Population and Sample Limitations**

The dataset reflects a specific population group and may underrepresent rare or atypical fetal conditions. Consequently, the model's performance in such edge cases remains untested and may be suboptimal.

### **1.7.6 Methodological Limitations**

The performance of DL models is highly dependent on the size, quality, and diversity of the training data. A lack of sufficient variation in fetal presentation, image quality, or annotation precision may introduce bias or limit robustness (LeCun et al., 2015). Additionally, variation in ultrasound machines and annotation tools used may influence the consistency and reliability of the inputs.

## **1.8 Conceptual and procedural definitions**

This study outlines the essential terms and concepts used in the study to ensure clarity and consistency in interpretation.

### **1.8.1 Nuchal Translucency**

A subcutaneous fluid-filled space located at the back of the fetal neck, measured via ultrasound between 11 and 13 weeks of gestation. NT thickness serves as a key marker in first-trimester screening for chromosomal abnormalities, particularly trisomy 21 (Down syndrome) (Nicolaidis, 2004).

### **1.8.2 Deep Learning**

A specialized branch of machine learning that utilizes multi-layered Artificial Neural Networks (ANNs) to extract hierarchical features from complex datasets, such as medical images (LeCun et al., 2015).

### **1.8.3 Chromosomal Abnormalities**

Genetic disorders caused by numerical or structural changes in chromosomes, such as Trisomy 21 (Down syndrome), Trisomy 18 (Edwards syndrome), and Trisomy 13 (Patau syndrome) (Nicolaidis, 2011).

### **1.8.4 Image Quality Assessment (for NT Ultrasound)**

The process of classifying fetal ultrasound images into standard versus non-standard quality for NT measurement, based on predefined anatomical and acquisition criteria (e.g., midsagittal plane, visibility of NT/nasal bone, neutral fetal position, alignment of head and neck). Standard images fulfill all criteria and are deemed diagnostically reliable for downstream segmentation and measurement, while non-standard images lack one or more such criteria (L. Liu et al., 2025).

### 1.8.5 NT Measurement Accuracy

Defined as the degree to which the deep learning model's NT measurements match those obtained by expert sonographers.

## 1.9 Thesis Structure

The thesis follows a structured and rigorous academic progression, as illustrated in Figure 1.2. It begins with the Introduction, defining the research problem, significance, objectives, and scope. The Literature Review then examines existing studies to identify gaps and position the research within its academic context. The Dataset and Preprocessing chapter describe data sources and preparation methods, followed by Exploratory Data Analysis (EDA) to reveal patterns and inform methodological choices. The Methodology section outlines the research design and selected techniques, while the Results and Discussion interpret findings in relation to objectives and prior work. The thesis concludes with the Conclusion and Future Work, summarizing insights and suggesting future directions, followed by the References section compiling all cited studies.

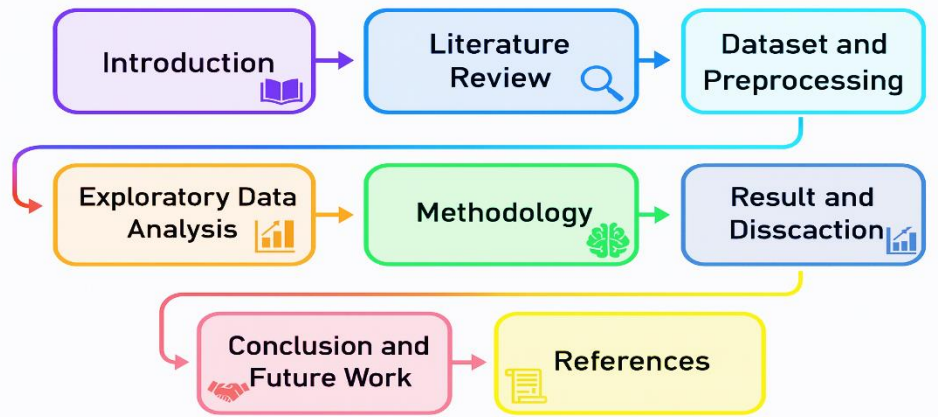


Figure (1.2) Thesis Structure

## 1.10 Conclusion

This chapter introduced the clinical significance and diagnostic value of NT measurement in first-trimester, emphasizing the limitations of manual methods such as operator dependency and image quality variability. It proposed DL as a promising approach to improve accuracy and consistency. The chapter also outlined the research problem, objectives, questions, hypotheses, and limitations, providing conceptual definitions. The next chapter review existing literature on traditional, semi-automated, and AI-based NT measurement and the clinical implications of NT abnormalities.

## Chapter 2. Literature Review

---

### 2.1 Introduction

Nuchal Translucency (NT) measurement during the first trimester is a fundamental component of modern prenatal screening and has been firmly established as a key predictor of chromosomal abnormalities, structural malformations, and adverse pregnancy outcomes. Since the pioneering work of Nicolaides and colleagues, NT has become an essential marker integrated into screening programs that combine ultrasound findings with maternal serum biochemistry to enhance risk prediction accuracy (Nicolaides, 2011, Salomon, et al., 2013).

Despite its clinical significance, NT measurement remains highly operator-dependent and sensitive to factors such as fetal position, image quality, and sonographer expertise, which contribute to substantial inter-observer variability and compromise diagnostic reliability. In recent years, machine learning and deep learning have emerged as promising solutions to improve consistency, automate NT detection and measurement, and reduce subjective dependence on manual interpretation (Yang et al., 2020, Topol, 2019).

This chapter provides a structured review of the literature on fetal development, screening technologies, NT physiology and pathology, and the evolution from manual to automated and AI-based NT measurement methods. The review highlights current limitations and identifies the research gaps that motivate the development of the deep learning framework proposed in this study.

### 2.2 Stages of Fetal Development

Before a baby is ready for the outside world, it undergoes several critical developmental phases. This chapter explores the stages of gestation, from fertilization to birth, as illustrated in Figure 2.1 below.

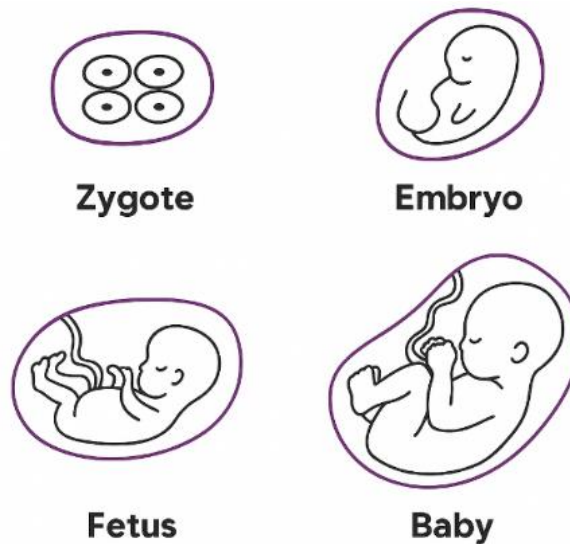


Figure (2.1) Fetal Development(Sumner, 2004)

- **Fertilization and Zygote Formation (Weeks 0-2)**

The process of human development begins when a sperm fertilizes an egg in the fallopian tube, resulting in the formation of a zygote. Over the following 24 hours, the zygote undergoes several rounds of cell division. Approximately five days after fertilization, it develops into a structure called a blastocyst. This blastocyst consists of two main parts: one part that will develop into the embryo and another that will form the placenta (Sumner, 2004).

- **Embryonic Stage (Weeks 3-8)**

- **Week 3:** The blastocyst implants into the uterine wall, initiating pregnancy.
- **Week 4:** Cell differentiation begins, forming the basic structure of the body.
- **Week 5:** The nervous and circulatory systems start developing, and the heart begins to form.
- **Week 6:** Facial features such as eyes, ears, and mouth start taking shape; limb buds appear.
- **Week 7:** Limb cartilage starts forming; hands begin developing.
- **Week 8:** Major organs and body systems are in place, and the umbilical cord fully forms. At the end of this stage, the embryo officially transitions into a fetus (Sumner, 2024).

➤ **Fetal Stage (Weeks 9-40)**

- **Weeks 9-12:** The fetus begins moving, practicing drinking and urinating amniotic fluid.
- **Weeks 13-16:** Fine hair grows, skin thickens, and fingerprints form.
- **Weeks 17-20:** The fetus is covered in lanugo hair and vernix for protection; movement becomes more noticeable.
- **Weeks 21-25:** The fetus can grasp, and lung development progresses significantly.
- **Weeks 26-30:** The nervous system nears completion; the fetus gains body fat and develops eyelashes.
- **Weeks 31-35:** The skin becomes opaque; major organs mature in preparation for birth.
- **Weeks 36-40:** The fetus gains final weight and positions itself for delivery. At 39 weeks, it is considered full-term and ready for birth (Sumner, 2024).

To visually illustrate these developmental milestones across gestation, Figure 2.2 presents a timeline infographic outlining the stages from fertilization to full-term birth.



Figure (2.2) Stages of Fetal Development (Kurbatova, n.d.)

In addition to the descriptive narrative and visual timeline, a summary table is presented to consolidate the key stages of prenatal development Table 2.1.

Table (2.1) Summary of Fetal Development Stages

| Stage                             | Time Frame (Weeks) | Key Developments                                                                           |
|-----------------------------------|--------------------|--------------------------------------------------------------------------------------------|
| <b>Zygote &amp; Fertilization</b> | 0–2                | Fertilization occurs; zygote divides into a blastocyst; implantation in the uterus.        |
| <b>Embryonic Stage</b>            | 3–8                | Organ formation begins; heart starts beating; limbs and facial features develop.           |
| <b>Fetal Stage</b>                | 9–40               | Growth and maturation; movement begins; lungs and brain develop; fetus prepares for birth. |

## 2.3 Screening

Screening is a structured method aimed at identifying individuals from a seemingly healthy population who may have a high enough risk of developing a specific health condition, thus justifying further investigation. Most screening tests are not diagnostic by themselves but are intended to flag potential cases. Individuals who test positive in screening are generally offered additional diagnostic procedures. However, such procedures cannot be extended to the entire population due to constraints like limited financial resources or potential risks associated with unnecessary testing (Äyräs, 2016).

The outcomes of a screening process are categorized into four groups: true positives, false positives, true negatives, and false negatives. Typically, individuals who initially test positive may later show normal results upon further diagnostic evaluation, which classifies them as false positives. On the other hand, those who are correctly identified as having the condition are considered true positives. False negatives are individuals who have the condition but were not identified by the screening, while true negatives are those without the condition and who receive a negative result in the screening (Äyräs, 2016).

Key measures in evaluating screening tests include **sensitivity**, which refers to the test's ability to correctly identify individuals with the condition (true positives), and **specificity**, which reflects its capacity to correctly identify those without the condition (true negatives). The **positive predictive value** indicates the probability that a person who tests positive actually has the condition (Äyräs, 2016), as illustrated in Table 2.2.

Table( 2.2) Confusion Matrix for Screening Test Outcomes (Äyräs, 2016)

|               | <b>Condition +</b> | <b>Condition -</b> |
|---------------|--------------------|--------------------|
| <b>Test +</b> | True positive (A)  | False positive (B) |
| <b>Test -</b> | False negative (C) | True negative (D)  |

The performance of a screening test can be quantitatively evaluated using the following metrics:

**True Positive (TP):** Cases in which the condition is genuinely present and the screening method correctly identifies it as positive (Parikh et al., 2008).

**False Positive (FP):** Cases where the screening test incorrectly classifies an individual as positive despite the absence of the condition (Altman & Bland, 1994).

**False Negative (FN):** Instances in which the condition is present but the test fails to detect it, leading to an incorrect negative result (Parikh et al., 2008).

**True Negative (TN):** Cases where the condition is genuinely absent and the screening test correctly identifies the result as negative (Altman & Bland, 1994).

➤ Sensitivity (Recall) =  $\frac{TP}{TP+FN}$

➤ Specificity =  $\frac{TN}{TN+FP}$

➤ Positive Predictive Value (PPV) =  $\frac{TP}{TP+FP}$

➤ Accuracy =  $\frac{TP}{TP+TN+FP+FN}$

➤ F1-score (F-measure) =  $\frac{2.TP}{2.TP+FP+FN}$

### 2.3.1 Prenatal Screening

Prenatal screening refers to the tests available for expectant parents who wish to gather information regarding the health of their pregnancy. While the majority of pregnancies are associated with a low risk of birth defects and genetic disorders, some women may have a family or medical history that categorizes their pregnancy as "high risk." It is important to note that even in "low risk" cases, approximately two to three out of every 100 infants may be born with a birth defect. Prenatal screening tests are designed to identify pregnancies that could be at an elevated risk for specific conditions, thereby necessitating specialized care. Commonly offered screening tests during pregnancy include the first trimester screen, the second trimester screen (quad screen), and a third screening (Zhytnik et al., 2020), which are visually summarized in Figure 2.3.

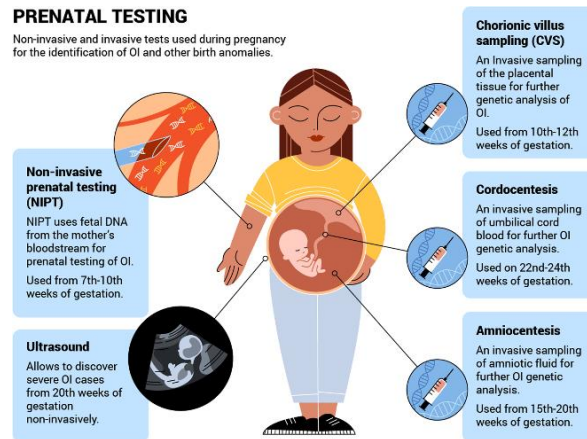


Figure (2.3) Overview of Prenatal Testing Options for Members of Families with OI Risk. NIPT – non-Invasive Prenatal Testing, Ultrasound, CVS - Chorionic Villus Sampling, Cordocentesis, Amniocentesis (Zhytnik et al., 2020)

### 2.3.2 First Trimester Screening

During all subsequent prenatal appointments, the gynecologist will conduct a physical examination and record the patient's weight and blood pressure. The first-trimester ultrasound, often the woman's initial ultrasound, is typically performed between 11 and 13 weeks of gestation and is usually conducted vaginally.

The purposes of this ultrasound include:

- Determining the number of embryos and gestational sacs.
- Detecting the embryo's heartbeat.
- Measuring the embryo to estimate gestational age in conjunction with the Last Menstrual Period (LMP).
- Assessing Nuchal Translucency.
- Evaluating the embryo's morphology.
- Examining the uterus and surrounding structures (Zaira et al., 2022).

### 2.3.3 Triple Screening

First trimester combined screening, commonly referred to as triple screening, is a biochemical assessment routinely conducted for all expectant mothers to evaluate the risk of congenital anomalies in their infants. This screening method integrates multiple markers to estimate risk levels, and its components are visually represented in Figure 2.4.

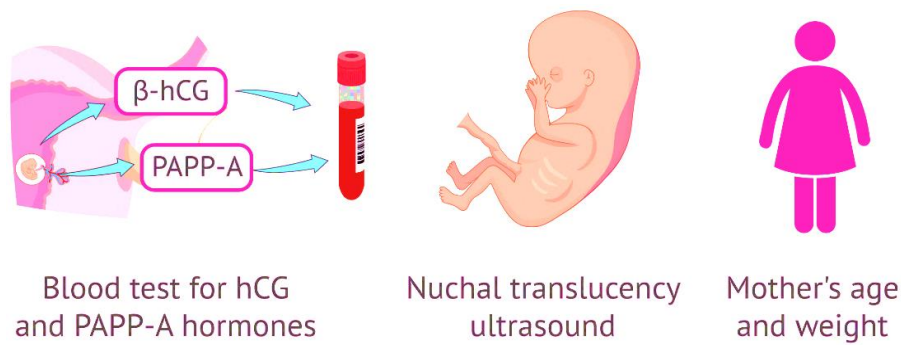


Figure (2.4) Triple Screening (Zaira et al., 2022)

This risk assessment is based on several factors collected by the gynecologist during the first-trimester blood test and ultrasound, including:

- Nuchal fold measurement or Nuchal Translucency.
- Biochemical indicators: (PAPP-A) and free beta-hCG.
- Maternal age, weight, number of fetuses, and other relevant factors.

The results are then compared against established reference tables that reflect normal values corresponding to the gestational age (Zaira et al., 2022). Table 2.3 below shows detection and false-positive rates for different screening methods.

Table (2.3) Performance of Different Methods of Screening for Trisomy 21(K. H. Nicolaides, 2011)

| Method of Screening                                     | Detection Rate (%) | False-Positive Rate (%) |
|---------------------------------------------------------|--------------------|-------------------------|
| MA                                                      | 30                 | 5                       |
| MA + Fetal NT                                           | 75–80              | 5                       |
| MA + Serum free $\beta$ -hCG + PAPP-A                   | 60–70              | 5                       |
| MA + NT + free $\beta$ -hCG + PAPP-A<br>(Combined Test) | 85–95              | 5                       |
| Combined Test + NB//DVF /Tricuspid Flow                 | 93–96              | 2.5                     |

## 2.4 First-Trimester Anatomy Assessment

During the first trimester, ultrasound assessment focuses on identifying key anatomical structures that support accurate NT measurement, including the mid-sagittal plane, nasal bone, brainstem, and facial profile. Proper visualization ensures compliance with FMF and ISUOG guidelines, reducing measurement

errors and improving screening accuracy. Anatomical assessment therefore serves as a prerequisite to reliable NT measurement and subsequently impacts the performance of automated detection and segmentation models. Figure 2.5 provides a visual representation of these anatomical landmarks as seen during the ultrasound scan, highlighting the importance of early and detailed imaging in ensuring accurate diagnosis and timely intervention (Sepulveda et al., 2013).

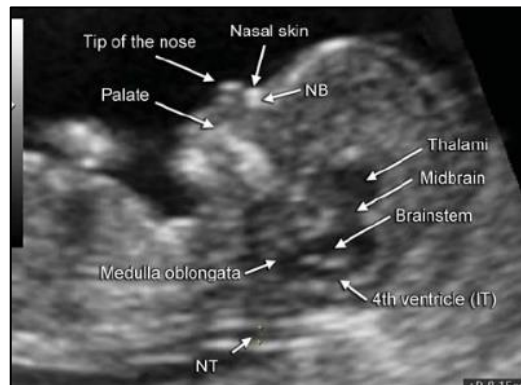


Figure (2.5) Normal First-Trimester Anatomic Features of the Fetal Brain, Face and Neck as Depicted by Two-Dimensional Ultrasound from the Midsagittal Section (Sepulveda et al., 2013)

- **Tip of the Nose**

The anterior part of the fetal nose helps evaluate facial development. Abnormalities here may suggest midfacial hypoplasia or craniofacial syndromes (Sivan et al., 1997).

- **Nasal Skin and Nasal Bone**

The nasal bone's presence, length, and echogenicity are significant markers for aneuploidies, including trisomy 21. NB absence or hypoplasia is a high-risk indicator (Cicero et al., 2001).

- **Palate**

Evaluating the palate for continuity is crucial for identifying cleft palate, which is associated with several syndromes like Pierre Robin sequence (M. C. Lee et al., 2003).

- **Thalami**

These paired structures in the brain are assessed for symmetry and abnormalities, which may indicate neural tube defects or holoprosencephaly (Persico et al., 2010) .

- **Midbrain**

The midbrain controls motor functions and visual/auditory pathways. Abnormal development in this region is linked to severe brain malformations (G & K.H., 1999).

- **Brainstem**

Assessed to rule out abnormalities affecting autonomic regulation, such as Joubert syndrome or Chiari malformations (Volpe et al., 2021).

- **Medulla Oblongata**

This part of the brainstem regulates involuntary processes like respiration and heart rate. Abnormalities may suggest brainstem compression or Arnold-Chiari malformations (Glenn & Barkovich, 2006).

- **4th Ventricle (IT)**

Visualized to assess posterior fossa abnormalities, such as Dandy-Walker malformation or ventriculomegaly (Aubry et al., 2003).

## **2.5 Nuchal Translucency**

Nuchal Translucency refers to the fluid-filled space beneath the skin at the back of the fetal neck, measured during the first trimester of pregnancy. The measurement is performed between 11 weeks and 13 weeks + six days of gestation, or when the Crown-Rump Length (CRL) ranges between 45- and 84-mm. NT assessment plays a crucial role in estimating the risk of chromosomal abnormalities, such as Down syndrome (DS), and may also serve as an indicator of other fetal anomalies. For optimal screening accuracy, NT measurement is most effective when combined with maternal age and biochemical markers in maternal blood, including (hCG) and (PAPP-A) (Abuhamad, 2014). The anatomical position and measurement technique of the NT region, as visualized in sagittal ultrasound imaging, are illustrated in Figure 2.6, which depicts the placement of calipers between the inner borders of the nuchal space.



Figure (2.6) Fetal NT Thickness (Khatter et al., 2023)

## 2.6 NT Measurement Success Rates Across Gestational Ages

The success rate of obtaining a reliable NT measurement is closely tied to gestational age. A study involving 101 singleton pregnancies affected by DS revealed that the feasibility of NT measurement significantly varies between six and 16 weeks of gestation. Overall, the failure rate was 9%, but it showed clear trends across different gestational windows.

- **Early Gestation (6–9 weeks):** NT measurement is often unsuccessful due to the small size and difficult visualization of the fetus. Failure rates were highest at 6–7 weeks (83%), slightly decreased at 8 weeks (61%), and improved to 33% by 9 weeks.
- **Optimal Window (10–12 weeks):** The most successful NT measurements occur during this period. Failure rates decreased to 13% at 10 weeks and reached the lowest point at 12 weeks (7%). This window corresponds to the CRL range of 45–84 mm, recommended for accurate NT assessment, where success rates can reach 98–100% in optimal conditions.
- **Later Gestation (13–16 weeks):** Success rates decline again after 12 weeks. The failure rate was 8% at 13 weeks but increased significantly to 14% at 14 weeks, 32% at 15 weeks, and 65% at 16 weeks. These challenges are attributed to changes in fetal positioning and the natural resolution of NT as the lymphatic system matures.

This analysis underscores the importance of timing NT measurements within the 11–13+6-week window to maximize diagnostic accuracy and minimize failure rates (Wald et al., 2003), as illustrated in Figure 2.7.

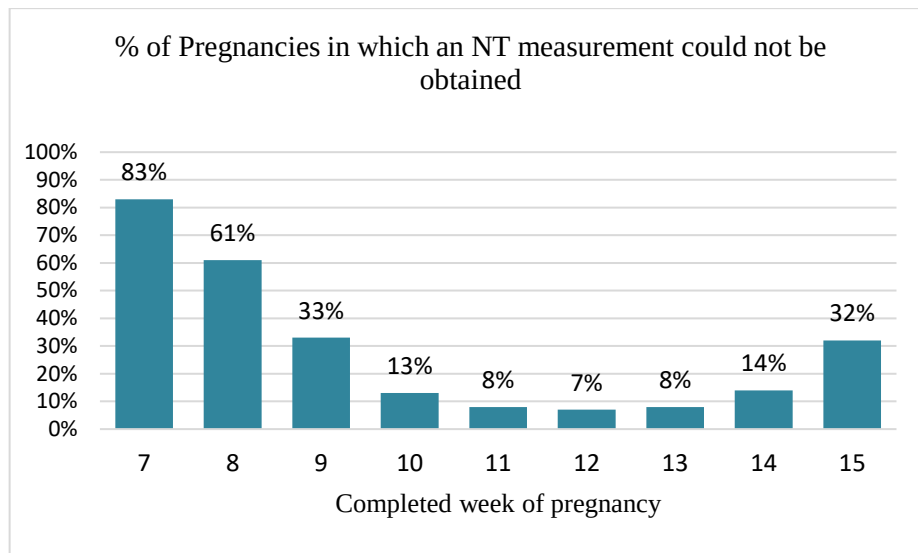


Figure (2.7) Week of Pregnancy (Chitty, 2003)

## 2.7 The Transient Nature of NT between 10-14 weeks

During this period, the fetal lymphatic system is in the process of development, and the placenta exhibits high peripheral resistance. This combination leads to the accumulation of excess fluid behind the neck, resulting in an increased NT measurement. After 14 weeks, the lymphatic system matures sufficiently to effectively drain this excess fluid into the placental circulation. Consequently, any fetal abnormalities that may cause fluid accumulation tend to resolve on their own, making them less detectable through NT measurement (Souka et al., 2005).

### 2.7.1 Reason for selecting up to 14 weeks for NT

- It allows for safer and earlier termination options for women if the fetus is found to be affected.
- The occurrence of increased NT is less common between 14 and 18 weeks compared to before 14 weeks of gestation in affected fetuses.
- The fetus typically assumes a horizontal position during the 10 to 14-week period, which is optimal for accurate measurement, achieving a success rate of 98% to 100%. Beyond 14 weeks, obtaining suitable images becomes challenging as the fetus often shifts to a vertical position (Benacerraf et al., 1987, K. H. Nicolaides et al., 1992, Comas et al., 2002).

### **2.7.2 Reason for selecting 10 weeks as lower limit**

- Prior to this period, Chorionic Villus Sampling (CVS) was the primary diagnostic test, and performing CVS before 10 weeks was linked to transverse limb reduction defects.
- Fetal structures become visible after 10 weeks, allowing for the identification of any abnormalities during the NT scan.

## **2.8 Pathological correlation between increased NT and euploid fetuses**

Several studies have investigated the pathological implications of increased NT in euploid fetuses, highlighting its correlation with adverse pregnancy outcomes. Based on data from three studies encompassing 4,991 euploid fetuses with increased NT measurements, the risk of miscarriage or fetal demise has been shown to rise significantly with NT thickness. Specifically, while the prevalence of fetal loss is approximately 1.6% in cases where NT measures between the 95th and 99th percentiles, it escalates to nearly 20% when NT reaches or exceeds 6.5 mm. (Souka et al., 2005, Bilardo et al., 2007). Furthermore, the relationship between NT thickness and fetal demise appears to follow an exponential trend, though existing estimates may be conservative. This is partly due to the fact that a substantial proportion of cases with progressive fetal hydrops, approximately 30%, result in elective termination before intrauterine fetal death occurs (Bilardo et al., 2007).

In addition to pregnancy loss, an increased NT measurement has been associated with a higher likelihood of congenital abnormalities requiring medical or surgical intervention, as well as neurodevelopmental impairments. The prevalence of these conditions has been reported to increase from 1.6% in fetuses with NT below the 95th percentile to 2.5% (Michailidis & Economides, 2001), in those between the 95th and 99th percentiles, reaching approximately 45% in cases where NT measures 6.5 mm or greater. These findings underscore the significance of NT assessment in early pregnancy as a critical marker for both structural and functional fetal anomalies (Souka et al., 1998, Souka et al., 2001).

## **2.9 Etiology of Increased NT**

The exact mechanism underlying NT remains incompletely understood. The phenomenon of increased NT is a transient finding observed in the first trimester, yet its underlying etiology is complex and multifactorial. Several potential mechanisms have been proposed, each contributing to NT enlargement through distinct yet possibly overlapping pathways.

### **2.9.1 Hemodynamic Disturbances and Cardiac Dysfunction**

A widely supported hypothesis is that even a minor degree of cardiac dysfunction in early fetal life can lead to NT enlargement. In the first trimester, the fetal heart is particularly sensitive to hemodynamic imbalances. A small impairment in diastolic function possibly due to developmental or structural abnormalities of the heart and great arteries—may result in venous stasis. This is observed as abnormal DVF (with absent or reversed a-wave during atrial contraction) and is documented in both chromosomally abnormal fetuses and those with isolated cardiac defects. Additionally, increased expression of cardiac peptides, such as atrial and brain natriuretic peptides, in trisomic fetuses suggests that even subtle cardiac strain might contribute to fluid accumulation in the nuchal region (Haak & van Vugt, 2003).

### **2.9.2 Alterations in the Extracellular Matrix**

Another line of evidence implicates changes in the extracellular matrix (ECM) as a significant factor. In particular, studies have shown that the nuchal skin of fetuses with chromosomal abnormalities (for example, trisomy 21) has increased levels of hyaluronic acid—a hydrophilic glycosaminoglycan that has a high capacity to bind water. This increased deposition can lead to substantial fluid retention within the ECM. Moreover, many of the ECM component proteins, including various types of collagens, are encoded on chromosomes 21, 18, and 13. Consequently, gene dosage effects in these syndromes can lead to alterations in collagen metabolism and ECM structure, contributing to the observed increase in NT (Haak MC, 2003).

### 2.9.3 Perturbed Lymphangiogenesis and Endothelial Dysfunction

Abnormal development of the lymphatic system has also been proposed as an etiological factor. Under normal conditions, lymph angiogenesis involves the budding of lymphatic channels from the venous system, followed by their maturation and eventual separation from the venous circulation except in areas such as the jugulo-axillary sacs, which continue to drain into the venous system. A delay or disturbance in this developmental process, including endothelial dysfunction, may result in persistent dilation of the jugular lymphatic sacs and impaired drainage of interstitial fluid from the nuchal area. In conditions such as Turner syndrome, for example, there is documented lymphatic hypoplasia, which further supports the role of lymphatic dysgenesis in NT enlargement (Haak MC, 2003).

### 2.9.4 Additional Contributing Factors

Beyond the primary mechanisms discussed above, several additional factors have been associated with increased NT:

- **Venous Congestion:** Mechanical factors that lead to venous congestion in the head and neck have been identified. For instance, in cases of superior mediastinal compression due to diaphragmatic hernia or a narrow thoracic cavity as seen in skeletal dysplasias, increased NT can occur secondary to impaired venous return.
- **Fetal Anemia and Hypoproteinemia:** Conditions such as fetal anemia or hypoproteinemia can alter the osmotic balance and blood viscosity, further contributing to the accumulation of interstitial fluid in the nuchal region.
- **Congenital Infections:** Infections, including parvovirus B19, may indirectly increase NT by inducing cardiac dysfunction or anemia. Although less frequently encountered in the first trimester, these infectious etiologies provide an additional pathway through which NT can be elevated (Nicolaidis KH H. V., 2002).

## **2.10 The Factors Assess at the Frist Trimester Nuchal Transcalency Scan**

The first-trimester NT scan involves assessing several fetal and maternal factors that enhance the accuracy of early screening. These parameters complement the NT measurement and help refine the prediction of chromosomal and structural abnormalities. Together, they provide a more complete picture of fetal development during early gestation and improve risk-stratification for clinical decision-making (Nicolaides, 2004; Salomon et al., 2013).

### **2.10.1 Crown Rump Length**

The relationship between NT thickness (mm) and CRL (mm) is depicted using percentile distributions. The NT measurement, a critical parameter in first-trimester ultrasound screening, is plotted against CRL, representing embryonic growth. The graph delineates the 5th, 25th, 50th, 75th, and 95th percentiles, demonstrating the natural variation in NT thickness as fetal size increases. As shown in Figure 2.8, NT thickness gradually increases with CRL, reflecting normal physiological changes during early gestation. The 50th percentile (median) represents the expected NT value for a given CRL, while the 5th and 95th percentiles define the lower and upper reference ranges, respectively. Clinically, NT measurements exceeding the 95th percentile are considered significantly elevated and are associated with an increased risk of chromosomal abnormalities, particularly trisomy 21 (Down syndrome), trisomy 18, and Turner syndrome (Knipe et al., 2008).

This percentile-based reference system is essential for risk assessment in prenatal screening programs. It aids in distinguishing normal from abnormal NT values, facilitating early detection of potential fetal anomalies. ML algorithms aimed at automating NT measurement should integrate percentile-based thresholds to enhance diagnostic accuracy and reliability (Knipe et al., 2008).

This percentile-based reference system is essential for risk assessment in prenatal screening programs. It aids in distinguishing normal from abnormal NT values, facilitating early detection of potential fetal anomalies. ML algorithms aimed at automating NT measurement should integrate percentile-based thresholds to enhance diagnostic accuracy and reliability.

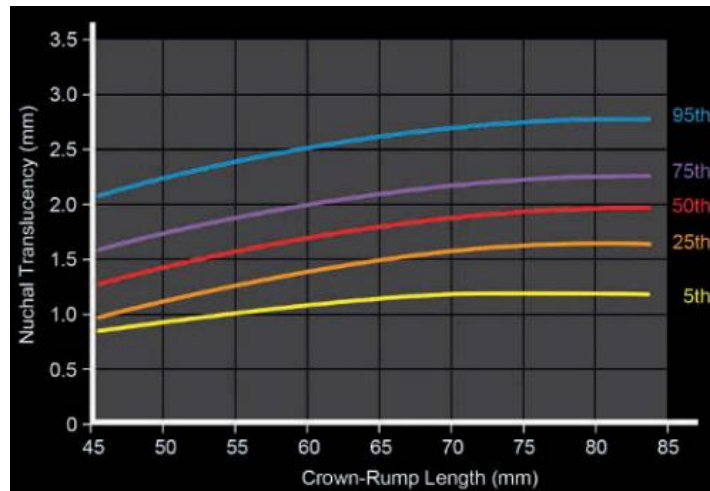


Figure (2.8) NT Measurement Percentiles by Crown-Rump Length (CRL) in First-Trimester Screening (Woodward, et al., 2016)

### 2.10.2 Gestational Age

The relationship between CRL (mm) and Gestational Age (weeks) during the first trimester of pregnancy is illustrated in Figure 2.9. CRL is a fundamental biometric parameter in early pregnancy ultrasound assessments, primarily used to estimate gestational age, evaluate fetal growth, and identify potential developmental abnormalities (Salomon et al., 2011).

The graph demonstrates a nonlinear increase in CRL as gestational age advances, with growth appearing exponential during the early weeks. The central curve represents the median CRL value, while the red dashed lines indicate the range of expected variability, likely corresponding to the 5th and 95th percentiles or a standard deviation range. This percentile-based distribution is essential for determining whether fetal growth falls within normal parameters.

CRL measurement is considered the most reliable method for gestational age estimation before 14 weeks of pregnancy, with an accuracy of  $\pm 3$  to 5 days (Hadlock et al., 1992). Deviations from expected CRL values may indicate Intrauterine Growth Restriction (IUGR), embryonic developmental delay, or chromosomal anomalies, necessitating further evaluation through additional sonographic and biochemical markers (Verburg et al., 2008).



Figure (2.9) Crown-Rump Length by Gestational Age in Weeks During the First Trimester (Westerway et al., 2000)

### 2.10.3 Maternal Age

In 1909, Shuttleworth was the first to observe that mothers of children with DS tended to be older than the average. Chromosomally abnormal fetuses have a higher likelihood of dying in utero compared to normal fetuses; consequently, the risk diminishes with advancing gestational age while it increases with the age of the mother, as shown in Figure 2.10.

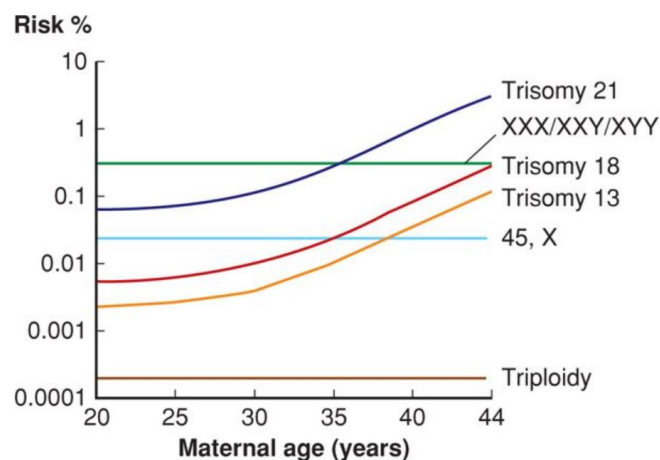


Figure (2.10) Maternal Age Related for Chromosomal Abnormalities (*OBSTETRICS*, 2017)

Previously, the screening for Down's Syndrome relied primarily on advanced maternal age. Under this criterion, only 30% of infants with Down's Syndrome were identified (Souka et al., 2005). The estimated risk for trisomies 21, 18, and 13 in relation to maternal age and gestation is illustrated in Figure 2.11 (Snijders et al., 1999).

| Maternal age (yrs) | Trisomy 21 Gestation (wks) |      |      |      | Trisomy 18 Gestation (wks) |      |      |       | Trisomy 13 Gestation (wks) |       |       |       |
|--------------------|----------------------------|------|------|------|----------------------------|------|------|-------|----------------------------|-------|-------|-------|
|                    | 12                         | 16   | 20   | 40   | 12                         | 16   | 20   | 40    | 12                         | 16    | 20    | 40    |
| 20                 | 1068                       | 1200 | 1295 | 1527 | 2484                       | 3590 | 4897 | 18013 | 7826                       | 11042 | 14656 | 42423 |
| 25                 | 946                        | 1062 | 1147 | 1352 | 2200                       | 3179 | 4336 | 15951 | 6930                       | 9778  | 12978 | 37567 |
| 30                 | 626                        | 703  | 759  | 895  | 1456                       | 2103 | 2869 | 10554 | 4585                       | 6470  | 8587  | 24856 |
| 31                 | 543                        | 610  | 658  | 776  | 1263                       | 1825 | 2490 | 9160  | 3980                       | 5615  | 7453  | 21573 |
| 32                 | 461                        | 518  | 559  | 659  | 1072                       | 1549 | 2114 | 7775  | 3378                       | 4766  | 6326  | 18311 |
| 33                 | 383                        | 430  | 464  | 547  | 891                        | 1287 | 1755 | 6458  | 2806                       | 3959  | 5254  | 15209 |
| 34                 | 312                        | 350  | 378  | 446  | 725                        | 1047 | 1429 | 5256  | 2284                       | 3222  | 4277  | 12380 |
| 35                 | 249                        | 280  | 302  | 356  | 580                        | 837  | 1142 | 4202  | 1826                       | 2576  | 3419  | 9876  |
| 36                 | 196                        | 220  | 238  | 280  | 456                        | 659  | 899  | 3307  | 1437                       | 2027  | 2691  | 7788  |
| 37                 | 152                        | 171  | 185  | 218  | 354                        | 512  | 698  | 2569  | 1116                       | 1575  | 2090  | 6050  |
| 38                 | 117                        | 131  | 142  | 167  | 272                        | 393  | 537  | 1974  | 858                        | 1210  | 1606  | 4650  |
| 39                 | 89                         | 100  | 108  | 128  | 208                        | 300  | 409  | 1505  | 654                        | 922   | 1224  | 3544  |
| 40                 | 68                         | 76   | 82   | 97   | 157                        | 227  | 310  | 1139  | 495                        | 698   | 927   | 2683  |
| 41                 | 51                         | 57   | 62   | 73   | 118                        | 171  | 233  | 858   | 373                        | 526   | 698   | 2020  |
| 42                 | 38                         | 43   | 46   | 55   | 89                         | 128  | 175  | 644   | 280                        | 395   | 524   | 1516  |

Figure (2.11) Estimated Risk for Trisomies 21, 18 and 13 (1/number given in the table) in Relation to Maternal Age and Gestation (*Nicolaides H, 2004*)

In 1990, the screening process for Trisomy 21 was enhanced by integrating maternal age with fetal NT measurements taken between 10 and 14 weeks of gestation. This combined approach increased the detection rate of affected fetuses with Trisomy 21 to 75%, as illustrated in Figure 2.12.

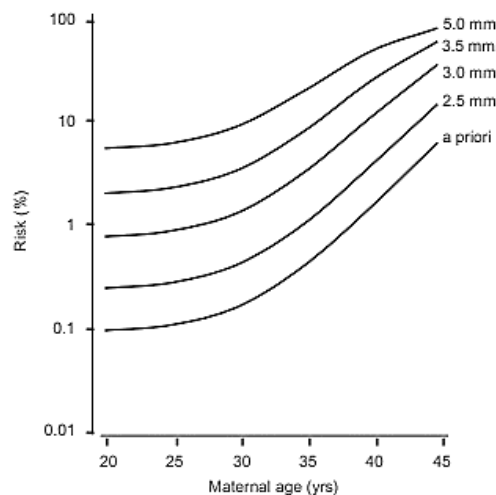


Figure (2.12) Maternal Age-Related Risk for Trisomy 21 at 12 Weeks of Gestation (a priori) and the Effect of Fetal Nuchal Translucency Thickness (NT) (*Nicolaides, 2004*)

#### 2.10.4 Co-variables

NT measurements extend beyond their commonly recognized role in DS screening. An increased NT is associated with a variety of both structural and genetic abnormalities. Since the advent of first-trimester screening, a considerable

body of data has been gathered on the use of ultrasound and biochemical markers in early pregnancy assessments. This thesis investigates several co-variables that impact the effectiveness of first-trimester screening, with particular emphasis on factors such as fetal gender and maternal smoking. Additionally, the potential role of ductus venosus and hepatic artery measurements as markers in contingent screening programs is explored and discussed, further enhancing the predictive capabilities of early prenatal assessments.

#### **2.10.5 Gender**

This study aimed to investigate the influence of fetal gender on pregnancy outcomes in fetuses with enlarged NT. The researchers retrospectively reviewed the pregnancy outcomes of all women who underwent NT measurement at their institution between January 2000 and November 2007. Separate analyses were conducted for fetuses with normal NT and those with enlarged NT ( $\geq 95$ th percentile).

Among the fetuses with enlarged NT, 365 were male (57.4%) and 271 were female (42.6%), a statistically significant difference ( $P = 0.001$ ). In this group, a normal pregnancy outcome was observed in 187 out of 332 male fetuses (56.3%) and 98 out of 249 female fetuses (39.4%), with a significantly better outcome for male fetuses ( $P < 0.001$ ). The Relative Risk (RR) for an adverse outcome in males was 0.72. For chromosomally normal male fetuses with enlarged NT, 80% had an uneventful pregnancy outcome. This percentage increased to 90% when considering only male fetuses with NT measurements between the 95th and 99th percentiles and normal karyotype (RR for adverse outcome for male gender, 0.47). The study concluded that there are significantly more male fetuses with enlarged NT. Male fetuses with enlarged NT and normal chromosomes have nearly twice the chance of a favorable pregnancy outcome compared to females. The authors suggest that a minimal degree of NT enlargement in male fetuses without genetic or structural anomalies may reflect either accelerated growth or a maturational delay in the cardiovascular system, which is more common in males and can lead to moderately increased nuchal fluid accumulation (Timmerman et al., 2009). The main findings are summarized in Table 2.4.

Table (2.4) Pregnancy Outcomes in Male and Female Fetuses with Enlarged NT

| <b>Outcome</b>                                             | <b>Male Fetuses</b> | <b>Female Fetuses</b> | <b>P-value</b> | <b>Relative Risk (RR)</b> |
|------------------------------------------------------------|---------------------|-----------------------|----------------|---------------------------|
| <b>Total Cases</b>                                         | 365<br>(57.4%)      | 271<br>(42.6%)        | 0.001          | —                         |
| <b>Normal Outcomes (All Enlarged NT Cases)</b>             | 56.3%               | 39.4%                 | < 0.001        | 0.72                      |
| <b>Normal Outcomes (Mildly Enlarged NT, Chrom. Normal)</b> | 90%                 | Lower                 | —              | 0.47                      |

### 2.10.6 Maternal Smoking

This study aimed to evaluate the impact of fetal gender and maternal smoking on the performance of first-trimester DS screening in the Netherlands. The researchers examined the prenatal database of a Fetal Medicine Unit, focusing on singleton pregnancies that underwent the first trimester combined test (NT, free  $\beta$ -hCG, and PAPP-A) to assess the risk of DS between January 2000 and October 2008. The study analyzed differences in screening performance between male and female fetuses using Chi-squared and Mann-Whitney U tests, and compared smokers and non-smokers using likelihood ratios, corrected for a-priori risk. The study included 4,538 fetuses: 2,350 male (52%) and 2,188 female (48%) fetuses. It was found that women carrying female fetuses had a 22% higher free  $\beta$ -hCG level compared to those carrying male fetuses. The NT was 2% larger in male fetuses. While female fetuses were 1.2 times more likely to be screen positive, this difference was not statistically significant ( $P=0.08$ ). Smoking was reported by 249 women (6%). Women who smoked had a 21% lower PAPP-A level and were more likely to be classified as screen positive compared to non-smokers (likelihood ratio = 1.3,  $P=0.002$ ). The study concluded that maternal smoking should be considered in first-trimester risk assessment to prevent smokers from being unnecessarily classified as high risk for DS. Additionally, information on fetal gender could be incorporated as further refinement in the screening algorithm, especially in a sequential screening approach. The study highlighted that co-variable in first-trimester screening, such as maternal smoking, can significantly impact risk assessment, particularly for women with intermediate-high a priori risk for DS. These women are more sensitive to changes in risk and thus more likely to be classified as screen positive (Miron et al., 2008).

### **2.10.7 Plasma Protein-A (PAPP-A), free $\beta$ -hCG—and birth Weight**

This study aimed to investigate the association between first-trimester screening variables NT, Pregnancy-Associated Plasma Protein-A (PAPP-A), and free  $\beta$ -hCG—and birth weight, with a particular focus on predicting macrosomia. The researchers reviewed the database of a Fetal Medicine Unit, selecting singleton pregnancies that underwent first-trimester DS screening, while excluding those with chromosomal or structural defects. Birth weight percentiles were adjusted for gestational age at delivery, parity, and gender. The study included 6,503 fetuses. Among them, 315 (4.8%) had an enlarged NT ( $\geq 95$ th percentile), and 6.6% had low PAPP-A ( $\leq 5$ th percentile). The median birth weight was significantly lower in fetuses with low PAPP-A (41st percentile vs. 52nd percentile). Macrosomia, defined as an excessively high birth weight, occurred more frequently in fetuses with an enlarged NT (11% vs. 7%). The study concluded that, in fetuses with normal outcomes, both PAPP-A multiples of the median (MoM) and NT MoM are significantly associated with birth weight. Specifically, an enlarged NT was found to be linked with a higher likelihood of macrosomia (Timmerman et al., 2014).

### **2.10.8 Ductus Venosus**

This study evaluates the use of the Ductus Venosus Pulsatility Index (DVPI) as an additional marker in first-trimester antenatal screening for Down's syndrome (DS), comparing its effectiveness when used with the Combined test (NT, PAPP-A, and free  $\beta$ -hCG) and the Integrated test (first-trimester NT and PAPP-A combined with second-trimester markers). The study analyzed data from 66 DS and 7,184 unaffected pregnancies, finding that the median DVPI in DS pregnancies was 1.55 times higher than in unaffected pregnancies. As a standalone marker, DVPI achieved a 62% Detection Rate (DR) for a 5% False-Positive Rate (FPR). When added to the Combined and Integrated tests, DVPI significantly reduced the FPR—for example, at a 90% DR, the FPR dropped from 8.5% to 4.6% for the Combined test and from 2.0% to 1.1% for the Integrated test. This improvement in screening performance also reduced the number of procedure-related fetal losses and maintained cost-effectiveness. The study

concludes that incorporating DVPI into first-trimester screening enhances efficacy and safety, particularly in high-risk populations, and recommends its use in routine screening practice where feasible (Borrell et al., 2009).

## **2.11 Abnormalities Associated with Increased NT**

The study found that 5.9% of fetuses had increased NT thickness ( $\geq 95$ th percentile), which is slightly higher than the general population (4.9%) (H. Zhang et al., 2023).

Increased NT during the first trimester of pregnancy is an important ultrasound marker used to assess the risk of fetal abnormalities. While NT measurement is often associated with chromosomal abnormalities, it can also indicate a range of other fetal defects, some of which may be unrelated to genetic conditions. About 20% of fetuses with increased NT will have a chromosomal abnormality, with the incidence increasing with NT thickness: approximately 7% for NT between the 95th and 99th centile (3.5 mm), 20% for NT of 3.5–4.4 mm, 50% for NT of 5.5–6.4 mm, and 75% for NT of 8.5 mm or more (Bekker et al., 2006).

### **2.11.1 Cardiac Defects**

Major defects include aortic narrowing and septal or valvular anomalies, with a detection rate of 37.5%. Specialist fetal echocardiography, along with ductus venosus and tricuspid regurgitation assessments, can improve detection (De Domenico et al., 2011).

### **2.11.2 Sporadic Abnormalities**

Body stalk anomaly is a lethal condition seen in 84% of cases with increased NT. NT is also elevated in 40% of diaphragmatic hernia cases and 85% of chromosomally abnormal exomphalos cases. Megacystis is linked to increased NT in 75% of chromosomally abnormal and 30% of normal cases. While rare syndromes may be associated, a definitive link remains unclear (De Domenico et al., 2011).

### 2.11.3 Trisomy 21 (Down Syndrome)

Down syndrome results from an extra copy of chromosome 21 (trisomy 21) and is characterized by growth retardation, intellectual disability, and distinct craniofacial features, including upward-slanting eyes, epicanthal folds, a flattened facial profile, and small ears. About 75% of trisomy 21 fetuses have increased Nuchal Translucency NT thickness. Common complications include congenital heart defects, hypotonia, increased susceptibility to leukemia, immune dysfunction, thyroid disorders, and premature aging (Hyland & Kennedy, 2023). These features are visually depicted in Figure 2.13.

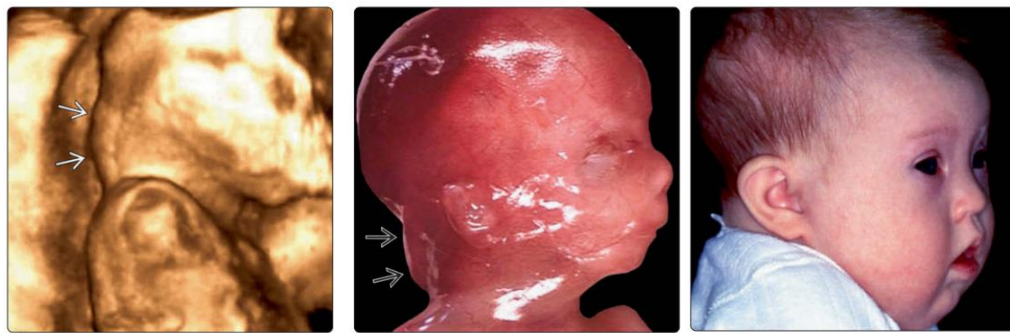


Figure (2.13) Child with Down Syndrome (Paula, 2021)

### 2.11.4 Trisomy 18 (Edwards Syndrome)

Edwards syndrome, or trisomy 18, is a severe genetic disorder caused by an extra copy of chromosome 18, leading to multisystem abnormalities and significant developmental impairment. It is characterized by congenital heart defects, growth restriction, clenched hands, prominent heels, low-set ears, micrognathia, renal anomalies, and skeletal malformations. The condition affects approximately 1 in 5,000 live births, with a high prenatal mortality rate—85% of affected pregnancies result in loss between 10 weeks of gestation and term. Among those born alive, most do not survive beyond two months, with only about 5% living beyond one year (Fiadjoe & Kennedy, 2023). These clinical characteristics are illustrated in Figure 2.14.



Figure (2.14) Child with Trisomy 18 (Paula, 2021)

#### 2.11.5 Trisomy 13 (Patau Syndrome)

Patau syndrome, or trisomy 13, is a severe genetic disorder caused by an extra copy of chromosome 13, leading to significant developmental and structural abnormalities. Key features include intellectual disability, holoprosencephaly, congenital heart defects, deafness, cleft lip and palate, and ocular anomalies such as microphthalmia, anophthalmia, and coloboma. The incidence is approximately 1 in 20,000 live births, with a high neonatal mortality rate—over 90% of affected infants die within the first month, and only about 5% survive beyond one year (NHS, 2023). These characteristics are illustrated in Figure 2.15.

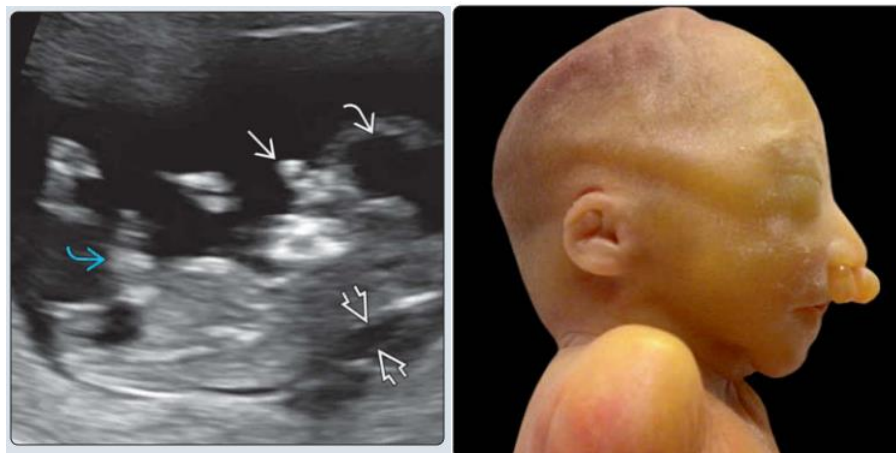


Figure (2.15) Child with Trisomy 13 (Paula, 2021)

### 2.11.6 Turner Syndrome

Turner syndrome is a genetic disorder affecting approximately 1 in 2,000 female births, caused by the presence of a single X chromosome (45, X) instead of two. It is the only monosomy compatible with life, though 98% of affected fetuses are spontaneously aborted. Surviving individuals exhibit distinct features, including short stature, gonadal dysgenesis (absence of functional ovaries), webbed neck, lymphedema, skeletal abnormalities, and a broad chest with widely spaced nipples. Approximately 55% have complete X monosomy, primarily due to paternal nondisjunction (80%), while others have structural X chromosome abnormalities or mosaicism due to mitotic errors (NHS, 2021). A visual representation of these features is provided in Figure 2.16 .



Figure (2.16) Patient with Turner syndrome (*Sadler, 2019*)

### 2.12 Risks Associated with an Increased NT

This table presents the association between NT thickness and the prevalence of major fetal abnormalities, fetal death, and chromosomal defects. The data indicate a strong positive correlation between increased NT thickness and higher rates of adverse outcomes. For instance, NT measurements  $>6.5$  mm are associated with a 46.2% risk of major fetal abnormalities, 19.0% risk of fetal death, and 64.5% risk of chromosomal defects. Conversely, NT thickness below the 95th centile shows significantly lower risks (1.6%, 1.3%, and 0.2%, respectively). These findings underscore the clinical importance of NT measurements in prenatal screening for identifying high-risk pregnancies and guiding further diagnostic evaluations (FMF). The summarized data are shown in Table 2.5.

Table (2.5) NT Thickness and the Frequency of Chromosomal Problems (AP, 2005)

| NT Thickness       | Major Fetal Abnormalities | Fetal Death | Chromosomal Defects |
|--------------------|---------------------------|-------------|---------------------|
| >6.5 mm            | 46.2%                     | 19.0%       | 64.5%               |
| 5.5–6.4 mm         | 24.2%                     | 10.1%       | 50.5%               |
| 4.5–5.4 mm         | 18.5%                     | 3.4%        | 33.3%               |
| 3.5–4.4 mm         | 10.0%                     | 2.7%        | 21.1%               |
| 95th–99th centiles | 2.5%                      | 1.3%        | 3.7%                |
| <95th centile      | 1.6%                      | 1.3%        | 0.2%                |

### 2.13 Historical View

Screening for DS has undergone significant advancements since its introduction in the seventies. Initially, screening relied exclusively on maternal age, with women aged 36 or older being offered invasive diagnostic procedures such as amniocentesis or CVS, despite their associated miscarriage risks (K. H. Nicolaides, 2004).

In the eighties, the development of the triple test, which measures serum markers (AFP, hCG, and estriol), allowed for a more effective second-trimester (K. H. Nicolaides, 2004) screening, achieving a 60% detection rate for DS.

A major advancement in the 1990s was the discovery that NT measurement at 11–14 weeks could serve as a non-invasive early screening tool for chromosomal abnormalities (K. H. Nicolaides, Azar, Byrne, et al., 1992). This innovation enabled first-trimester screening, aligning with the preference of pregnant women for earlier detection. The combination of NT measurement and maternal age provided a detection rate of approximately 75%, with a false positive rate of around 5%.

By the late 1990s, first-trimester screening was further optimized through the inclusion of maternal serum markers Pregnancy-Associated Plasma Protein-A (PAPP-A) and free  $\beta$ -hCG alongside NT thickness measurement, significantly improving detection rates to 85–90% (Nicolaides, 2004).

In the Netherlands, NT screening was first introduced by the Fetal Medicine Unit at the Academic Medical Centre in 1994. Initially performed in a research setting, it became available to any woman requesting the test by 1998. However,

nationwide access to the combined test (NT + serum markers) was not implemented until 2007 (Souk a et al., 2005).

In recent years, researchers have explored additional first-trimester ultrasound markers, such as NB absence, DVF abnormalities, and tricuspid regurgitation, to further enhance screening accuracy while reducing FPR. Moreover, studies indicate that increased NT thickness is not solely linked to DS; even in fetuses with a normal karyotype, an enlarged NT is associated with adverse pregnancy outcomes, including congenital heart defects and genetic syndromes(Bilardo et al., 2007) . A visual summary of the historical development of DS screening strategies is provided in Figure 2.17.

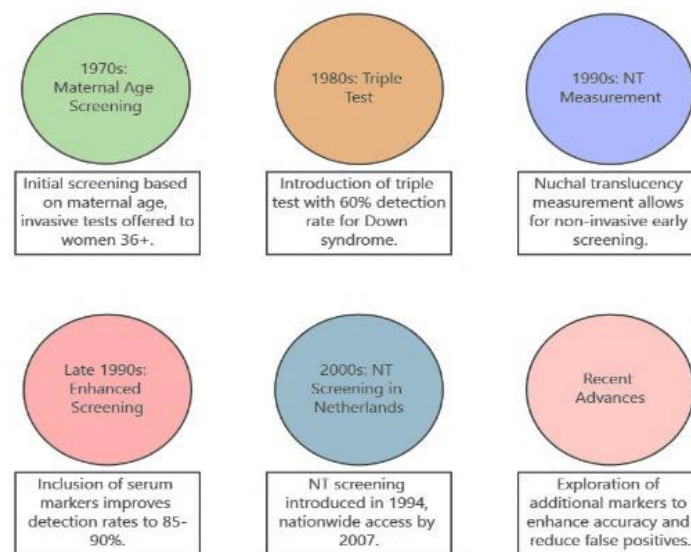


Figure (2.17) History for First-Trimester Screening

### 2.13.1 Pioneering Research on Nuchal Translucency Measurement and Fetal Aneuploidy Detection

Several landmark studies in the 1990s and early 2000s established the clinical significance of NT measurement in first-trimester screening for fetal aneuploidy:

- **Nicolaides et al. (1994)** conducted a study on 1,273 pregnant women with singleton pregnancies, performing first-trimester ultrasound at 10–14 weeks of gestation. They found that NT measurements  $\geq 3\text{mm}$  were associated with fetal aneuploidy, with a sensitivity of 85%, specificity of 95.9%, and a positive predictive value of 35.5%, while maintaining a false positive rate of 5% (K. H. Nicolaides et al., 1994).

- **Pandya et al. (1995)** studied 1,763 pregnant women and reported that NT measurements  $\geq 2.5\text{mm}$  had a sensitivity of 75%, specificity of 92%, and a false positive rate of 8% (Pandya et al., 1995).
- **Bewlerg et al. (1995)** analyzed 1,127 pregnancies and observed that an NT thickness  $\geq 3\text{mm}$  had a sensitivity of 40%, specificity of 94%, and a false positive rate of 6.1% (Bewley et al., 1995).
- **Taipale et al. (1997)** conducted a large-scale study involving 10,010 pregnant women and found that an NT measurement  $\geq 3\text{mm}$  had a sensitivity of 62%, specificity of 99%, a positive predictive value of 24%, and a negative predictive value of 100% (Taipale et al., 1997).
- **Economides et al. (1998)** studied 2,281 pregnant women and concluded that an NT measurement above the 99th percentile for gestational age had a sensitivity of 81%, specificity of 99%, and a false positive rate of only 0.4% (Economides et al., 1998).
- **Snijders et al. (1998)** focused on maternal age as a risk factor, evaluating 127 women with singleton pregnancies. NT measurements taken at 10–14 weeks of gestation yielded a sensitivity of 77% and a false positive rate of 4.4% (Snijders et al., 1998b).
- **Hafner et al. (1998)** conducted research on 4,233 pregnancies, demonstrating that an NT measurement  $> 2.5\text{mm}$  had a sensitivity of 65%, specificity of 98.5%, a positive predictive value of 14.8%, a negative predictive value of 99.9%, and a false positive rate of 1.5% (Hafner et al., 1998).
- **Schwarzler et al. (1999)** studied 4,523 pregnancies in an unselected population, examining first-trimester screening for aneuploidy and congenital heart defects through maternal age and NT measurement. Their findings indicated a sensitivity of 76% with a false positive rate of 4.7% (Schwärzler et al., 1999).

These studies collectively laid the foundation for the widespread clinical adoption of NT measurement as a key parameter in early pregnancy screening. They also influenced the development of modern first-trimester screening protocols that incorporate maternal serum markers and other ultrasound features to improve screening accuracy.

## 2.14 Method of Measurement of Nuchal Translucency

Nuchal Translucency (NT) can be assessed using either transabdominal or transvaginal ultrasound techniques, and it is crucial that the scan be performed by a qualified radiologist. The ultrasound system should be of high quality, employing a 5 MHz transabdominal probe or an 8 MHz transvaginal probe, and must be capable of generating video loops with precision to one decimal place. Typically, the NT scan lasts 10–20 minutes, and attempts to obtain an accurate measurement should continue for at least 20 minutes before discontinuing the procedure (Braithwaite & Economides, 1995).

- During the scan, the fetus must be visualized in a mid-sagittal plane, with the thalamus (t), medulla oblongata (mo), and pons (p) clearly identifiable. Proper fetal positioning is essential especially ensuring a neutral neck posture—to maintain measurement accuracy.
- The Fetal Medicine Foundation (FMF) criteria provide the standard technique for NT assessment: a hyperextended neck can increase the NT measurement by approximately 0.6 mm, while a flexed neck can decrease it by about 0.4 mm (Morris et al., 1999, Whitlow et al., 1998).

These principles are visually demonstrated in Figures 2.18–2.21, where Figure 2.18 shows the correct NT measurement that meets FMF criteria with precise “on-to-on” caliper placement (Raksha et al., 2019). In contrast, Figure 2.19 illustrates inadequate magnification for accurate measurement (Hyett, 2018), Figure 2.20 depicts a hyperextended neck unsuitable for NT evaluation (Hyett, 2018), and Figure 2.21 presents an overly flexed neck, also inappropriate for accurate measurement (Hyett, 2018).

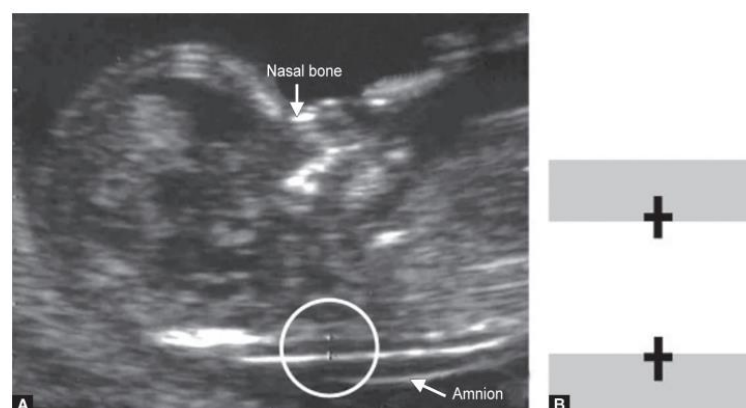


Figure 2.18 Nuchal Translucency Measurement that Meets Fetal Medicinal Foundation Criteria (t=thalamus, mo=medulla oblongata, p=pons). The Correct Caliper Placement is “on-to-on” (Raksha et al., 2019)



Figure 2.19 Magnification is too Small for Accurate Measurement of NT



Figure 2.20 Neck is Hyperextended, Not Correct for Measuring NT (Hyett, 2018)



Figure 2.21 Neck is too Much Flexed, Not Suitable for Measuring NT (Hyett, 2018)

- The gestational age should range from 10 to 14 weeks, with the fetal CRL measuring between 38 and 84 mm. Recent updates in fetal medicine indicate that the optimal gestational age is between 11 and 13 weeks plus six days, with the fetal CRL corresponding to 45 to 84 mm (Braithwaite et al., 1996, Whitlow & Economides, 1998).
- The fetal image must occupy 75% of the display, focusing solely on the fetal head and upper thorax (Edwards et al., 2003).
- To ensure precision, the image should be magnified as much as possible, allowing for only a 0.1 mm variation in measurement with slight movements of the calipers. The calipers should be accurately positioned on the inner edge of the nuchal fold and aligned perpendicularly to the long axis of the fetal body, utilizing a cross-shaped (+) caliper for "on to on" placement (Whitlow & Economides, 1998). This technique is illustrated in Figure 2.22.

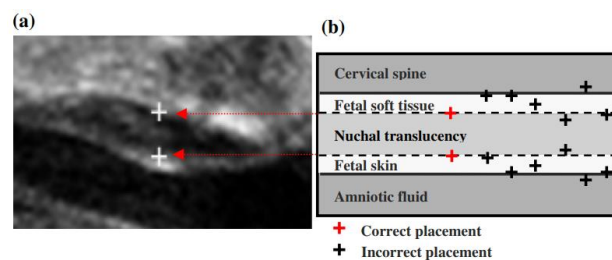


Figure (2.22) Correct Placement of Calipers (Y.-B. Lee et al., 2007)

- It is essential to differentiate between the fetal skin and the amnion, as both are thin membranes. This distinction can be made when there is spontaneous fetal movement, causing the fetus to shift away from the amniotic membranes. Additionally, this differentiation can occur when the mother coughs or when the abdomen is gently tapped (K. H. Nicolaides, Azar, Byrne, et al., 1992).
- The maximum thickness of the subcutaneous translucency between the skin and the soft tissue covering the cervical spine is measured. Multiple measurements are taken, and the highest value is recorded for greater accuracy (K. H. Nicolaides, et al., 1992). This measurement approach is demonstrated in Figure 2.23.

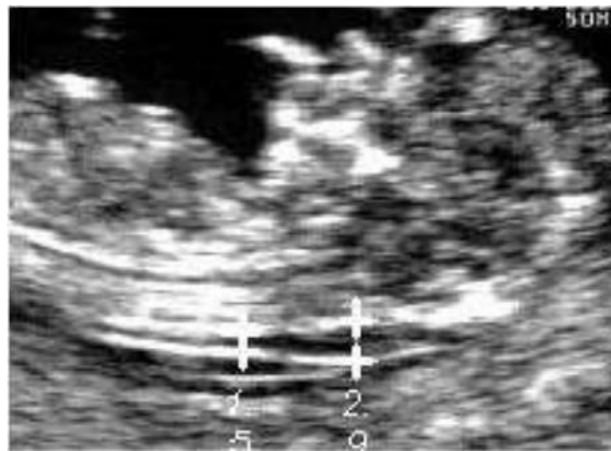


Figure (2.23) Maximum Thickness (Sebire et al., 1997)

In 5% of instances, the umbilical cord encircles the fetal neck, leading to inaccurately elevated NT readings. When this occurs, the NT measurements taken above and below the cord will vary; therefore, it is advisable to calculate the average of these two measurements. This scenario is illustrated in Figure 2.24.



Figure (2.24) Umbilical Cord Surrounds the Fetal Neck that Disrupts NT Measurement (Sebire et al., 1997)

## 2.15 Variability in NT Cutoff Values

Despite the widespread use of NT screening, there is no global consensus on the optimal cutoff value for defining increased NT thickness. Different countries and institutions use varying thresholds, such as 2.5 mm, 3.0 mm, 3.5 mm, or percentiles like the 95th or 99th percentile of the NT distribution (Su et al., 2021). This inconsistency has led to challenges in standardizing prenatal screening protocols and interpreting results across different populations. A recent study investigated the predictive performance of various NT cutoff values in detecting chromosomal abnormalities in singleton pregnancies. The study analyzed a sample of 3,223 invasive cases, categorized into five NT thickness groups: 2.5 mm, 95th percentile, 99th percentile, 3.0 mm, and 3.5 mm, alongside a control group of 2,301 samples with normal NT.

The findings highlighted that the NT 95th percentile cutoff was the most effective for first-trimester screening, offering a reliable balance between sensitivity and specificity. This threshold reduces false positives while ensuring the detection of true positives, making it an optimal criterion for guiding invasive diagnostic procedures.

These results align with prior research, such as Nicolaides et al (K. H. Nicolaides, et al., 1992), who initially established NT as a marker for chromosomal defects, and Kagan et al (Kagan et al., 2006), who affirmed its predictive utility for adverse pregnancy outcomes. However, the lack of standardized NT cutoff values remains a challenge, as thresholds often vary across regions based on local guidelines and population-specific characteristics.

Conducted (Acácio et al., 2001) a diagnostic validation study to determine the optimal fixed cutoff point for NT thickness as a screening marker for chromosomal abnormalities, particularly trisomy 21, in a South American population. The study analyzed 230 pregnant women (mean age: 35.8 years) at 10–14 weeks of gestation, measuring NT via ultrasound and confirming fetal karyotype. The mean NT thickness was 2.18 mm, with an increased NT correlating with a higher prevalence of aneuploidy (10% overall). Using Receiver Operating Characteristic (ROC) curve analysis, the study identified 2.5 mm as the best cutoff for NT screening, achieving a sensitivity of 69.5% and a specificity of

87.4% for all chromosomal abnormalities, and 75% sensitivity with 84.9% specificity for trisomy 21. The study supports NT measurement as a reliable ultrasound marker for aneuploidy screening, with higher accuracy for trisomy 21, but also highlights the need for risk-adjusted screening models integrating maternal age and gestational age for improved diagnostic precision.

Conducted (H. Zhang et al., 2023) a retrospective study at a Chinese tertiary medical center to assess the relationship between increased NT thickness and chromosomal abnormalities or structural defects in fetuses. The study included 264 fetuses with NT  $\geq$  95th percentile at 11–14 weeks of gestation, with a median NT measurement of 2.41 mm. Among these cases, 132 underwent invasive prenatal diagnosis, revealing 16 chromosomal abnormalities (6 cases of trisomy 21, 4 cases of trisomy 18, and 4 cases of Copy Number Variations (CNVs)). The study found a significant correlation between NT  $\geq$  2.5 mm and chromosomal abnormalities, with incidence rates of 8.8% for chromosomal abnormalities and 28.9% for structural defects compared to 1.3% and 6%, respectively, in fetuses with NT < 2.5 mm. The most common structural defects associated with increased NT included hydrops (6.4%), cardiac anomalies (3%), and urinary anomalies (2.7%). The findings underscore the importance of NT thickness as a predictive marker for both chromosomal and structural abnormalities, supporting the use of the 95th percentile as a more sensitive cutoff value for screening compared to a fixed threshold of 2.5 mm.

## **2.16 Case Studies: Increased Fetal Nuchal Translucency**

Increased NT is a key marker in early prenatal screening and may indicate chromosomal or structural anomalies. The following case studies highlight varied clinical outcomes associated with elevated NT, emphasizing its importance in guiding further diagnostic assessment.

### **2.16.1 Case Study 1: Increased NT and Esophageal Atresia**

This case study presents the association between increased fetal NT and the subsequent diagnosis of esophageal atresia with Tracheoesophageal Fistula (TEF).

A 32-year-old woman underwent first-trimester screening at 12 weeks, during which an NT measurement of 3.0 mm was recorded, indicating an elevated risk for fetal trisomy. Subsequent tests, including karyotyping and fetal echocardiography, showed normal results.

At 31 weeks, the woman exhibited an increased symphysis-fundal height and was diagnosed with polyhydramnios (AFI of 33 cm). By 32 weeks, persistent stomach emptiness suggested potential esophageal atresia, which was later confirmed at delivery. The infant underwent surgical repair of the esophagus within 24 hours after birth.

This case highlights the importance of increased NT as a marker for potential fetal abnormalities and emphasizes the need for thorough prenatal evaluations and early recognition of conditions like esophageal atresia to improve neonatal outcomes through timely interventions (Brown & Nicolaides, 2000).

#### **2.16.2 Case Study 2: Increased NT with Normal Outcome**

In this case study, a 26-year-old first-time mother (primigravida) was referred for an ultrasound at 12 weeks of gestation, which revealed an increased NT measurement above the 99th percentile. Despite the concerning NT measurement, genetic testing using Non-Invasive Prenatal Testing (NIPD) indicated a normal karyotype, and there was no family history of chromosomal abnormalities or congenital malformations reported.

The biochemical screening results were also reassuring, showing a low risk for chromosomal anomalies (PAPP-A=0.92 MoM and fb-HCG=1.13 MoM). After discussing the findings and potential implications in detail, the parents chose to proceed with the pregnancy, despite the stress caused by the elevated NT.

Throughout the pregnancy, the mother underwent regular ultrasound assessments every four weeks. At 19 weeks and 3 days, the NT measured 7.1 mm, but further evaluations, including detailed sonography and fetal echocardiography, found no structural malformations, and growth of the fetus was normal.

The pregnancy culminated in delivery at 40 weeks, resulting in a healthy baby girl weighing 40-30 grams, with an APGAR score of 8/9. At a nine-month follow-up, the child exhibited normal physical and neurological development (Jasovic-Siveska, et al., 2018). Figure 2.25 illustrates a detailed ultrasound image.

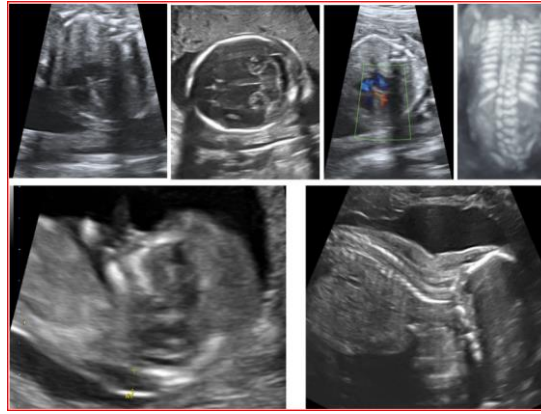


Figure (2.25) Detailed Ultrasound Image (Jasovic-Siveska, et al., 2018)

### 2.16.3 Case Study 3: NT Screening Outcomes in a Prospective Study

This prospective study was conducted at Coimbatore Medical College Hospital from August 2013 to August 2014 and included 50 pregnant women between 10–14 weeks of gestation with reliable LMP and singleton intrauterine pregnancies. Women with unreliable LMP or multiple pregnancies were excluded. Participants were randomly selected and counseled about trimester-wise screening. First-trimester ultrasound assessed fetal viability, multiple gestations, and measured NT, while second-trimester scans at 20–22 weeks evaluated fetal growth and structural anomalies. Clinical examinations were also performed. Of the women, 50% were aged 21–25, 56% were primigravida, and 52% were scanned between 12.1–13 weeks. Most had no past obstetric issues (78%) or present risk factors (78%). NT measurements showed 96% had values below the 95th percentile; only 4% had increased NT, with one fetus diagnosed with a cardiac defect (ASD) and the other found normal. Regarding outcomes, 66% had vaginal deliveries, and 34% had cesarean sections; 58% of newborns weighed 2.6–3.0 kg, and 8% required NICU admission due to IUGR, MAS, preterm birth, or CHD. Diagnostic metrics for NT were: sensitivity 50%, specificity 97.92%, PPV 50%, and NPV 97.92%, with a statistically significant correlation between increased NT and fetal abnormalities ( $p = 0.017$ ). The study concludes that NT is a valuable, non-invasive, and effective first-trimester screening tool for early detection of chromosomal and structural anomalies, though larger studies are needed to strengthen screening guidelines (R. P. Devi, 2015).

## 2.17 Trends in Research Publications on Fetal Health Monitoring

Figure 2.26 illustrates the number of publications per year related to fetal health monitoring using ultrasound images, as sourced from ScienceDirect. From 2011 to 2021, there is a clear and significant upward trend in research interest, with the number of publications increasing steadily over the decade. This growth reflects the growing importance of fetal health monitoring in prenatal care and the increasing adoption of advanced technologies, such as ML and DL, in this field. The sharp rise in publications, particularly in recent years, underscores the rapid advancements in imaging techniques, computational methods, and their applications in early detection of fetal abnormalities. This trend highlights the critical need for continued research and innovation to improve the accuracy, efficiency, and accessibility of prenatal screening tools.

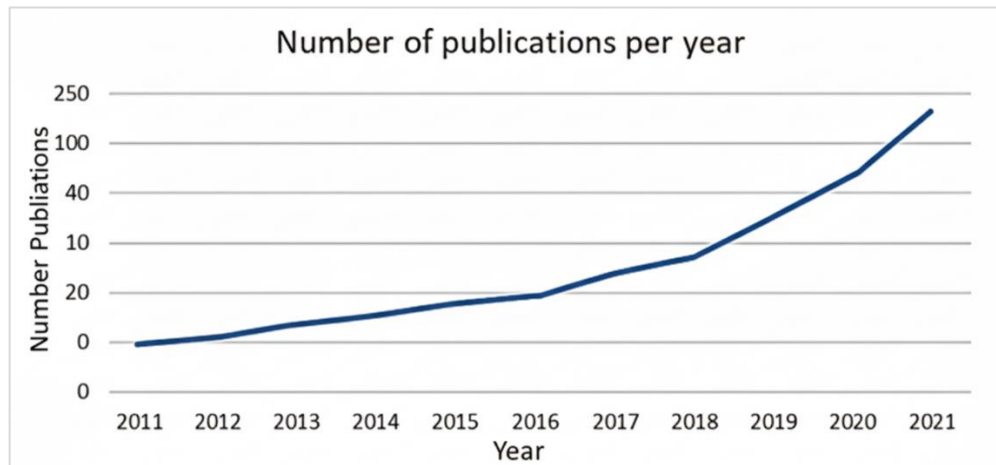


Figure (2.26) Number of Publications about Learning Methods Used in Fetal Health Monitoring, within Last Decade by Year. Numbers are Obtained from ScienceDirect

## 2.18 Manual Methods for Measuring Nuchal Translucency

The measurement of NT involves identifying the subcutaneous fluid-filled space at the back of the fetal neck in a mid-sagittal ultrasound view. This is typically performed between 11 and 13+6 weeks of gestation when the CRL ranges from 45 to 84 mm (K. H. Nicolaides, et al., 1992). The process requires the placement of calipers on the inner borders of the NT region, ensuring that the fetus is in a neutral position and the amnion is distinguished from the NT line, this procedure is illustrated in Figure 2.27.

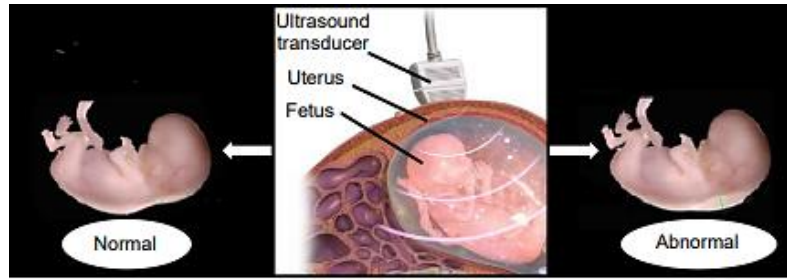


Figure (2.27) Nuchal Translucency Detection by Ultrasound Scan (Verma et al., 2022)

### 2.18.1 Limitations of Traditional Methods

- **Operator Dependency:** The accuracy of manual NT measurement depends heavily on the skill and experience of the sonographer (Snijders et al., 1996). Variations in technique can lead to inter- and intra-observer inconsistencies.
- **Image Quality:** Factors such as poor resolution, artifacts, and fetal positioning can affect the ability to accurately visualize and measure NT (Souka et al., 2001).
- **Error-Prone Process:** Studies indicate that manual measurements can suffer from subjectivity and visual constraints, resulting in variability in results (Müller et al., 2004).

## 2.19 Semi-Automated Methods for Measuring Nuchal Translucency

There are a lot of various studies in the literature that specifically address the measurement of NT; this is one of the first works that attempted to automate the procedure. Their procedure is semi-automatic because the user has to identify by hand the membranes that surround the translucency (Bernardino et al., 1998). The authors in the study of (Y.-B. Lee et al., 2007) proposed a semi-automatic method for measuring fetal NT thickness in ultrasound images, addressing challenges related to weak edges, noise, and operator variability. The method integrates Coherence-Enhancing Diffusion (CED) filtering for noise reduction and border enhancement, followed by Dynamic Programming (DP) for optimal NT border detection. A cost function, combining intensity, gradient, and continuity constraints, guides the DP-based segmentation process. Quantitative evaluation demonstrated a high correlation (0.99) between semi-automatic and manual measurements, outperforming traditional edge-detection techniques.

The authors in the study (Wee et al., 2010) developed an automated system to detect and measure NT in fetal ultrasound images, aiming to reduce observer variability in prenatal screenings for chromosomal abnormalities. The methodology involved using a multilayer perceptron neural network to identify the NT region, achieving a detection accuracy of 93.33%. Subsequently, the Bidirectional Iteration Forward Propagation algorithm was employed to measure NT thickness, resulting in a high correlation (0.98) between automated and manual measurements. The study utilized a dataset of 150 ultrasound images, acquired with a KNOTRON Sigma 330 Expert ultrasound machine equipped with a 3.5 MHz convex transducer.

The study in (Nie et al., 2017) presented an automatic method for measuring fetal NT using Three-Dimensional (3D) ultrasound data, addressing the limitations of manual Two-Dimensional (2D) NT measurement in prenatal screening. Traditional NT thickness measurement is performed manually in the mid-sagittal plane, leading to inter-operator variability and potential measurement errors. The proposed approach utilizes dynamic programming and a novel cost function to detect NT thickness and area in 2D, while also extending the analysis to 3D NT volume measurement, providing a more comprehensive marker for fetal abnormalities. Experimental validation on 68 clinical 3D ultrasound datasets demonstrates that this method achieves higher accuracy and robustness compared to previous automated techniques.

Table 2.6 presents the statistical results of the NT thickness measurement, comparing manual measurement with other methods (2007 Lee, 2010 Wee, and the proposed method).

Table (2.6) The Statistic Result of The NT Thickness

| Method             | <i>M</i> | <i>σ</i> | D NT Len | D border |
|--------------------|----------|----------|----------|----------|
| Manual measurement | 1.5185   | 0.3374   | -        | -        |
| 2007 Lee           | 1.5067   | 0.4033   | 0.2833   | 0.2893   |
| 2010 Wee           | 0.7075   | 0.0757   | 0.8119   | 0.8111   |
| Proposed method    | 1.5076   | 0.4035   | 0.2825   | 0.2678   |

This study present (Catanzariti et al., 2009) proposed a semi-automated method that detects NT borders after manual region selection, using cost function minimization to improve accuracy and efficiency. This approach reduces measurement time, enhances reproducibility, and minimizes operator errors, making NT assessment more reliable and standardized in clinical practice. Integrating such semi-automated methods into prenatal screening lays the foundation for fully automated ML-based approaches, further improving diagnostic precision and early detection.

This study present (Sciortino et al., 2017) automatic method for detecting and measuring NT in ultrasound images using wavelet-based multi-resolution analysis. The approach utilizes the *à trous* algorithm for wavelet decomposition, thresholding for NT region identification, and mathematical morphology for refining boundaries and measuring NT thickness. Tested on 382 ultrasound frames from 12 patients, the method achieved a 99.95% true positive detection rate and an average measurement error of 1.28 pixels (0.128 mm), demonstrating high accuracy and reliability.

Table (2.7) Summarizing the References, Technologies, Tasks, And Performance Metrics for Different NT Measurement Methods

| References                | Technologies                                                   | Tasks                                                                                | Performances                                         |
|---------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------|------------------------------------------------------|
| <b>Lee et al.</b>         | Coherence-enhancing diffusion filter, Dynamic programming      | Automatic measurement of NT with manual ROI.                                         | Correlation coefficient: 0.99                        |
| <b>Catanzariti et al.</b> | Dynamic programming                                            | Automatic measurement of NT with manual ROI                                          | No quantitative analysis; performs better than Lee.  |
| <b>Siqing et al.</b>      | Dynamic programming, Hessian plate filter, Deep belief network | Automatic identification of the mid-sagittal plane, detection, and measurement of NT | $\sigma = 0.40$ , $dNTlen = 0.28$ , $dborder = 0.27$ |
| <b>Sciortino et al.</b>   | Wavelet, Multi-resolution analysis                             | Automatic identification of the mid-sagittal plane, detection, and measurement of NT | Sensitivity: 99.95%                                  |

This study proposed (Catanzariti, 2009) a semi-automated method for measuring NT in ultrasound images, reducing the time required for measurement from 15–35 minutes to a shorter duration while improving accuracy and reducing operator dependency. The system employs a dynamic programming algorithm to detect NT boundaries based on gradients, continuity, and positional constraints, calculating the maximum vertical distance between boundaries within a manually selected ROI.

The study by (Y.-H. Deng et al., 2008) integrates morphological filtering, Gradient Vector Flow (GVF) snake, and dynamic programming to enhance contour extraction and parameter estimation. Initially, a morphological filter is applied to reduce speckle noise, followed by manual selection of the ROI to isolate the NT region. Contour extraction is performed through a threshold-based binary image transformation and GVF snake, which refines the boundary by capturing concavities. Dynamic programming is then employed to optimize contour selection, ensuring accurate delineation of the NT region. NT thickness is measured as the maximum distance between contour points, while NT area is calculated by pixel enumeration within the extracted boundary. The method demonstrates superior performance on both real and simulated ultrasound images, outperforming traditional noise reduction techniques such as Speckle Reducing Anisotropic Diffusion (SRAD). The findings suggest that NT area may serve as a more sensitive diagnostic marker than NT thickness, highlighting the potential of automated NT measurement systems in prenatal screening for chromosomal abnormalities such as trisomy 21, trisomy 18, and Turner syndrome.

This study presents (Lakshmi et al., 2018) an automated algorithm for measuring both NT thickness and Front Maxillary Facial (FMF) angle in ultrasound images to enhance first-trimester screening for trisomy 21. The method applies image processing techniques, including OTSU thresholding, contour extraction, and line fitting, to segment the NT region and compute the FMF angle. The algorithm follows a structured pipeline: (1) ROI segmentation, where the fetal head and neck regions are isolated; (2) image thresholding, using OTSU's method to differentiate NT and FMF structures; (3) contour extraction, detecting NT

boundaries and FMF landmarks; (4) feature measurement, calculating the NT thickness as the maximum vertical distance between contours and the FMF angle using slope-based line fitting; and (5) risk classification, identifying fetuses at risk for DS based on  $NT > 2.5$  mm and  $FMF > 85^\circ$ . Tested on clinical ultrasound images, the method achieved 90% accuracy for NT measurement and reliably extracted FMF angles.

This study proposed (Sonia & Shanthi, 2024) a semi-automated method for measuring NT thickness, a critical marker for DS detection in the first trimester of pregnancy. The method employs morphological operations with Otsu thresholding to segment the NT region from ultrasound images, following pre-processing with a Lee filter and manual ROI selection. Tested on a dataset of 80 fetal ultrasound images (50 normal, 30 abnormal), the approach demonstrated that normal fetuses had an NT thickness of  $1.99 \pm 0.62$  mm, whereas abnormal fetuses exhibited an NT thickness of  $4.10 \pm 0.90$  mm. The study highlights the effectiveness of image processing techniques in standardizing NT measurement, reducing inter-operator variability, and supporting early prenatal screening for chromosomal abnormalities.

The authors in (Abele et al., 2010) evaluated the intra- and interoperator reliability of manual and semi-automated NT measurements among sonographers with varying levels of experience, using SonoNT software. The study, conducted on 70 high-quality NT ultrasound images, involved four operators (two experts, one general consultant, and one medical student) performing manual and semi-automated NT measurements. The results demonstrated that semi-automated methods significantly reduced measurement variability, particularly for non-expert operators, aligning their reliability with expert-level performance. Intraoperator reliability improved with higher Intraclass Correlation Coefficients (ICCs) in semi-automated measurements, while interoperator variability was minimized, especially when using the inner-inner caliper placement method.

The authors of (Moratalla et al., 2010) developed a semi-automated system to measure NT thickness, comparing its performance with the traditional manual method. The dataset consisted of 48 ultrasound images from 12 fetuses at 11–13

weeks of gestation. Twenty sonographers, each with a certificate of competence in first-trimester scanning, measured NT twice using both methods. The semi-automated system required the operator to place an adjustable box over the fetal neck, after which the system identified the maximum vertical distance between the nuchal membrane and the soft tissue overlying the cervical spine. Results showed that the semi-automated method significantly reduced measurement variability, with a between-operator Standard Deviation (SD) of 0.0149 mm compared to 0.109 mm in the manual method. Similarly, the within-operator SD was reduced to 0.05 mm from 0.126 mm. ICCs were 0.98 for the semi-automated method and 0.85 for the manual method, demonstrating superior consistency.

The authors of (P. W. Chen et al., 2012) evaluated the impact of image settings on NT measurements using a semi-automated system, aiming to assess whether changes in pre- and post-processing parameters affect measurement accuracy. The dataset included 66 archived NT images for post-processing analysis and 91 live scans of the Posterior Tibial Artery (PTA) to simulate NT conditions during pre-processing. Measurements were obtained using the SonoNT™ function on a Voluson E8 ultrasound system. Results showed that most setting changes led to statistically significant differences in NT measurements, although the absolute differences were small ( $< 0.1$  mm) and likely clinically negligible. Notably, high contrast settings decreased NT measurements by 0.139 mm, while turning off Tissue Harmonic Imaging (THI) increased measurements by 0.120 mm.

This study (Grangé et al., 2011) investigated the feasibility and reproducibility of a Semi-Automated Adjusted Measurement (SAAM) technique for NT thickness during the first trimester of pregnancy. The dataset included 160 archived NT images, assessed by three operators (two experts and one less experienced) using an ultrasound system with SAAM functionality. The method involved manually placing a frame around the NT area, allowing the software to automatically detect NT borders and provide multiple measurements (maximum, minimum, mean, and median NT thickness). Results demonstrated high intra- and interobserver reproducibility, with ICCs ranging from 0.76 to 0.95 across different NT measurements, and the highest reproducibility observed for maximum NT.

The authors of (Lai et al., 2011) explored the use of 3D ultrasound imaging to enhance the accuracy and consistency of NT measurements for early detection of trisomy 21. The proposed method involved reconstructing NT from multiple 2D ultrasound images using an open-source Visualization ToolKit (VTK), enabling interactive 3D exploration of NT structure. A virtual slider cutting plane allowed precise identification of NT boundaries, while measurements were calculated using the Euclidean distance formula in 3D space. The dataset consisted of 150 DICOM ultrasound frames acquired with a 3.5 MHz abdominal probe. Results demonstrated improved accuracy and reduced inter- and intra-observer variability compared to traditional 2D methods, with the 3D approach minimizing errors related to manual caliper placement and image positioning.

## **2.20 Machine Learning Methods for Measuring Nuchal Translucency**

This study (Chinnaiyan & Alex, 2021) explored the use of ML techniques to improve the early diagnosis and prediction of fetal abnormalities during the first trimester of pregnancy. The proposed framework analyzed ultrasound fetal images through a four-step process: segmentation, image enhancement, feature extraction, and classification. Various ML models, including SVM, Random Forest, and Ensemble Learning, were evaluated for their performance in detecting abnormalities such as anencephaly, spina bifida, and congenital heart defects. The study emphasized the potential of ML to enhance fetal health assessments by providing a reliable, automated support system that reduces dependency on expert sonographers.

Traditional first-trimester screening for trisomy 21 relies on maternal age and NT, with limited accuracy. To address this (Sun et al., 2021) proposed an individualized nomogram integrating fetal NT and facial profile markers analyzed via ML. Using a retrospective dataset of 624 fetal ultrasound images (302 trisomy-21, 322 euploid), the proposed nomogram incorporates eight key markers identified via LASSO regression. These include fetal NT, the prenasal-thickness-to-nasal-bone-length ratio, the frontomaxillary facial angle, and others (facial profile line, frontonasal facial angle, mandibulomaxillary facial angle, maxilla-nasion-mandible angle and d2 distance between the anterior edge of the prefrontal

skin and the mandibulomaxillary line). Results indicated strong performance in both the training set (AUC = 0.983) and validation set (AUC = 0.979), outperforming traditional NT-based screening.

This study proposes (V. Devi & Ramachandran, 2021) an advanced ML approach for automating DS detection by enhancing the accuracy of NT measurement in prenatal ultrasound imaging. Building on the modified AdaBoost algorithm integrated with Backpropagation Neural Networks (BPNN), we aim to improve classification performance and reduce operator dependency. A previous study by Vincy Devi and Rajesh (2021) demonstrated the effectiveness of this approach, achieving 98.67% accuracy on a dataset of 82 images, outperforming traditional SVM and LBP classifiers.

The study (J. Deng et al., 2025) utilized a dataset comprising 5,519 singleton pregnancies from the University Malaya Medical Centre, Malaysia, to develop a nuchal thickness reference chart and evaluate its predictive value for Small-for-Gestational-Age (SGA) infants. Two models were employed: logistic regression, which achieved an AUC of 0.75 in the Malaysian cohort using five variables (abdominal circumference, femur length, nuchal thickness, maternal age, and ultrasound-confirmed gestational age), and SVM, which demonstrated superior performance in the Singapore cohort with an AUC of 0.81 when incorporating all available variables.

Another study proposed (Ramanathan et al., 2018) applied ML techniques to predict the probability of DS using first-trimester prenatal screening data. The dataset comprised 442 prenatal screening reports, with features including Nuchal Translucency,  $\beta$ -hCG, PAPP-A, CRL, maternal age, smoking habits, diabetes, NB presence, and previous trisomy history. Data preprocessing involved handling missing values, normalization, and addressing class imbalance using the Adaptive Synthetic Sampling Approach (ADASYN). Several models were evaluated, including Naive Bayes, Random Forest, and Artificial Neural Networks (ANN). Among these, Naive Bayes achieved the highest accuracy at 92.63%, outperforming Random Forest and ANN, which achieved 73.5% and 64%, respectively. Table 2.8 summary the results of the NT thickness measurement.

Table (2.8) Supervised Learning Technique

| Supervised Learning Technique | Accuracy |
|-------------------------------|----------|
| Naive Bayes                   | 92.63    |
| Random Forest                 | 73.5     |
| Artificial Neural Network     | 64.0     |

In this study (Li et al., 2019) we propose a Cascaded Voting Isolation Forests and Logistic Regression (CVIFLR) framework, specifically designed to address the challenges of highly imbalanced and feature-correlated datasets in the context of DS prediction. The framework integrates a three-stage cascaded approach, comprising pre-judgment using Isolation Forests (IF), model ensemble via voting strategy, and final judgment through Logistic Regression (LR), to effectively identify DS cases with high accuracy. Experimental results demonstrate that the CVIFLR framework achieves an Area Under the Receiver Operating Characteristic (AUROC) of 0.99, outperforming state-of-the-art methods such as BalanceBagging, BalanceCascade, and SMOTE-based techniques. Furthermore, the optimal combination of input features—PAPP-A MoM,  $\beta$ -hCG MoM, uE3 MoM, and MA—was identified, yielding the lowest FPR across varying True Positive Rates (TPR).

The proposed (Park et al., 2013) fully automatic NT measurement system demonstrates a significant advancement over existing methods by integrating Hierarchical Detection Network (HDN) for fetal head detection, Dijkstra's shortest path algorithm for NT region approximation, and Graph Cut segmentation for precise boundary refinement. The results, validated on 196 DICOM ultrasound scans, highlight its high accuracy, with an NT measurement error as low as 0.24 mm. Compared to prior approaches such as dynamic programming, mean shift analysis, wavelet analysis, and Chan-Vese segmentation this model achieves greater reliability and robustness, minimizing manual input while ensuring consistent and reproducible NT assessments.

Figure 2.28 illustrates the segmentation process of the NT region. (a) The pose of the detected NT region is employed to delineate the segmentation area, ensuring precise localization of the ROI. (b) Initial edges of the NT region are extracted by applying Dijkstra's shortest path algorithm, where the inverse gradient magnitude serves as the cost function to identify optimal boundaries. (c) The detected NT edges are utilized to define foreground (highlighted in green) and background (highlighted in red) seed points, which are critical for subsequent segmentation. (d) The final segmentation result is achieved by executing a Graph Cut segmentation algorithm, which optimally partitions the image based on the defined seeds and edge information.

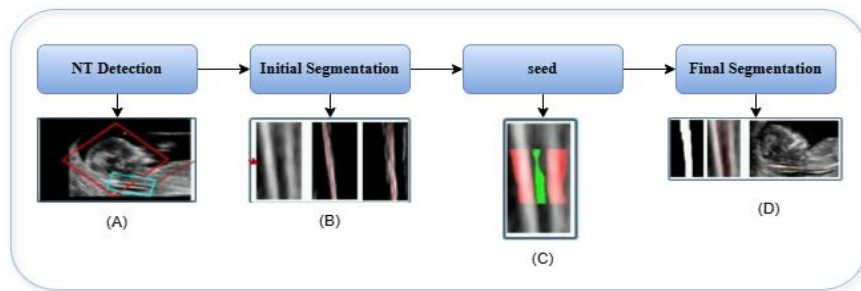


Figure (2.28) Algorithmic Workflow for NT Segmentation Utilizing NT Region Detection Results (Park et al., 2013)

The work in (Y. Deng et al., 2012) employs three discriminative classifiers trained with Gaussian pyramids to identify the NT region, fetal head, and body while ensuring spatial consistency through a spatial model. SVM were applied for feature classification, leveraging the Histogram of Oriented Gradients (HOG) to counteract speckle noise in ultrasound images. To further refine detection, the model integrates the generalized distance transform and dynamic programming, optimizing the identification of anatomical structures.

The model's performance was evaluated on 690 clinical NT ultrasound images. Results showed that with 50% of the data used for training, the NT detection accuracy reached 55.9%, while the detection accuracy of the fetal head and body achieved 89.4% and 100%, respectively. The "OR" node, designed to resolve fetal orientation, correctly selected the fetus's direction with up to 94.0% accuracy when using 10% of the training data. Notably, the spatial model improved NT detection by approximately 5.68% compared to using the SVM classifier alone.

This study (Sciortino et al., 2017) proposed a fully automated, unsupervised method for detecting and measuring NT in first-trimester ultrasound images, addressing the limitations of manual and semi-automated approaches. Their method leverages wavelet analysis and feedforward neural networks to identify the Mid-Sagittal Section (MSS), ensuring accurate NT segmentation and measurement. Initially, a symmetry transform is applied to detect the correct MSS, followed by À Trous wavelet transform for multi-scale edge enhancement, allowing precise delineation of NT boundaries. A two-stage feedforward neural network is then employed: the first network filters out incorrect sagittal planes, while the second network classifies facial components, refining NT localization. Tested on real clinical ultrasound images, the method achieved 97.4% accuracy, with an average measurement error of just 0.3 mm, demonstrating its robustness against image noise and artifacts.

This study (Anzalone et al., 2013) introduces an automated system for measuring NT thickness from ultrasound video streams. By integrating sagittal view selection, ROI detection, and NT measurement modules, the system eliminates manual intervention, improving accuracy and repeatability. The system employs advanced algorithms, including anisotropic filtering for noise reduction, template matching with K-means clustering for ROI detection, and dynamic programming-based border detection for precise NT thickness measurement. Experimental results demonstrate robust performance, with 93% accuracy in sagittal view selection and 61% in ROI detection, validated against expert annotations.

The author (Nirmala & Palanisamy, 2009) introduced a computer-aided system to measure NT thickness in first-trimester ultrasound images, aiming to enhance the detection of chromosomal abnormalities such as DS. The methodology involved preprocessing steps like speckle noise reduction using a median filter and ROI selection, followed by segmentation techniques including mean shift analysis and Canny edge detection to delineate the NT region. Subsequent blob analysis facilitated precise NT thickness measurement. The system was evaluated on 50 ultrasound images, revealing that at 14 weeks of gestation, the average NT thickness in normal fetuses was  $1.87 \pm 0.25$  mm.

The study (Y. Deng et al., 2010) presents a hierarchical structural model for the fully automated detection of the NT region in fetal ultrasound images, addressing the limitations of manual measurement, which is prone to variability. The model employs SVM classifiers trained on HOG features to identify the NT, head, and body regions, while a Gaussian pyramid ensures robust detection across different scales. A spatial model incorporating anatomical constraints enhances accuracy, and dynamic programming with a generalized distance transform optimizes the detection process. The model's performance was evaluated on 345 clinical NT ultrasound images. The results demonstrated high precision in detecting fetal structures: body detection achieved 100 accuracies across all configurations, while head detection reached up to 99.4, stabilizing around 97.2 in higher configurations. For NT detection, the model attained an accuracy of 59.0 under optimal conditions. Notably, the spatial model improved NT detection performance by approximately 3.6 compared to using the SVM classifier alone.

The proposed (Sciortino et al., 2017) methodology for measuring NT in ultrasound images involves several advanced techniques to automate the process and improve accuracy. The system uses anisotropic filters and the à trous wavelet transform to preprocess the images and highlight key fetal structures. Neural networks are employed to detect the jaw bone and vertical branch, ensuring the correct identification of median sagittal sections. The choroid plexus is analyzed using a modified circular object detector, and mathematical morphology is applied to enhance the edges of the NT region. The NT thickness is measured as the maximum distance between the upper and lower edges of the detected NT region. The method was validated on a dataset of 3000 ultrasound frames, showing a high correctness rate of 99.95 for locating the NT bounding box and an error of up to 0.3 mm in 97.4 of the measurements compared to manual ground truth.

The work in (Chaudhari & Oza, 2021) proposed a fully automated approach for NT segmentation and thickness measurement. The methodology leverages the Scale-Invariant Feature Transform (SIFT) for key point detection and General Regression Neural Networks (GRNN) for segmentation and measurement. The approach begins with detecting the fetal head using SIFT to identify key

anatomical landmarks, followed by segmentation of the NT region using SIFT features and morphological processing to refine boundaries and reduce noise. The segmented pixels are then converted to actual NT thickness measurements using standard reference charts, ensuring consistency with clinical practices.

The model was trained and validated on a dataset of 500 first-trimester ultrasound images, achieving a significantly lower error rate (0.02) compared to other traditional models such as SIFT-SVM (0.22), SIFT-ANN (0.08), and SIFT-KNN (0.44). As illustrated in Figure 2.29.

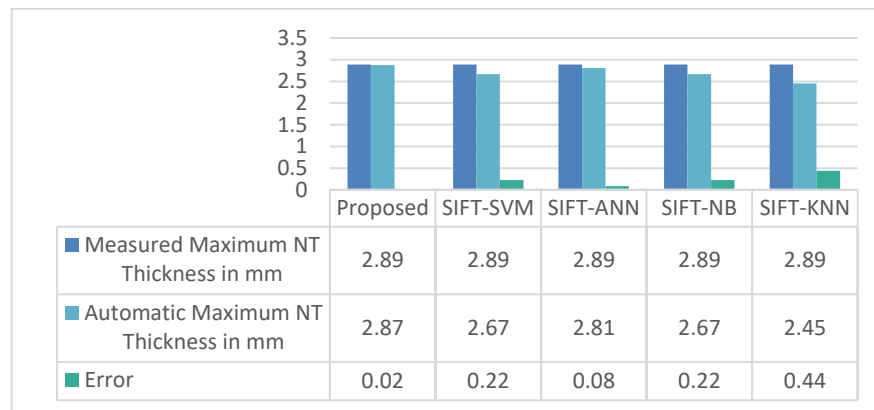


Figure (2.29) Proposed Method Maximum NT Thickness Measurement

The Figure 2.30 compares the accuracy of different methods for measuring minimum NT thickness, showing the proposed method's automatic measurement of 0.77 mm closely matches the measured value of 0.75 mm with a minimal error of -0.02 mm, outperforming SIFT-SVM (0.14 mm error), SIFT-ANN (0.41 mm error), SIFT-NB (0.19 mm error), and SIFT-KNN (0.16 mm error). This highlights the proposed method's superior precision in determining NT thickness.

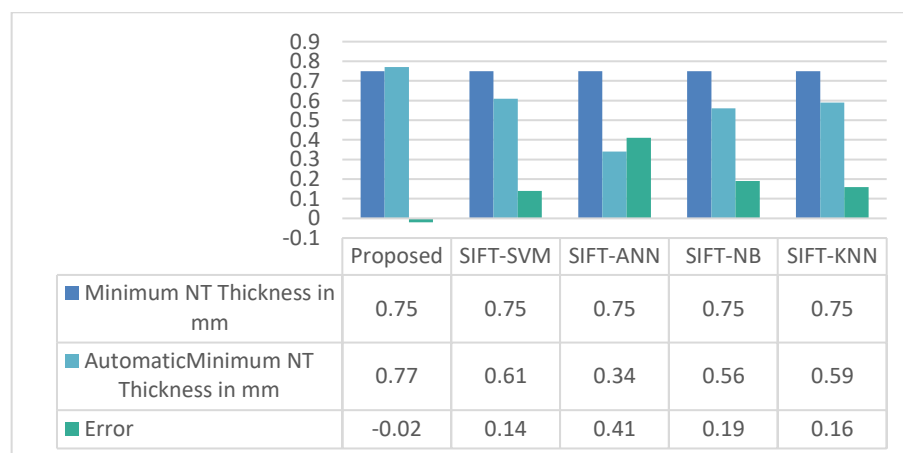


Figure (2.30) Proposed Method Minimum NT Thickness Measurement

This study (Choudhari & Oza, 2024) proposes an automated system for measuring NT thickness in first-trimester ultrasound images using a Markov Random Field (MRF) model. The method combines Fuzzy Clustering (FCM) and the Level-Set Method to segment and refine NT regions. The proposed method achieves a maximum NT thickness error of 0.03 mm and a minimum error of 0.04 mm. The system was validated on a dataset of 200 ultrasound images annotated by specialists, demonstrating its ability to reduce reliance on clinician expertise and improve diagnostic efficiency.

### Maximum NT Thickness Measurements

This Figure 2.31 compares different methods based on their measured and automatically determined maximum NT thickness, along with the error between these measurements. The proposed method shows the smallest error, indicating higher accuracy in measuring the NT thickness compared to GRNN, RBED, and HSM.

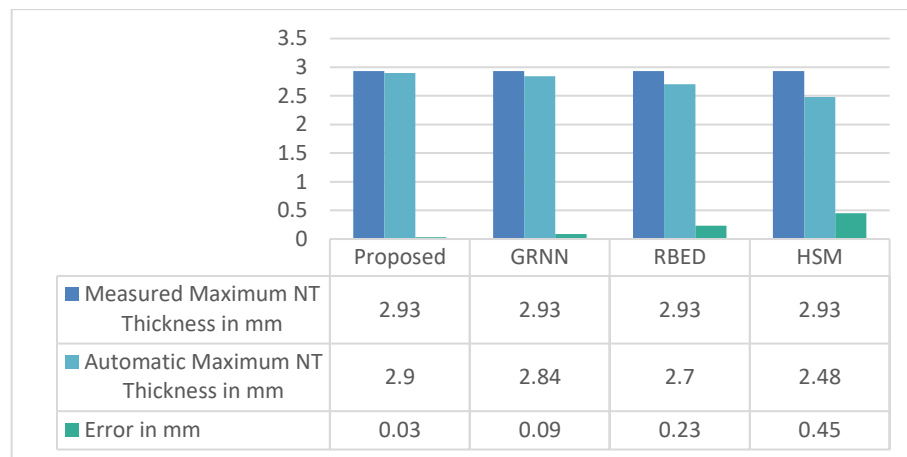


Figure (2.31) Comparison of Maximum NT Thickness

### Minimum NT Thickness Measurements

This Figure 2.32 compares different methods based on their measured and automatically determined minimum NT thickness, along with the error between these measurements. The proposed method shows the smallest error, indicating higher accuracy in measuring the NT thickness compared to GRNN, RBED, and HSM.

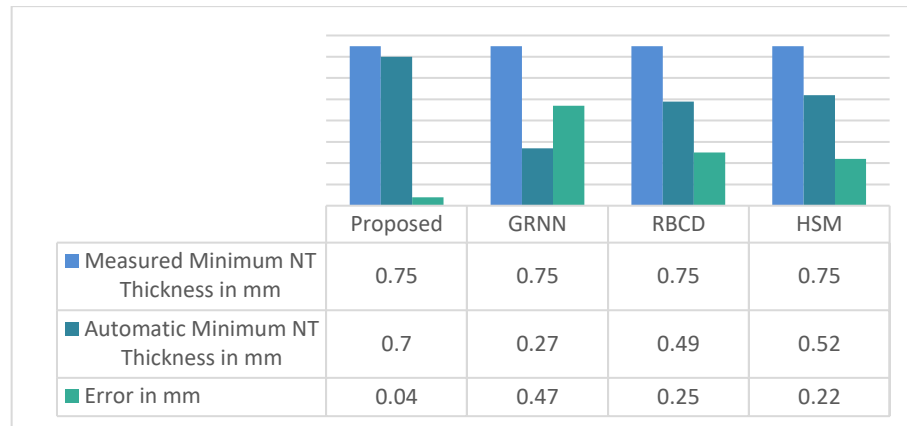


Figure (2.32) Comparison of Minimum NT Thickness

The study (Sonia & Shanthi, 2015) presents an automated method for classifying ultrasound fetal images based on NT thickness using Discrete Wavelet Transform (DWT) and a SVM classifier. The approach includes preprocessing steps such as noise removal using a Lee filter and semi-automatic extraction of the ROI to isolate the NT region. Features are extracted from the images using 2D Haar Wavelet Transform, and an SVM classifier is trained to distinguish between normal and abnormal NT. The method achieved an accuracy of 84, with high sensitivity 93.8 for detecting abnormal NT, although specificity 74.4 was lower, indicating a higher rate of false positives.

The paper (Verma, 2022) introduces an innovative framework for classifying the risk of fetal abnormalities based on NT thickness using the Adaptive Stochastic Gradient Descent Algorithm (ASGDA). challenge of accurately detecting developmental defects, with fetal abnormalities affecting approximately three out of every 1000 pregnancies. Ultrasound scans, commonly used for prenatal screening, can be subject to inconsistencies in manual measurements, which the paper addresses through automation. By using a dataset of 100 fetal ultrasound images, the authors applied preprocessing techniques, including hybrid maxpool matrix histogram analysis, and segmentation via a two-step generic grouping algorithm. The ASGDA model, combining max-pooling and average-pooling, achieved a remarkable accuracy of 98.64%, outperforming traditional methods like CNN and XGBoost. With precision, recall, and F1-score values of 97.2%, 97%, and 96.2%, respectively, the model demonstrates high

performance. This comparative analysis, as visualized in Figure 2.33.

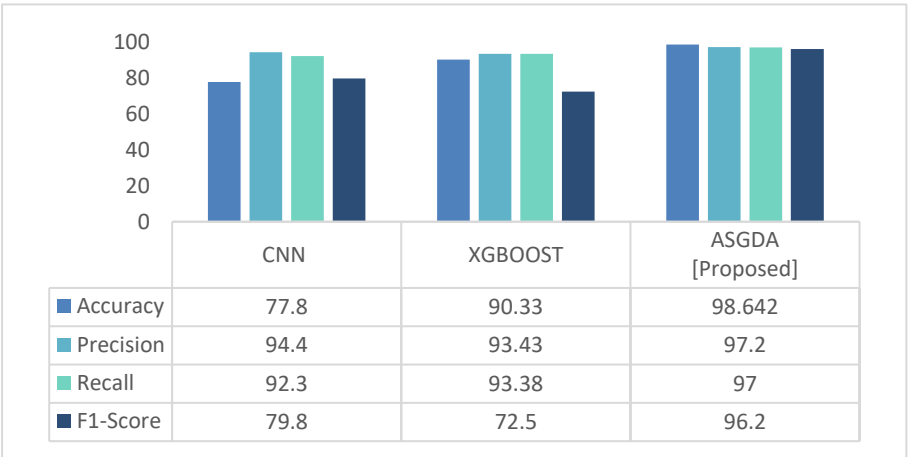


Figure (2.33) Comparative Analysis of Classification Model

## 2.21 Deep Learning Methods for Measuring Nuchal Translucency

The paper (Liu et al. 2025) proposed an intelligent framework for automatic image quality assessment of fetal ultrasound images used in Nuchal Translucency (NT) measurement, addressing the critical issue of diagnostic reliability in early pregnancy screenings. Utilizing a multi-center dataset of 7,728 NT ultrasound images labeled as either standard or non-standard, the study employed deep learning models—VGG16, ResNet50, and Vision Transformer (ViT)—to classify image quality. Among them, VGG16 achieved the best results with an accuracy of 98.1%, sensitivity of 97.8%, and specificity of 98.4%. The study also applied Grad-CAM visualizations to validate that model attention aligned with clinically relevant fetal features, such as the head contour and NT region.

The study (Zhao et al. 2025) introduced a novel deep learning framework—Deep Content and Contrastive Perception (DCCP)—to automate the quality assessment of fetal ultrasound images used for Nuchal Translucency (NT) measurement. The study leveraged two perception streams: one focusing on global content via deep convolutional features, and the other capturing local anatomical fidelity using contrastive learning between key structures like the nasal bone, mandible, and NT region. The dataset comprised 6,402 first-trimester NT images manually labeled by clinical experts as standard or non-standard. The proposed DCCP model significantly outperformed baseline

CNNs, achieving 99.3% accuracy, with improved robustness across diverse imaging conditions. Crucially, the model was also validated using visual interpretability techniques (e.g., saliency maps), confirming that it emphasized clinically relevant features when judging image quality.

The author (Elger et al. 2025) introduced a lightweight and interpretable deep learning framework for automatic quality assessment of fetal NT ultrasound images. The study proposed a novel model called Graph-Based Projection Unit (GPA-Net), which integrates graph-based attention and local attention mechanisms to improve feature extraction and enhance the model's focus on relevant anatomical structures. The dataset consisted of 6,066 ultrasound images labeled as either standard or non-standard, following the FMF (Fetal Medicine Foundation) guidelines. GPA-Net achieved an accuracy of 98.52%, surpassing conventional CNN baselines while maintaining fewer parameters and better generalization. The authors also highlighted the importance of model explainability, using Grad-CAM visualizations to demonstrate that GPA-Net consistently attended to the NT region, mandible, and nasal bone. This study affirms the value of integrating attention mechanisms and graph-based reasoning in prenatal image quality assessment workflows.

The study (Hynek, et al., 2011) proposed a framework to automatically assess the quality of NT ultrasound images using handcrafted features and machine learning models. The study introduced a 12-dimensional feature vector that captured clarity, anatomical visibility, and fetal positioning, and then evaluated its performance using multiple classifiers including SVM, Random Forest, and Logistic Regression. A dataset of 1,290 NT images annotated by clinical experts was used for evaluation, and the best performing model (Random Forest) achieved an accuracy of 92.7% in distinguishing standard from non-standard images. The paper emphasized that combining interpretable features with lightweight classifiers offers a reliable alternative to deep learning.

The author (Zhang et al. 2025) developed a real-time deep learning model for Nuchal Translucency (NT) image quality assessment and measurement in first-trimester ultrasound. Their pipeline begins with image classification using YOLOv7 to detect standard NT planes based on FMF

guidelines. The detected regions are then passed to a ResNet-based regression model to estimate NT thickness directly from the cropped region of interest (ROI). The study used a private dataset of 3,816 ultrasound images with expert annotations and achieved a detection accuracy of 95.5% and a measurement error of 0.15 mm, showcasing the framework's clinical applicability. Unlike prior approaches, this study emphasizes end-to-end automation by integrating plane identification, ROI detection, and NT thickness measurement within a unified architecture.

The study (Palomaki et al. 2008) presented a deep learning model for quality assessment of fetal NT ultrasound images by formulating the task as a binary classification problem distinguishing standard from non-standard views. The study utilized a curated dataset of 6,065 NT plane images, annotated according to established clinical criteria, and trained both 2D CNNs and 3D CNNs to explore spatial and temporal feature representation. Among the models, a 3D ResNet-18 achieved the highest performance, with an accuracy of 95.8%, precision of 96.5%, and recall of 94.6%. The authors emphasized the importance of standardized quality filtering in first-trimester NT screening.

A study (Lin et al., 2022) explored the application of DL models to improve the accuracy and efficiency of fetal head characterization. The proposed Fetus Framework (FF) utilized a multi-step CNN-based approach to detect key fetal intracranial structures, including the thalami, midbrain, palate, fourth ventricle, cisterna magna, NT, nasal tip, nasal skin, and NB. The study aimed to automate the classification of standard (S) vs. non-standard (NS) sagittal views, which are critical for NT measurement and DS risk assessment. The dataset consisted of 1842 2D sagittal-view ultrasound images collected from two medical centers, with expert radiologists providing ground truth annotations. The FF outperformed conventional DL models and human experts, achieving an AUC of 0.996 in internal testing and 0.974 in external validation. Additionally, the framework demonstrated superior sensitivity and specificity compared to first-, third-, and fifth-year clinicians, reducing inter-observer variability and improving the accuracy of standard image selection.

This study (Kasera et al., 2024) presents a deep-learning-based method using the U-Net convolutional Neural Network to automatically segment NT and fetal regions in first-trimester ultrasound images. The goal is to enable early detection of increased NT thickness, a marker for genetic and structural anomalies. The model was trained and tested on a dataset of 560 standardized mid-sagittal ultrasound images from pregnancies between 11–14 weeks, including 88 cases (15.7%) with NT > 3.5 mm. The dataset was split into 451 training images (55 thick NT cases) and 109 test images (33 thick NT cases). Images were annotated by fetal medicine specialists and rated for quality on a Likert scale (1–5), with scores of 4–5 considered high quality. The U-Net architecture, with and without a pretrained ResNet-18 encoder, achieved a Dice score of 0.94 for both NT and fetal segmentation. The NT-to-Fetus Area Ratio (NT: Fetus) demonstrated high diagnostic accuracy, with an AUROC of 0.96 for detecting thick NT (sensitivity: 94%, specificity: 91%). Performance was superior for high-quality images (AUROC: 0.98) compared to low-quality ones (AUROC: 0.90).

This study proposes (M. C. Thomas & Arjunan, 2021) an automated approach for segmenting the NT region in ultrasound images using the Chan-Vese segmentation method. The authors employ a region-based active contour model to address challenges such as weak boundaries and noise. Preprocessing enhances image clarity, and the algorithm iteratively evolves an initial contour to accurately delineate the NT region. The method was validated on a dataset of 100 clinical ultrasound scans (50 abnormal, 50 normal). The methodology involved preprocessing ultrasound images using the Lee filter to reduce speckle noise while preserving structural details, followed by ROI isolation and segmentation. Performance metrics demonstrated the superiority of Chan-Vese segmentation, achieving the highest accuracy (Dice coefficient: 0.79284, Jaccard Index: 0.67055) compared to Level Set and Region Growing techniques.

The author (Liwen et al., 2022) developed and validated a DL model, Trisomy21Net, to screen for Trisomy 21 using first-trimester nuchal ultrasonographic images. The dataset consisted of 822 participants, with 274 fetuses diagnosed with Trisomy 21 and 548 euploid controls, collected from two

hospitals in China between 2009 and 2020. The proposed model, a shallow CNN with 11 layers, was trained on 550 cases (2140 images) and validated on 272 cases (1163 images), focusing on the fetal head region to enhance feature extraction. Trisomy21Net demonstrated superior performance compared to traditional methods that relied solely on NT thickness and maternal age, achieving an AUC of 0.98 in the training set and 0.95 in the validation set, with higher accuracy, sensitivity, and specificity.

In this study (Reshi et al., 2024) proposed model uses a CNN classifier to categorize fetal ultrasound images into two binary classes: "Normal" and "Risky" (indicating a potential risk for DS). The CNN classifier is trained on a dataset consisting of 1,120 pre-processed 2D sagittal-view ultrasound images from fetuses between 11 to 14 weeks of gestation. The model, which consists of 11 layers, employs a self-defined residual block for feature extraction and utilizes dropout techniques to mitigate overfitting. During training, the CNN model achieves high classification performance, with an Area Under the Curve (AUC) score of 0.97 on the training set and 0.94 on the validation set, demonstrating its effectiveness in distinguishing between the "Normal" and "Risky" classes.

Another study (Sciortino et al., 2016) investigated the use of wavelet analysis and neural network classifiers to enhance the automatic detection of MSS in fetal ultrasound videos during the first trimester of pregnancy. The dataset consisted of 3000 frames extracted from 10 real clinical ultrasound videos acquired between the 11th and 13th weeks of gestation. The proposed method aimed to improve the accuracy of NT measurements by automating the identification of key anatomical features, such as the jawbone and choroid plexus, using wavelet transforms and neural networks. The performance evaluation showed promising results, with the model achieving a true positive rate of 87.26%, a true negative rate of 94.98%, and a balanced accuracy of 91.12%. This approach reduces intra- and inter-observer variability, providing a more objective and consistent evaluation for NT assessment.

The authors of (Wee et al., 2010) investigated the application of ANNs to improve the noninvasive prediction of chromosomal abnormalities during the first trimester of pregnancy. Utilizing a dataset of 51,208 singleton pregnancies, the researchers employed nine key parameters, including maternal age, Nuchal Translucency thickness, biochemical markers, and ultrasound findings, to train and evaluate ANN models. The results demonstrated that the ANN-based system achieved a 100% detection rate for trisomy 21 with a 3.9% false positive rate and effectively identified other chromosomal abnormalities with an 84.2% detection rate and a 6.4% false positive rate. The study highlighted the potential of computational intelligence to enhance early prenatal screening.

The authors of (M. Thomas & Arjunan, 2022) developed a deep learning-based framework combining SegNet for NT segmentation and AlexNet for DS classification, aiming to improve early and non-invasive prenatal screening. Using a dataset of 100 ultrasound images (50 DS, 50 healthy), the study applied Wiener filtering for noise reduction, resizing, and normalization to enhance image quality. SegNet achieved high segmentation accuracy 0.98 with a Jaccard Index of 0.96, while AlexNet classified DS cases with 91.7 accuracy and 100 sensitivities. The model effectively reduced operator dependency and human error, overcoming challenges like weak NT edges and ultrasound noise. Despite its success, the study faced limitations such as a small dataset and the time-consuming nature of expert annotations.

As part of the growing field of automated prenatal diagnostics, the paper (Dhumal et al., 2023) introduces a highly efficient deep learning-based system for detecting and measuring NT. The dataset consisted of 500 clinical ultrasound images, annotated by expert sonographers for ground-truth validation. The proposed system utilizes a modified U-Net architecture with auxiliary classifiers, pyramid pooling, and concatenation strategies to enhance the precision and reliability of NT segmentation. Achieving a true positive rate of 99.98% and an error margin of only 0.02mm in NT thickness measurement, the model demonstrates exceptional accuracy and outperforms traditional methods that often rely on manual intervention.

This study (Chaudhari & Oza, 2021) proposed CAD-LNet, a Context Aware DL Network for automatic NT detection in ultrasound images. The framework addresses the limitations of manual NT measurement, which is prone to errors due to variable imaging conditions. Unlike traditional methods, CAD-LNet employs a Region Proposal Network (RPN) to efficiently identify NT regions and uses bounding box regression for precise NT segmentation. The model was trained on 400 ultrasound images and tested on 100 images, achieving a low average error of 0.42, significantly outperforming existing NT measurement techniques. The study demonstrated that CAD-LNet provides a more reliable and faster approach to NT measurement while minimizing the influence of lighting, positioning, and scanning conditions. However, the study acknowledges the limitation of dataset size (500 images) and emphasizes the need for larger datasets to enhance generalization.

Figure 2.34 comparison between manual and software-based NT thickness measurements across different methods. The proposed method demonstrates the smallest deviation between manual and automated measurements, showcasing its high accuracy and consistency in NT assessment.

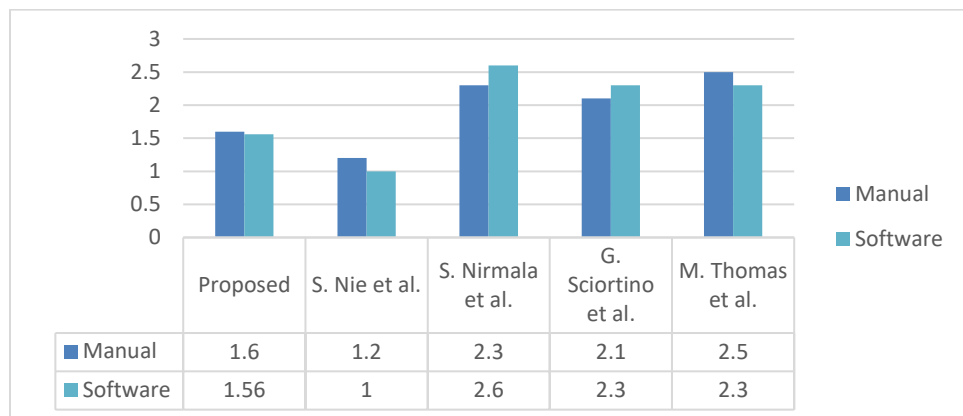


Figure (2.34) Comparison of NT Thickness Obtained with Manual & Proposed Method (Chaudhari, 2021)

- S. Nie et al. utilized dynamic programming with a novel cost function to measure NT thickness, area, and volume from 3D ultrasound data, improving accuracy (Nie et al., 2017).
- S. Nirmala et al. applied Mean Shift Analysis for segmentation, Canny Edge Detection for boundary enhancement, and Blob Analysis for NT measurement (Nirmala & Palanisamy, 2009).

- G. Sciortino et al. leveraged wavelet analysis and neural networks to automate NT detection, reducing human intervention while maintaining accuracy.
- M. Thomas et al. implemented Chan-Vese segmentation, a region-based method for NT identification and segmentation in early gestation to detect DS (M. C. Thomas & Arjunan, 2021).

This paper (T. Liu et al., 2019) presents a deep learning-based approach for direct detection and measurement of NT in ultrasound images, eliminating the need for additional fetal body part annotations. The method includes NT region detection using a CNN (VGG16), segmentation via a modified U-Net, and thickness measurement with PCA. Evaluated on a dataset of 770 labeled ultrasound images, the system achieved a recall rate of 97.01% for NT detection, an IoU of 0.7199 for segmentation, and a mean measurement error of 0.21 mm. The study highlights its potential for improving prenatal screening accuracy and efficiency.

The study by (Chaudhari & Oza, 2024) introduces F-LNET, a DL framework designed to improve the detection and measurement of NT in ultrasound images, aiding early identification of chromosomal abnormalities. F-LNET utilizes a RPN for ROI extraction, followed by NT segmentation through the DL network. The model was trained on 400 images and tested on 100, achieving an NT measurement of 1.66 mm, which outperforms previous studies. With a low error rate of 0.42 mm, F-LNET demonstrates high accuracy, low computational overhead, and clinical utility, reducing variability in manual measurements and providing reliable results even for less experienced practitioners.

Figure 2.35 compares NT thickness measurements obtained through manual methods and various software-based approaches, including the proposed method and those from other studies. The proposed method shows close agreement between manual and software measurements (1.7 vs. 1.66), indicating high accuracy and reliability. In contrast, other studies exhibit notable discrepancies, such as Nie et al. (1.24 vs. 1.02), Nirmala et al. (2.3 vs. 2.64), Sciortino et al. (2.2

vs. 2.43), and Thomas et al. (2.71 vs. 2.45). These variations highlight the challenges in automated NT measurement and underscore the effectiveness of the proposed method in achieving consistent and accurate results, which is essential for reliable prenatal screening.

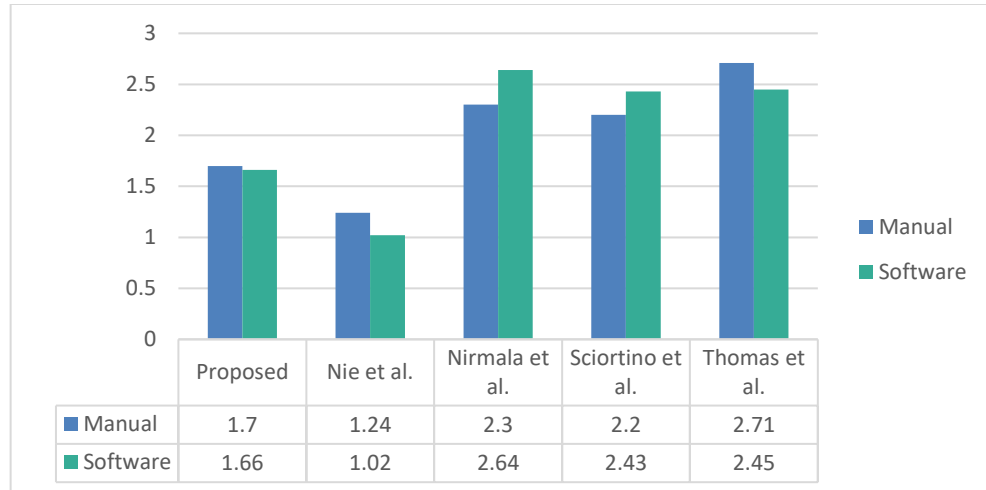


Figure (2.35) Thickness Measurements

The study by (Mehmood et al., 2020) introduces a DL framework that utilizes CNN to predict cardiovascular diseases with 0.97 accuracy, based on a dataset from the UCI repository containing 14 critical heart health attributes. By employing LASSO regression for feature selection and addressing data imbalance, the authors demonstrate the effectiveness of CNN in outperforming traditional ML methods.

The study by (Waheed et al., 2020) proposes an optimized Dense Convolutional Neural Network (DenseNet) for corn leaf disease recognition, achieving 98.06 accuracy on a dataset of 12,332 images divided into four classes: common rust, gray leaf spot, northern leaf blight, and healthy leaves. The model, with only 77,612 parameters, outperforms existing CNN architectures like EfficientNet and VGG19Net in both accuracy and computational efficiency.

The study by (Doan et al., 2022) proposes UATC-DenseNet, for underwater acoustic target classification, achieving 98.85 accuracy at 0-dB SNR on a dataset of 192,000 frames from 12 classes (11 targets and ambient noise). The model, leveraging skip-connections and dense architecture, outperforms traditional ML and other CNNs like ResNet and SqueezeNet in both accuracy and computational

efficiency. This research demonstrates the potential of DL for underwater surveillance, particularly in low-SNR conditions, with future work focusing on advanced time-frequency analysis for further improvements.

The paper (Pregitha et al., 2022) presents a DenseNet approach for classifying DS markers, specifically NT and NB length, in ultrasound images. Early detection of DS, caused by trisomy 21, is crucial, and ultrasound is a non-invasive method for identifying key markers. The study used 100 ultrasound images, including 28 confirmed trisomy 21 cases, with preprocessing through Anisotropic Diffusion Filtering (ADF) to enhance image quality. The DConN, with a 10-layer architecture, achieved a remarkable accuracy of 99.01%. The model showed high recall, specificity, and F1 scores, and ROC curve analysis indicated strong classification performance. The authors conclude that DConN offers significant potential for improving early DS detection in clinical settings.

The proposed method was evaluated against existing approaches to demonstrate its superior efficiency. Performance was assessed using metrics such as specificity, sensitivity, recall, F1 score, and accuracy.

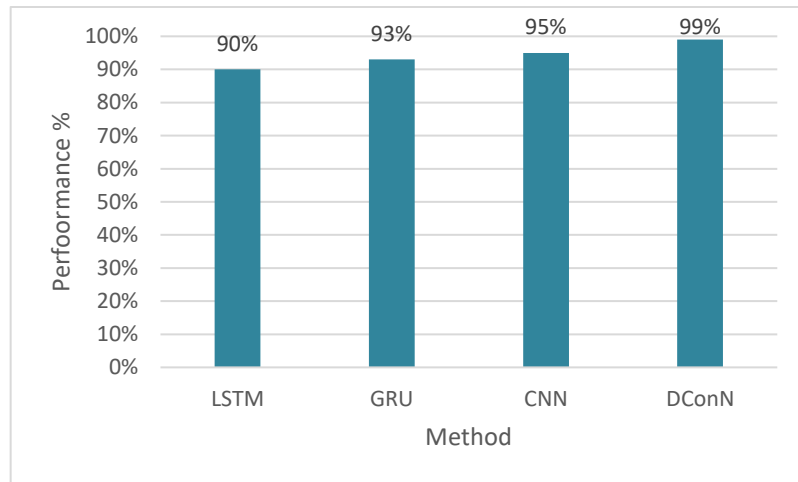


Figure (2.36) Comparison of Traditional DL Methods

Figure 2.36 the performance comparison between traditional DL models and the proposed DConNN reveals that DConNN achieves a maximum efficiency of 99.01%. In contrast, existing models like LSTM, GRU, and CNN achieve efficiencies of 94.55%, 95.6%, and 90.5%, respectively.

Figure 2.37 compared to traditional ML models, which rely on manual feature extraction and are often limited by image noise and variability, the proposed Dense Convolutional Neural Network (DConN) achieves superior accuracy (99.01%), outperforming Cascade ML (95.05%), SIFT+GRNN (97.4%), and Modified AdaBoost (98.6%). While Cascade ML suffers from error propagation and struggles with ultrasound textures, SIFT+GRNN depends on manual feature extraction, making it less adaptable to complex fetal images. Modified AdaBoost, though effective in handling class imbalance, requires extensive tuning. In contrast, DConN leverages deep feature learning, dense connectivity, and optimized convolutional layers to enhance classification precision, offering a 3.9% improvement over Cascade ML, 1.6% over SIFT+GRNN, and 0.41% over Modified AdaBoost, making it a clinically viable tool for automated DS detection.

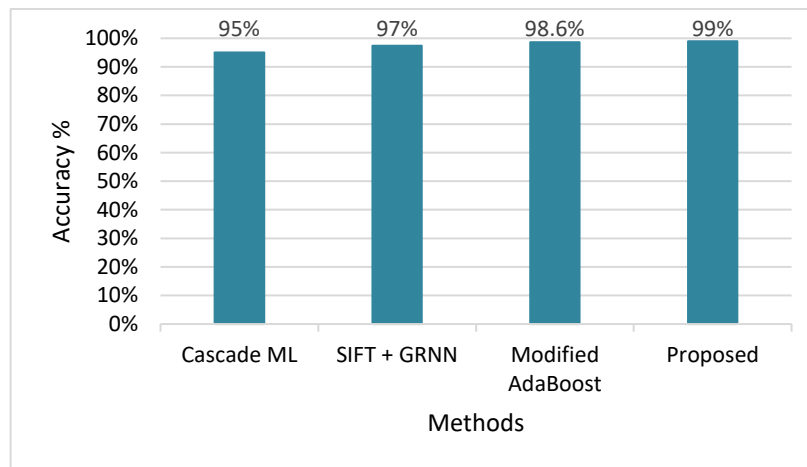


Figure (2.37) Traditional ML Model

Figure 2.38 compared to other deep CNN architectures like CH-DCNN, Optimized DenseNet, and UW-DCNN, the proposed DConN achieves the highest classification accuracy for detecting DS markers. While Optimized DenseNet and UW-DCNN benefit from dense connections, they are not specifically optimized for ultrasound images. The DConN architecture, with its tailored dense blocks and advanced regularization techniques, excels in capturing subtle features like NT and NB length, leading to superior performance in DS detection.

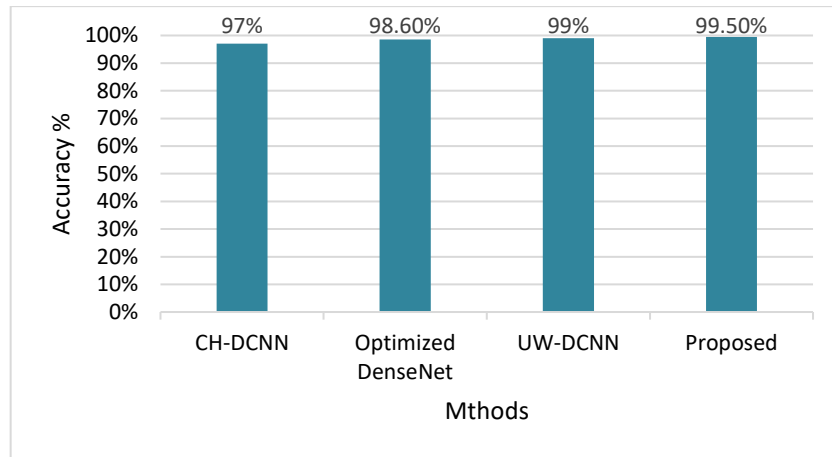


Figure (2.38) Deep CNN Architectures

The author (Yasrab et al., 2022) present a deep learning-based approach for automating CRL and NT segmentation in first-trimester fetal ultrasound images. Using a two-stage architecture Yolo-v5 for sagittal plane detection and a Nested Hourglass (NHG) network for segmentation—the model was trained on 250 ultrasound scans from the PULSE study at Oxford University Hospitals NHS Trust. It achieved a Mean Intersection-Over-Union (mIoU) of 80.74%, outperforming U-Net and SegNet, with a Pearson Correlation Coefficient of  $\rho = 0.93$  against manual segmentation. A weighted loss function improved accuracy by 4.27%, and post-processing via a dense Conditional Random Field (dCRF) enhanced boundary delineation. With 32.5% fewer parameters than U-Net, the model is computationally efficient and suited for clinical deployment, highlighting deep learning's potential in fetal biometry automation.

This study (Yekdast, 2019) proposes an intelligent deep learning-based approach for automated DS detection using CNN optimized with Particle Swarm Optimization (PSO). The authors emphasize the limitations of traditional manual feature extraction techniques and propose a fully automated feature learning and classification system. The dataset, consisting of 300 ultrasound images (171 DS, 129 normal), was processed through a PSO-optimized ConvNet, which achieved a recognition accuracy of 99.38%, significantly outperforming conventional classifiers such as Probabilistic Neural Networks (PNN), Multi-Layer Perceptron (MLP), and Adaptive Neuro-Fuzzy Inference System (ANFIS).

The work (Verma et al., 2022) proposed framework for detecting fetal NT abnormalities using Hybrid Maxpool Matrix Histogram Analysis (HMMHA) represents a significant advancement in prenatal screening, particularly for identifying genetic disorders such as DS. By leveraging ML techniques, this method automates the detection process, reducing the reliance on manual measurements, which are often time-consuming and prone to errors. The framework employs a linear contour size filter for noise reduction and HMMHA for image enhancement and feature extraction, followed by ensemble ML for classification. Tested on a dataset of 33 ultrasound images, the method achieved an error rate of 28.21%, with high performance metrics including sensitivity (90.4%), accuracy (95.8%), and F1-score (96.3%). Compared to established DL models like U-Net, ResNet-18, and DenseNet-121, HMMHA has demonstrated superior classification accuracy and a reduced error rate.

The researchers (Walker et al., 2022) developed a DenseNet169 CNN trained on 289 sagittal ultrasound images (129 cystic hygroma cases and 160 normal Nuchal Translucency controls). The model achieved an overall accuracy of 0.93, with a sensitivity of 0.92 and specificity of 0.94, supported by an area under the ROC curve of 0.94. To ensure interpretability, Grad-CAM heat maps were used, revealing that the model primarily focused on the fetal posterior cervical region, which is clinically relevant for diagnosing cystic hygroma. However, misclassifications (10 false positives and 10 false negatives) occurred, often when the fetus was close to the placental membrane, causing the model to focus on irrelevant areas.

The author (Chaudhari & Oza, 2020) introduced a pose-invariant deep learning-based framework for measuring NT thickness in first-trimester ultrasound images. Their proposed Phase Invariant Architecture (PIA) combines Context-Aware Circumstance Independent Searching (CIS) and a Multi-Level Decision System to enhance NT localization accuracy while handling scale and pose variations. The dataset consisted of 500 fetal ultrasound images, and the model demonstrated superior performance over existing NT measurement methods (OxNNet, Deng, Sciortino), achieving minimal errors (<1 mm in most cases).

The authors in the study of (Tang et al., 2023) introduced a three-stage DL model, FGD5, for the automated detection of genetic disorders in fetuses using ultrasound images. The model integrates an improved YOLOv5 network for ROI extraction, CNNs for disease-specific feature extraction, and an ensemble learning classifier (XGBoost) for final genetic disorder prediction. Trained on a dataset of 2554 ultrasound images, including 1061 cases of genetic abnormalities (e.g. Trisomy 18, and Trisomy 13), FGD5 demonstrated robust performance with an AUC of 0.903, sensitivity of 0.753, and specificity of 0.889 on internal test sets. On external test sets, it maintained an AUC of 0.844, showcasing its generalizability.

The authors in the study of (Aher et al., 2025) proposed an automated DL approach integrating GoogLeNet for NT segmentation and AlexNet for classification. The study utilized 250 first-trimester ultrasound images, including 50 confirmed trisomy 21 cases, with anisotropic diffusion filtering applied for noise reduction. The GoogLeNet model achieved precise NT segmentation, while AlexNet, using transfer learning, classified DS cases with 98.85% accuracy, 99.65% specificity, and 98.05% sensitivity, surpassing conventional methods. This approach enhances early, non-invasive prenatal screening, minimizing operator dependency and improving diagnostic reliability. Figure 2.39 Comparison with another Model.

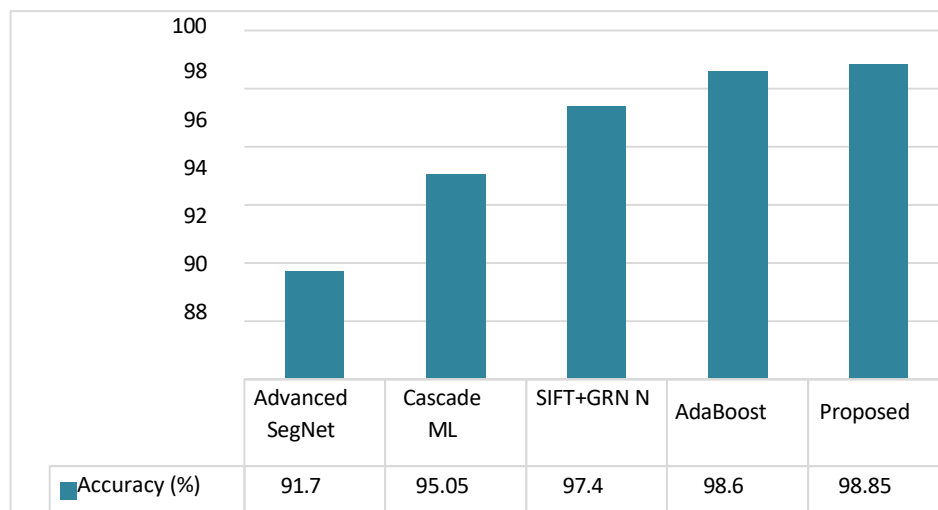


Figure (2.39) Comparison with the Another Model

This study (Ashok & B S, 2024) proposed an Automated Nuchal Cord Graphical Neural Network (NCGNN) for fetal abnormality prediction using DL techniques. The study utilized a dataset of 33 fetal ultrasound images obtained from Children’s Hospital, applying Kalman transloop filtering for noise reduction and Prolonged Wrapper Component Analysis (PWCA) for feature extraction. The NCGNN model was trained to classify NT abnormalities with exceptional accuracy (99.8 sensitivity, 99.7 precision, and 99.2 F1-score), significantly outperforming conventional ML approaches. This method introduces fully automated NT volume analysis, providing a more comprehensive risk assessment compared to traditional 2D-based NT thickness evaluations. Table 2.9 summarizes the comparative performance of the proposed method.

Table (2.9) Performance Comparison of Different Methods in NT Detection

| Method                                  | Sensitivity (%) | Correlation Coefficient (%) | Accuracy (%) | Precision (%) | Recall (%)  | F1-score (%) |
|-----------------------------------------|-----------------|-----------------------------|--------------|---------------|-------------|--------------|
| ML Algorithm (MLA)                      | 74.8            | 64.5                        | 73.8         | 68.4          | 65.5        | 68.9         |
| Feed Forward Neural Network (FFNN)      | 80.2            | 78.3                        | 83.4         | 84.7          | 70.2        | 82.3         |
| Ultrasound Imaging Techniques (UIT)     | 68.4            | 70.9                        | 71.2         | 72.4          | 74.6        | 74.6         |
| Adaptive Risk Prediction System (ARPS)  | 82.6            | 76.4                        | 86.7         | 87.6          | 87.6        | 84.6         |
| Hybrid MaxPool Matrix Histogram (HMMHA) | 90.4            | 92.6                        | 95.8         | 98.1          | 97.2        | 96.3         |
| <b>Proposed</b>                         | <b>99.8</b>     | <b>99.7</b>                 | <b>99.0</b>  | <b>98.0</b>   | <b>99.0</b> | <b>99.2</b>  |

## 2.22 Summary of NT Measurement Research

Table (2.10) presents an overall summary of research on automated Nuchal Translucency (NT) measurement and prediction models. In general, these studies demonstrate the growing use of machine learning and deep learning techniques to enhance diagnostic accuracy and reduce human dependency.

Table (2.10) Summary of NT Measurement Research

| Reference                  | Dataset                                                   | Independent Features                                                                          | Target Feature                  | Prediction Algorithms                            | Best Model                | Outcome (%)                     |
|----------------------------|-----------------------------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------------|--------------------------------------------------|---------------------------|---------------------------------|
| <b>Sun Y (2021)</b>        | 624 fetal ultrasound images (302 trisomy-21, 322 euploid) | NT, prenatal-thickness-to-nasal-bone ratio, frontomaxillary facial angle.                     | Trisomy 21                      | LASSO regression                                 | Nomogram model            | AUC: 0.983                      |
| <b>Devi Vincy (2021)</b>   | 82 ultrasound images                                      | NT measurements                                                                               | Down syndrome (DS)              | Modified AdaBoost + BPNN                         | AdaBoost + BPNN           | 98.67                           |
| <b>Deng Jiaxuan (2025)</b> | 5,519 singleton pregnancies                               | Abdominal circumference, femur length, nuchal thickness, maternal age, gestational age        | Small-for-gestational-age (SGA) | Logistic Regression, SVM                         | SVM                       | AUC: 0.81                       |
| <b>Ramanathan (2018)</b>   | 442 prenatal screening reports                            | NT, $\beta$ -hCG, PAPP-A, CRL, maternal age, smoking, diabetes, nasal bone presence           | Down syndrome (DS)              | Naive Bayes, Random Forest, ANN                  | Naive Bayes               | 92.63                           |
| <b>Li (2019)</b>           | Imbalanced prenatal dataset                               | PAPP-A MoM, $\beta$ -hCG MoM, uE3 MoM, maternal age                                           | Down syndrome (DS)              | CVIFLR (Isolation Forests + Logistic Regression) | CVIFLR                    | AUROC: 0.99                     |
| <b>Park J (2013)</b>       | 196 DICOM ultrasound scans                                | Hierarchical Detection Network (HDN), Dijkstra's algorithm, Graph Cut segmentation            | NT thickness                    | Automated NT measurement system                  | Proposed system           | Low error (0.24 mm)             |
| <b>Deng Y. W. (2012)</b>   | 690 clinical NT ultrasound images                         | HOG features, SVM, generalized distance transform, dynamic programming                        | NT detection                    | SVM + Spatial Model                              | SVM + Spatial Model       | 55.9% (NT), 89.4% (head)        |
| <b>Sciortino (2017)</b>    | 3000 ultrasound frames                                    | Wavelet analysis, feedforward neural networks                                                 | NT measurement                  | Unsupervised wavelet + neural network            | Proposed method           | 97.4% accuracy                  |
| <b>Anzalone (2013)</b>     | Ultrasound video streams                                  | Anisotropic filtering, K-means clustering, dynamic programming                                | NT thickness                    | Automated ROI detection                          | Proposed system           | 93% (sagittal view)             |
| <b>Nirmala (2009)</b>      | 50 ultrasound images                                      | Speckle noise reduction, mean shift analysis, Canny edge detection                            | NT thickness                    | Blob analysis                                    | Automated method          | Reliable NT measurement         |
| <b>Valenti (2017)</b>      | 3000 ultrasound frames                                    | Anisotropic filtering<br>À trous wavelet transform<br>Jaw bone/vertical branch detection (NN) | NT thickness (mm)               | Unsupervised neural networks + edge enhancement  | Proposed automated system | 99.95% bounding box correctness |

|                               |                                                        |                                                    |                                         |                                 |                    |                                        |
|-------------------------------|--------------------------------------------------------|----------------------------------------------------|-----------------------------------------|---------------------------------|--------------------|----------------------------------------|
|                               |                                                        | Choroid plexus analysis<br>Mathematical morphology |                                         |                                 |                    | 97.4% measurements within 0.3 mm error |
| <b>Chaudhari (2021)</b>       | 500 ultrasound images                                  | SIFT keypoints, GRNN                               | NT segmentation                         | SIFT + GRNN                     | Proposed method    | Error: 0.02                            |
| <b>Choudhari (2024)</b>       | 200 ultrasound images                                  | Fuzzy Clustering (FCM), Level-Set Method           | NT thickness                            | MRF model                       | MRF model          | Max error: 0.03 mm                     |
| <b>Sonia (2015)</b>           | Ultrasound fetal images                                | DWT, SVM                                           | NT classification                       | SVM                             | SVM                | 84% accuracy                           |
| <b>Verma (2022)</b>           | 100 fetal ultrasound images                            | Hybrid maxpool matrix histogram analysis, ASGDA    | NT abnormality classification           | ASGDA                           | ASGDA              | 98.64% accuracy                        |
| <b>Li Q. a. (2021)</b>        | 1,842 2D sagittal-view ultrasound images               | Thalami, midbrain, palate, NT, nasal tip, etc.     | Standard vs. non-standard sagittal view | CNN-based Fetus Framework (FF)  | FF                 | AUC: 0.996                             |
| <b>Kasera (2024)</b>          | 560 mid-sagittal ultrasound images (88 NT > 3.5 mm)    | U-Net segmentation, NT-to-Fetus Area Ratio         | Thick NT detection                      | U-Net (ResNet-18 encoder)       | U-Net              | AUROC: 0.96                            |
| <b>Thomas (2021)</b>          | 100 clinical ultrasound scans (50 abnormal, 50 normal) | Chan-Vese segmentation                             | NT segmentation                         | Chan-Vese active contour        | Chan-Vese          | Dice: 0.79284                          |
| <b>Zhang L (2022)</b>         | 822 participants (274 Trisomy 21, 548 euploid)         | NT thickness, nasal bone length                    | Trisomy 21                              | Shallow CNN (Trisomy21Net)      | Trisomy21Net       | AUC: 0.98                              |
| <b>Reshi (2024)</b>           | 1,120 2D sagittal-view ultrasound images               | CNN classifier                                     | Down syndrome risk                      | CNN with residual blocks        | Proposed CNN       | AUC: 0.97                              |
| <b>Sciortino G. a. (2016)</b> | 3000 ultrasound frames                                 | Wavelet analysis, neural networks                  | NT measurement                          | Feedforward neural network      | Proposed method    | 91.12% accuracy                        |
| <b>Wee (2010)</b>             | 51,208 singleton pregnancies                           | Maternal age, NT thickness, biochemical markers    | Chromosomal abnormalities               | Artificial Neural Network (ANN) | ANN                | 100% detection (Trisomy 21)            |
| <b>Thomas M. a. (2022)</b>    | 100 ultrasound images (50 DS, 50 healthy)              | SegNet (segmentation), AlexNet (classification)    | Down syndrome (DS)                      | SegNet + AlexNet                | Proposed framework | 91.7% accuracy                         |

|                           |                                                                 |                                                                         |                          |                                            |                 |                   |
|---------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------|--------------------------|--------------------------------------------|-----------------|-------------------|
| <b>Dhumal (2023)</b>      | 500 clinical ultrasound images                                  | Modified U-Net with auxiliary classifiers                               | NT segmentation          | U-Net with pyramid pooling                 | Proposed U-Net  | 99.98% TPR        |
| <b>Chaudhari (2021)</b>   | 500 ultrasound images (400 train, 100 test)                     | Region Proposal Network (RPN), bounding box regression                  | NT detection             | CAD-LNet                                   | CAD-LNet        | Error: 0.42       |
| <b>Liu (2019)</b>         | 770 labeled ultrasound images                                   | CNN (VGG16), modified U-Net, PCA                                        | NT measurement           | VGG16 + U-Net                              | Proposed method | 97.01% recall     |
| <b>Chaudhari (2024)</b>   | 500 ultrasound images (400 train, 100 test)                     | RPN, DL segmentation                                                    | NT measurement           | F-LNET                                     | F-LNET          | Error: 0.42 mm    |
| <b>Pregitha (2022)</b>    | 100 ultrasound images (28 trisomy 21)                           | NT and Nasal Bone (NB) length                                           | Down syndrome (DS)       | Dense Convolutional Neural Network (DConN) | DConN           | 99.01% accuracy   |
| <b>Yasrab (2022)</b>      | 250 ultrasound scans                                            | Yolo-v5 (sagittal plane detection), Nested Hourglass (NHG) network      | CRL & NT segmentation    | Yolo-v5 + NHG                              | Proposed model  | mIoU: 80.74%      |
| <b>Yekdast (2019)</b>     | 300 ultrasound images (171 DS, 129 normal)                      | PSO-optimized ConvNet                                                   | Down syndrome (DS)       | PSO-ConvNet                                | PSO-ConvNet     | 99.38% accuracy   |
| <b>Verma Deept (2023)</b> | 33 ultrasound images                                            | Hybrid maxpool matrix histogram analysis (HMMHA)                        | NT abnormality detection | Ensemble ML                                | HMMHA           | 95.8% accuracy    |
| <b>Walker (2022)</b>      | 289 sagittal ultrasound images (129 cystic hygroma, 160 normal) | DenseNet169 CNN                                                         | Cystic hygroma detection | DenseNet169                                | DenseNet169     | 93% accuracy      |
| <b>Chaudhari (2020)</b>   | 500 fetal ultrasound images                                     | Phase Invariant Architecture (PIA)                                      | NT thickness             | PIA                                        | PIA             | <1 mm error       |
| <b>Tang (2023)</b>        | 2554 ultrasound images (1061 genetic abnormalities)             | Improved YOLOv5, CNN, XGBoost                                           | Genetic disorders        | YOLOv5 + CNN + XGBoost                     | FGD5            | AUC: 0.903        |
| <b>Sandip Aher (2024)</b> | 250 ultrasound images (50 trisomy 21)                           | GoogLeNet (segmentation), AlexNet (classification)                      | Down syndrome (DS)       | GoogLeNet + AlexNet                        | Proposed model  | 98.85% accuracy   |
| <b>Ashok (2024)</b>       | 33 fetal ultrasound images                                      | Kalman transloop filtering, Prolonged Wrapper Component Analysis (PWCA) | NT abnormality           | NCGNN                                      | NCGNN           | 99.8% sensitivity |

|                                              |                                                           |                                                                                      |                                                      |                                                |                      |                                                        |
|----------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------------------------------------|------------------------------------------------------|------------------------------------------------|----------------------|--------------------------------------------------------|
| <b>Liu et al. (2025)</b>                     | 7,728 fetal NT ultrasound images (multi-center)           | Pixel-level features, fetal head contour, NT region                                  | Image quality (Standard / Non-standard)              | CNNs (VGG16, ResNet50, ViT)                    | VGG16                | Accuracy: 98.1%, Sensitivity: 97.8%, Specificity: 98.4 |
| <b>Zhao et al. (2025)</b>                    | 6,402 first-trimester NT images                           | Global content features, local anatomical features (nasal bone, mandible, NT region) | Image quality (Standard / Non-standard)              | Deep Content and Contrastive Perception (DCCP) | DCCP                 | Accuracy: 99.3%                                        |
| <b>Palomaki et al. (2008)</b>                | 6,065 NT ultrasound plane images                          | Spatial and temporal features from 2D & 3D CNNs                                      | Image quality (Standard / Non-standard)              | 2D CNN, 3D ResNet-18                           | 3D ResNet-18         | Accuracy: 95.8%, Precision: 96.5%, Recall: 94.6        |
| <b>Elger et al. (2025)</b>                   | 6,066 NT ultrasound images (FMF-guided)                   | Graph-based and local attention features                                             | Image quality (Standard / Non-standard)              | CNN + Graph-Based Projection Unit (GPA-Net)    | GPA-Net              | Accuracy: 98.52%                                       |
| <b>Hynek, Stejskal, &amp; Zvárová (2011)</b> | 1,290 NT ultrasound images                                | 12 handcrafted features (clarity, anatomical visibility, fetal position)             | Image quality (Standard / Non-standard)              | SVM, Random Forest, Logistic Regression        | Random Forest        | Accuracy: 92.7%                                        |
| <b>Zhang et al. (2025)</b>                   | 3,816 first-trimester NT images (private dataset)         | Detected ROIs from YOLOv7, ResNet-based regression features                          | NT plane classification and NT thickness measurement | YOLOv7 + ResNet regression                     | YOLOv7 + ResNet      | Detection Accuracy: 95.5%, Measurement Error: 0.15 mm  |
| <b>Lin et al. (2022)</b>                     | 1,842 sagittal-view fetal ultrasound images (two centers) | CNN-based features of fetal intracranial structures (NT, NB, thalami, etc.)          | Image quality (Standard / Non-standard)              | Multi-step CNN (Fetus Framework)               | Fetus Framework (FF) | AUC: 0.996 (internal), 0.974 (external)                |

## 2.23 Fundamental Challenges in Clinical Implementation of ML/DL Systems

The integration of ML and DL models into healthcare, particularly for applications like prenatal screening, holds transformative potential for improving diagnostic accuracy and efficiency. However, several challenges impede their widespread adoption, spanning technical, ethical, and systemic domains. These barriers are critical to address to ensure ML and DL models can be effectively utilized in clinical settings for tasks such as NT measurement.

**First**, data availability and quality pose significant hurdles. ML and DL models require large, diverse, and well-annotated datasets to achieve robust performance. In healthcare, acquiring such datasets is challenging due to privacy regulations, such as the General Data Protection Regulation (GDPR), and the high cost of expert annotations (Esteva, 2019), (Rajkomar, 2018). Ultrasound images, for instance, vary in quality due to differences in equipment, operator skill, and patient factors, complicating model training and generalizability (Souka AP K. E., 2001).

**Second**, model interpretability, often referred to as the "black box" problem, limits clinical trust. DL models, such as CNNs, excel in complex tasks like image segmentation but often lack transparency in their decision-making processes. Clinicians require explainable outputs to validate predictions, particularly for high-stakes applications like detecting chromosomal abnormalities, where errors can have profound consequences (Samek, 2017) (Holzinger, 2017).

**Third**, integration into clinical workflows remains a challenge. Healthcare systems are often fragmented, with legacy infrastructure that is incompatible with advanced computational tools. Deploying ML/DL models requires seamless integration with Electronic Health Records (EHRs) and ultrasound devices, alongside training for clinicians to interpret outputs (Topol, 2019). Resistance to change among medical professionals, driven by concerns over job displacement or liability, further complicates adoption.

**Fourth**, ethical considerations are paramount. Bias in AI models, particularly those trained on demographically skewed data, can lead to unequal outcomes across ethnic or socioeconomic groups (Obermeyer, 2019). Informed consent and data governance are especially complex in prenatal care, where emotional and ethical implications are high (Morley, 2024).

**Fifth**, regulatory and validation barriers continue to delay deployment. Agencies such as the U.S. Food and Drug Administration (FDA) require rigorous testing and approval of AI tools, which may be complicated by the adaptive nature of DL models that evolve over time (He, 2019), (Wiens, 2019).

### 2.23.1 Current Problems and Limitations in Nuchal Translucency Models

ML and DL models for NT measurement in first-trimester ultrasound have shown promise, achieving segmentation accuracies above 95% and detection rates for chromosomal abnormalities up to 99% (Kasera, 2024), (Pregitha, 2022), (Zhang L, 2022). However, specific limitations hinder their reliability and clinical adoption, particularly in the context of NT assessment for conditions like trisomy 21.

- **Small and Non-Diverse Datasets:** Most studies rely on limited datasets, often ranging from 100 to 822 cases (Pregitha, 2022), (Zhang L, 2022). For example, (Kasera, 2024) used 560 ultrasound images, which may not capture the full spectrum of fetal presentations, ethnic variations, or imaging conditions. Small datasets increase the risk of overfitting, reducing model generalizability across diverse populations and clinical settings.
- **Sensitivity to Image Quality:** Ultrasound images are prone to noise, artifacts, and variability due to fetal positioning, maternal factors, and equipment differences (Souka AP K. E., 2001). DL models like U-Net perform optimally on high-quality images but degrade in accuracy with low-resolution or noisy data, which are common in resource-limited settings (Kasera, 2024). This limitation restricts scalability in global healthcare contexts.
- **Lack of 3D Ultrasound Integration:** Current NT models predominantly analyze 2D ultrasound images, overlooking the potential of 3D/4D ultrasound for volumetric NT analysis (Siqing, 2017). 3D data could provide more comprehensive markers, such as NT volume, but its computational complexity and scarcity of annotated 3D datasets limit adoption (Lai, 2011).
- **Inconsistent NT Cutoff Values:** The absence of a standardized NT thickness affects model training and evaluation (Su L, 2021). Models trained on varying cutoffs may produce inconsistent predictions, complicating comparisons and clinical decision-making (Zhang, 2023).
- **Limited Integration with Biochemical Markers:** Most DL models focus on imaging data, neglecting biochemical markers like pregnancy-associated plasma protein-A (PAPP-A) and free  $\beta$ -Human Chorionic Gonadotropin ( $\beta$ -hCG), which enhance screening accuracy when combined with NT (Nicolaidis K. , 2011).

- **Operator Dependency in Validation:** While DL aims to reduce operator variability, many studies still rely on manual annotations for ground truth, introducing potential biases (Moratalla J, 2010). Automated validation methods are underdeveloped, undermining claims of full automation.

## 2.24 Research Gap

The literature reveals significant progress in applying DL to NT measurement, yet several critical gaps remain that this thesis aims to address.

- **First,** current DL models often rely on small, homogeneous datasets—typically fewer than 600 images from single centers (Pregitha, 2022) and (Kasera, 2024). used datasets of 100–560 images, often from single institutions, limiting applicability to diverse populations and imaging conditions. Larger, multi-center datasets are needed to capture variability in fetal anatomy, and ultrasound quality.
- **Second,** the integration of biochemical markers such as PAPP-A and  $\beta$ -hCG with DL-based NT measurement remains underutilized. Although the combined test significantly improves detection rates (Nicolaidis K. , 2011), most DL studies focus solely on imaging data (Zhang, 2023). Developing hybrid models that incorporate both modalities could enhance diagnostic accuracy.
- **Third,** image quality remains a limiting factor. DL models like U-Net exhibit reduced performance when applied to noisy or low-resolution scans, which are common in resource-limited settings (Kasera, 2024). Improving model robustness to image quality variability is essential for equitable clinical implementation.

## 2.25 Conclusion

This chapter reviewed the development of NT measurement methods, tracing the evolution from manual, operator-dependent assessments to advanced DL models. While traditional and semi-automated techniques have helped reduce inter-observer variability, they still depend heavily on human input and lack scalability. Recent DL-based approaches, especially those utilizing architectures like U-Net and CNNs, demonstrate high accuracy in NT segmentation, yet practical clinical integration remains limited due to challenges such as limited dataset diversity, poor model robustness on low-quality scans, and lack of standardization across studies.

## Chapter 3. Methodology

---

### 3.1 Introduction

In the field of Artificial Intelligence (AI), data constitutes a foundational element, serving as the cornerstone for data-driven applications and enabling informed decision-making across diverse sectors (Williams, 2025).

With the advent of the information revolution, the healthcare industry generates vast quantities of data in various forms, including numerical measurements, textual records, and medical images. These data are instrumental in extracting meaningful insights, detecting patterns, and predicting trends that enhance clinical decision-making (Lee S. , 2025). In medical imaging, such data are leveraged to develop models for early disease detection and diagnosis, enabling timely interventions that improve patient outcomes and reduce healthcare costs. In this chapter, present the methodology underpinning the end-to-end pipeline for Nuchal Translucency (NT) analysis in first-trimester ultrasound. We describe the datasets and split strategy, the preprocessing and standardization steps, the model architectures for image quality assessment and ROI detection, segmentation, the measurement protocol used to derive NT thickness in millimeters, and risk classification.

To support the methodology of this research, a visual summary of the entire deep learning workflow is illustrated in Figure 3.1. This figure outlines the major stages involved in the analysis of ultrasound images for NT measurement, including data collection, preprocessing, augmentation, detection, segmentation, classification, and model selection. Each component in the figure highlights the step-by-step flow from raw data acquisition to the final prediction of fetal risk levels.

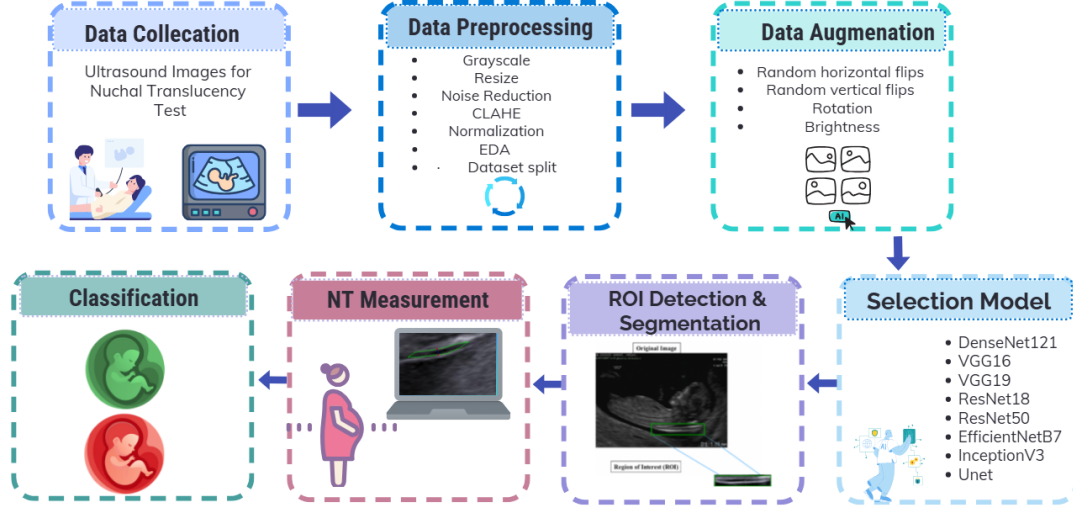


Figure (3.1) Overview of Proposed Pipeline

### 3.2 Data Collection

The development of robust deep learning models for medical image analysis relies heavily on the availability of high-quality, diverse datasets. For this thesis, which focuses on the assessment, detection, segmentation, measurement, and classification of NT in fetal ultrasound images, efforts were made to collect local ultrasound data from hospitals and maternal health centers in Palestine. However, these efforts were unsuccessful due to privacy restrictions, limited data retention, and the absence of standardized medical imaging archives.

As a result, this study uses two ethically approved datasets aligned with the pipeline stages. the public Fetus Framework dataset from Mendeley Data, used primarily for image-quality assessment (C. Chen & Fajin, 2022), and a first-hand (primary) clinical dataset collected directly by the National Hospital of Obstetrics and Gynecology (NHOG), Vietnam, used for ROI detection, segmentation, NT measurement, and final classification.

#### 3.2.1 First Dataset (Fetus Framework)

The Fetus Framework dataset was obtained from the Mendeley Data repository (C. Chen & Fajin, 2022). It contains 1684 sagittal 2D fetal ultrasound images, each labeled by clinical experts as either Standard or Non-Standard according to NT image acquisition quality criteria. These categorical labels make

it highly suitable for training and evaluating deep learning models focused on image quality classification the first step in the proposed methodology.

The dataset was downloaded as a compressed ZIP archive directly from the official Mendeley platform. Data were fully anonymized to ensure patient confidentiality and made publicly available for academic research.

#### **3.2.2.1 Size**

The dataset comprises a total of 1,684 2D sagittal-view ultrasound images collected from 1,519 pregnant women at Shenzhen People's Hospital (C. Chen & Fajin, 2022).

#### **3.2.2.2 Data Type, Format and Views**

The images in the Fetus Framework dataset are provided in widely supported formats, primarily PNG, allowing seamless integration into deep learning pipelines. Each image is labeled according to its view quality as either standard or non-standard, reflecting its suitability for assessment quality for ultrasound image. Standard-view images conform to established clinical guidelines and typically depict a well-aligned mid-sagittal plane, showcasing essential anatomical landmarks such as the Crown–Rump Length (CRL), nasal bone, midbrain, amniotic fluid, and a clearly defined NT region. These images also ensure appropriate fetal positioning, where the head is in a neutral position not hyperflexed or hyperextended and the fetus is magnified so that the head and upper thorax occupy a substantial portion of the frame.

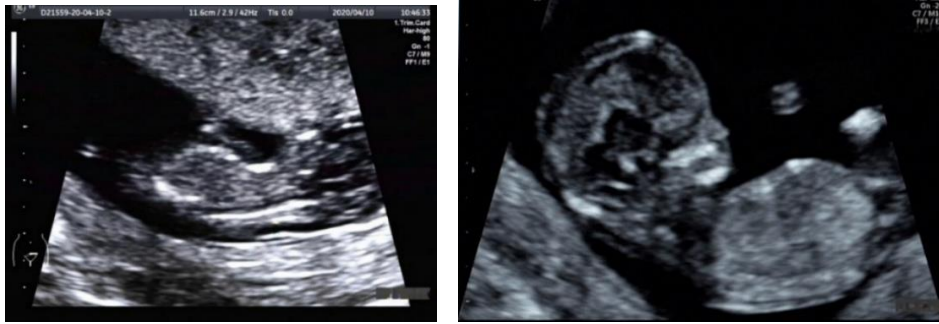
In contrast, non-standard images deviate from these criteria. They may suffer from poor orientation, motion blur, incomplete anatomical coverage, suboptimal magnification, or missing key landmarks, all of which compromise diagnostic reliability. This classification supports the training of quality assessment models aimed at filtering suboptimal images prior to automated NT measurement. A visual comparison of representative examples from each category is shown in Figure 3.2 and 3.3 illustrating the key distinctions between standard and non-standard NT ultrasound views.



(a)

(b)

Figure (3.2) Standard NT Image



(a)

(b)

Figure (3.3) Non-Standard NT Image

### 3.2.2.3 Characteristics

The Fetus Framework dataset provides several key features that make it well-suited for training and evaluating deep learning models in ultrasound image quality assessment:

- **Binary Quality Labels**

Each image is labeled as either Standard or Non-Standard, based on strict fetal NT image quality criteria, including mid-sagittal plane visibility, crown–rump length range, and fetal posture. These labels support supervised learning for classification models.

- **High-Quality 2D Sagittal Images**

The dataset contains 1,684 2D sagittal ultrasound images acquired during the 11–13+6-week gestational window under NT-screening protocols. Native resolutions are heterogeneous (no single fixed dimension); standardize all images to  $224 \times 224$  pixels for model development.

- **Expert Annotations**

Labels were assigned by clinical sonographers or researchers, ensuring reliability and domain accuracy. This is essential for training models that replicate expert-level decision-making.

- **Ethically Approved & Publicly Available**

The dataset is distributed under open academic licensing via Mendeley Data, with all required ethical standards satisfied. This ensures reproducibility and transparency in academic research.

- **Clinical Diversity**

Images were collected from multiple clinical environments, which introduces natural variability in image acquisition settings, ultrasound devices, and fetal anatomy presentation. This enhances the generalizability of models trained on this dataset.

Table 3.1 reports the class distribution in the Fetus Framework dataset: 1,197 Standard (71.1%) and 487 Non-Standard (28.9%), totaling 1,684 images.

Table (3.1 )Class Distribution of The Fetus Framework Dataset

| Class        | Count        | Percent     |
|--------------|--------------|-------------|
| Standard     | 1,197        | 71.1%       |
| Non-Standard | 487          | 28.9%       |
| <b>Total</b> | <b>1,684</b> | <b>100%</b> |

### 3.2.2 Second Dataset (NHOG)

The second dataset comprises 741 fetal ultrasound images collected at the National Hospital of Obstetrics and Gynecology (NHOG) in Hanoi, Vietnam. It was provided by **Mr. Hồ Văn Thép**, a student at the VNU University of Engineering and Technology, for academic and non-commercial research purposes. The dataset was specifically sought to support ROI detection, segmentation, NT measurement, and classification tasks. Native resolutions are

heterogeneous (no single fixed image dimension); standardize all images to 256×256 pixels for modeling.

The images were acquired during routine prenatal screenings and annotated using the Polygonal Segmentation method by experienced obstetricians and gynecologists, ensuring clinically accurate delineation of the NT region. To formally request access, an official letter of commitment was issued by my academic institution, guaranteeing adherence to ethical standards and restricting the use of data to academic purposes only. Upon approval, Mr. Thép shared a secure Google Drive link containing the dataset.

The dataset was transferred in compressed format, including both the original images and accompanying segmentation annotations. Files were reviewed and securely stored in a local research environment for further preprocessing and analysis. All ethical requirements were respected, with patient consent obtained, data anonymized, and strict usage limitations observed.

This collaboration provided access to a clinically rich and expertly labeled dataset that significantly enhanced the scope and depth of this thesis, particularly in evaluating segmentation accuracy and classification performance in real-world ultrasound imaging scenarios.

### **3.2.2.1 Data Collection & Ethics**

To ensure ethical integrity, transparency, and reproducibility in acquiring the private fetal ultrasound dataset used for Region of Interest (ROI) detection, segmentation, NT thickness measurement, and risk classification, a structured and ethically guided data collection pipeline was followed. The dataset was obtained from the National Hospital of Obstetrics and Gynecology (NHOG) in Vietnam, comprising high-resolution 2D ultrasound images from the first trimester. These images represent diverse fetal anatomical views and include pixel-level annotations or calibrated measurement markers, enabling accurate development and evaluation of deep learning models for automated NT analysis.

- **Identified Needs:** The first defined the specific requirements of the dataset, focusing on fetal NT ultrasound images suitable for segmentation, measurement, and classification tasks.

- **Collaborator Search:** Potential collaborators with access to high-quality clinical fetal ultrasound data were identified, and initial communication was established.
- **Official Request Submission:** A formal data access request was drafted and submitted to the relevant medical institution, outlining the research objectives, intended use, and data protection measures.
- **Dataset Approval Process:** The hospital's research or ethics committee reviewed the request and granted permission for data sharing, subject to ethical compliance.
- **Dataset Transfer:** Upon approval, the dataset was securely transferred, ensuring that all privacy protocols were maintained.
- **Data Verification:** The received dataset was reviewed to confirm completeness, correctness of annotations, and alignment with research goals.
- **Compliance and Ethics:** All data handling procedures complied with institutional ethics policies, including confidentiality, consent where applicable, and secure data storage.
- **Integration into Thesis:** Finally, the validated dataset was integrated into the experimental pipeline of this thesis, supporting the training and evaluation of the proposed deep learning.

### 3.2.2.2 Size

The NHOG dataset comprises 741 fetal ultrasound images acquired during routine prenatal examinations at NHOG.

### 3.2.2.3 Data Type, Format, and Views

The NHOG fetal ultrasound dataset consists of 2D grayscale images captured during the first trimester, saved in standard image formats such as PNG. These images represent sagittal views of the fetus and are intended for the assessment of NT and related anatomical structures. Standard-view images clearly show key fetal markers such as the NT space, nasal bone, and mid-sagittal plane with a neutral fetal position, suitable for clinical evaluation. A representative example of a standard-view image from the NHOG dataset is presented in Figure 3.4.



Figure (3.4) Sample Image from The NHOG Fetal Ultrasound Dataset

#### 3.2.2.4 Characteristics

The NHOG fetal ultrasound dataset offers several distinctive features that make it particularly valuable for deep learning tasks focused on Nuchal Translucency (NT) ROI detection, segmentation, measurement, and risk classification:

- **Precise Pixel-Level Annotations**

The dataset includes detailed polygonal segmentation masks of the NT region, manually annotated by experienced obstetricians and gynecologists. These high-precision annotations enable accurate training and validation of segmentation models, ensuring clinical relevance. Figure 3.5. Which showcases a representative NT ultrasound image alongside its corresponding segmentation mask.

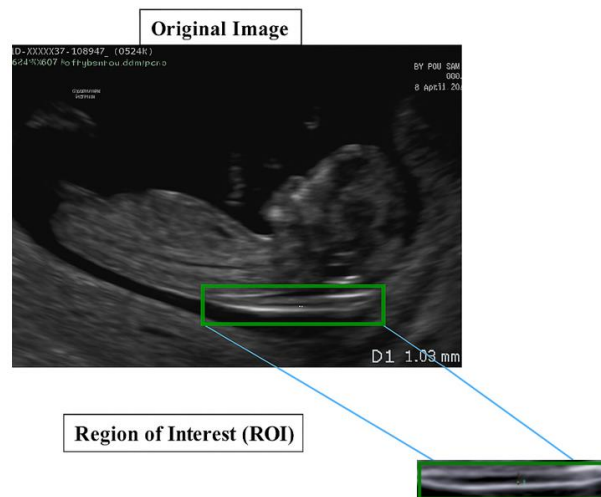


Figure (3.5) Original Images and Corresponding Rois Extraction by The Proposed Method

- **Measurement-Ready Data**

Each image is suitable for direct NT thickness measurement using automated image analysis techniques.

- **Real-World Clinical Variability**

The images were acquired during actual prenatal examinations, capturing diverse fetal positions, anatomical variations, and imaging conditions. This heterogeneity enhances the robustness and generalizability of machine learning models trained on this dataset.

- **High-Resolution**

The dataset includes images in both PNG formats with sufficient resolution to allow fine-grained analysis. This provides flexibility in preprocessing and augmentation strategies for deep learning pipelines.

- **Focused Sagittal Views**

All images capture the fetal profile in the sagittal plane, which is essential for NT assessment according to international guidelines. This consistency ensures the dataset is highly relevant for standardized NT measurement practices.

- **Ethical Approval and Academic Use**

The dataset was shared under strict academic and non-commercial terms, with informed consent obtained and all data anonymized. Its use complies fully with ethical research standards and institutional guidelines.

Table 3.2 shows NHOG annotation availability: 641 images (86.5%) have pixel-level masks and 100 (13.5%) are unlabeled.

Table( 3.2 )NHOG Dataset: Labeled Availability

| Status    | Count | Percent |
|-----------|-------|---------|
| Labeled   | 641   | 86.5%   |
| Unlabeled | 100   | 13.5%   |
| Total     | 741   | 100%    |

### 3.3 Data Preprocessing and Augmentation

Effective preprocessing of ultrasound data is a foundational step in ensuring accurate and meaningful analysis using machine learning and deep learning techniques.

#### 3.3.1 First Dataset-Fetus Framework Pre-processing

To ensure reliable feature extraction and robust model performance in classifying fetal ultrasound images into Standard and Non-Standard quality categories, a comprehensive preprocessing pipeline applies. This pipeline specifically tailors to address the unique challenges posed by ultrasound imaging namely, low contrast, speckle noise, and class imbalance—while preserving critical anatomical features necessary for accurate assessment of NT.

- **Dataset split**

The dataset was partitioned into training, validation, and testing subsets using a 70%–15%–15% split, respectively, to ensure balanced representation across the model development stages. This stratified division supports effective model training, hyperparameter tuning, and unbiased performance evaluation. The distribution of images across the three subsets is visually summarized in Figure 3.6 providing an overview of the dataset split used throughout the experiments.

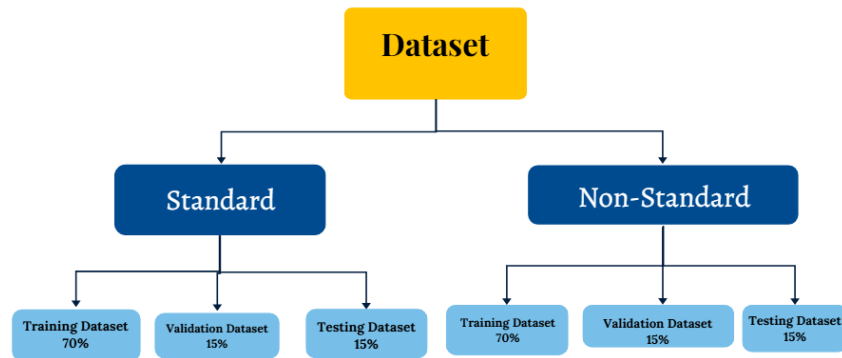


Figure (3.6) Overview of the Dataset Split

- **Image Loading and Color Conversion**

Ultrasound images were loaded from the dataset using the OpenCV library, which by default reads images in the BGR color format (Blue → Green → Red).

However, most deep learning frameworks and visualization libraries, including PyTorch and Matplotlib, require images in the RGB format (Red → Green → Blue) to ensure correct visualization and consistent input processing. Therefore, all images were converted from BGR to RGB.

- **Speckle Noise Reduction**

Ultrasound images are commonly affected by speckle noise, which reduces image quality and obscures critical anatomical details, particularly in the Nuchal Translucency (NT) region. To address this issue, a BM3D-inspired denoising approach was used. The algorithm removes noise while preserving fine structural boundaries, ensuring accurate NT detection and measurement.

Figure 3.7 illustrates the effect of BM3D denoising on a sample ultrasound image at different  $\sigma$  (sigma) values. As shown,  $\sigma = 10$  provides an optimal balance between noise reduction and edge preservation, successfully removing speckle noise while maintaining the NT region's boundaries. When higher sigma values such as  $\sigma = 50$  or  $\sigma = 75$  are used, over-smoothing occurs, resulting in loss of structural details. Therefore,  $\sigma = 10$  was selected for the proposed framework to ensure enhanced image clarity without compromising essential anatomical features.

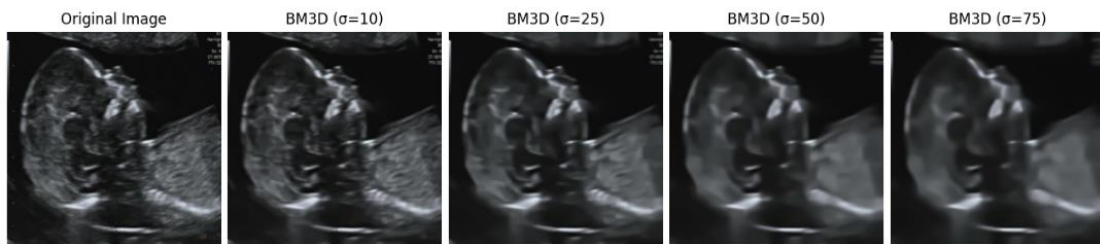


Figure (3.7) Effect of BM3D Denoising on Ultrasound Image Quality

- **Image Resizing**

All images were resized to a standard resolution of 224×224 pixels, to ensure uniformity and compatibility with the deep learning model, and maintaining a consistent input shape across the entire dataset. Figure 3.8 illustrates the resizing operation, showing the transformation of the original ultrasound image into the standardized input resolution.

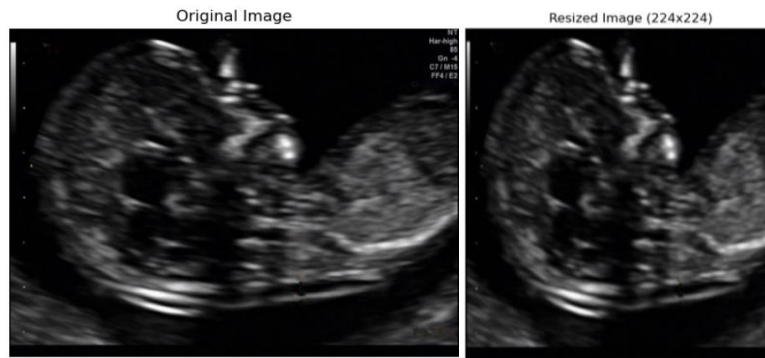


Figure (3.8) Original Image and Resizing Operation

- **Normalization**

Pixel values were normalized using. This standardization facilitates faster convergence during training and ensures compatibility with transfer learning from models pretrained on large-scale image datasets. Figure 3.9 demonstrates the effect of normalization, highlighting the difference between the resized ultrasound image and its normalized counterpart.

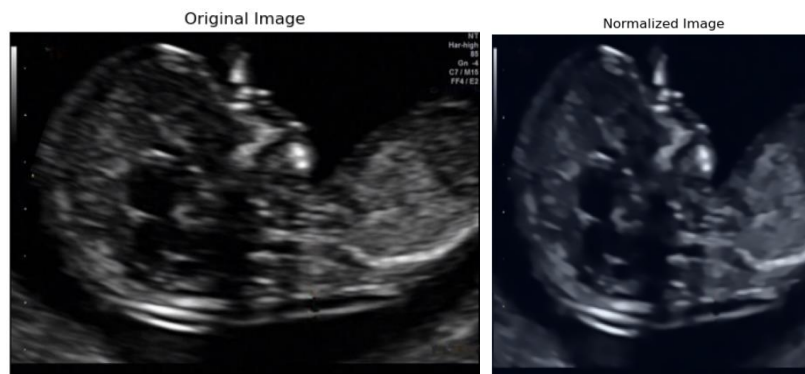


Figure (3.9) Original Image and Normalization Operation

Table (3.3) Summarizes the Preprocessing Steps for Frist Dataset

| Step                    | Method / Tool       | Objective                                      | Output Format    |
|-------------------------|---------------------|------------------------------------------------|------------------|
| <b>Image Loading</b>    | OpenCV              | Read images for processing                     | RGB format       |
| <b>Color Conversion</b> | cv2.cvtColor        | Convert from BGR to RGB                        | RGB images       |
| <b>Denoising</b>        | BM3D                | Remove speckle noise while preserving features | Denoised image   |
| <b>Resizing</b>         | cv2.resize          | Standardize size for model input               | 224 × 224 pixels |
| <b>Normalization</b>    | ImageNet statistics | Scale values for pretrained models             | Normalized image |

- **Data Augmentation (for Class Imbalance)**

To increase the diversity of the dataset and improve the generalization of the proposed framework, several data augmentation techniques were applied during training. These included **horizontal flipping, random rotations by multiples of 90 degrees, Gaussian noise addition, contrast adjustment, and Gaussian smoothing**. The augmentations were applied probabilistically to the training set, ensuring that the model was exposed to varied representations of the ultrasound images while preserving the anatomical characteristics of the Nuchal Translucency (NT) region. This process reduced the risk of overfitting, enhanced the robustness of the framework, and ensured reliable performance across diverse imaging conditions. Figure 3.10 presents examples of augmented images generated from a sample ultrasound scan.

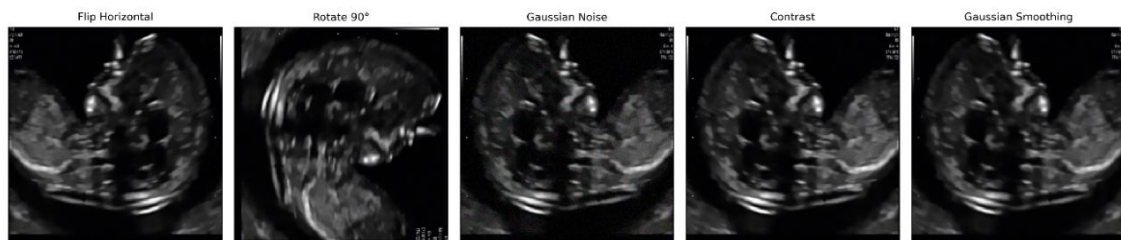


Figure (3.10) Ultrasound Image After Augmentation

### 3.3.2 Second Dataset-NHOG Pre-processing

To ensure reliable feature extraction and robust model performance in the classification of fetal ultrasound images into Normal and High-Risk categories, a comprehensive preprocessing pipeline was developed and applied. This pipeline was specifically tailored to address the unique challenges of ultrasound imaging, namely low contrast and the presence of speckle noise, while simultaneously preserving the critical anatomical features essential for accurate NT classification.

- **Dataset split**

Data partitioning is essential for evaluating a model's performance on unseen data. A common practice is to split the dataset into a training set (used for learning) and a test set (used for unbiased evaluation), typically with a ratio like 80:20. The overall distribution of these subsets is illustrated in Figure 3.11.

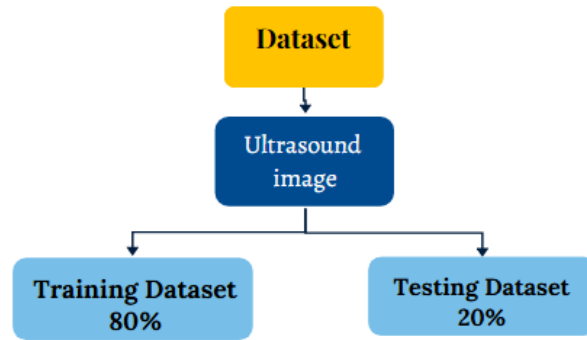


Figure (3.11) Overview of the Dataset Split

### • COCO Annotation Processing

COCO-formatted annotations were converted into binary masks for Nuchal Translucency (NT) segmentation. Annotations were extracted and transformed into binary masks, saved as images in a designated directory. This process ensured a synchronized image-mask dataset for training NT detection models. The Figure 3.12 illustrates the sequential steps involved in preparing segmentation masks from annotated data.

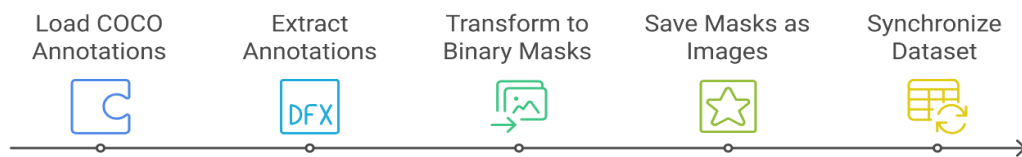


Figure (3.12) Workflow of Annotation Processing and Dataset Synchronization

### • Dataset Integrity and Management

1. Mismatch Detection: Sorted image and mask file lists were systematically compared to identify and report naming discrepancies, ensuring dataset reliability.
2. Batch Renaming: Files were renamed with a consistent prefix (e.g., nt-) and sequential numbering, preserving original extensions to facilitate dataset organization.
3. Selective Copying: Masks were selectively copied to align with existing images, with operations tracked to Verify completeness.

4. **Verification Visualizations:** Subplot visualizations of image-mask pairs and contour comparisons were generated and saved for rigorous quality assurance.

- **Resizing**

All ultrasound images were anisotropically resized to a fixed spatial dimension of 256x256 pixels. This operation is essential for standardizing the input to the Convolutional Neural Network (CNN), as most architectures require a fixed input size. By ensuring all images have the same dimensions, we enable efficient batch processing and streamline the subsequent layers of the network.

- **Normalization**

The pixel intensity values of the resized images, originally in a range such as  $[0, 255]$ , were normalized to a floating-point range of  $[0, 1]$  by dividing each pixel value by 255. This normalization procedure serves two primary purposes: it ensures that all input features have a comparable scale, which mitigates the risk of gradient instability during backpropagation, and it facilitates faster convergence of the optimization algorithms.

- **Noise Reduction**

Given the dominance of speckle and related artifacts in ultrasound, the preprocessing pipeline applies Speckle-Reducing Anisotropic Diffusion (SRAD) to attenuate speckle while preserving anatomical edges.

- **Contrast-Limited Adaptive Histogram Equalization (CLAHE)**

Used to locally enhance contrast and reduce over-amplification of noise, particularly effective for NT region visualization.

- **Data Augmentation Techniques**

To increase dataset diversity and reduce overfitting, augmentation was applied during training:

1. **Rotation** ( $-90^\circ$  to  $+90^\circ$ )

Simulates slight fetal orientation changes during scanning.

2. **Horizontal and Vertical Flipping**

Introduces image variability while preserving anatomical correctness.

### 3. Scaling and Zooming

Helps models generalize across varying NT region sizes.

### 4. Brightness and Contrast Shifts

Accounts for variability in scanning equipment and exposure settings.

Figure 3.13 presents examples of augmented images generated from a sample ultrasound scan.

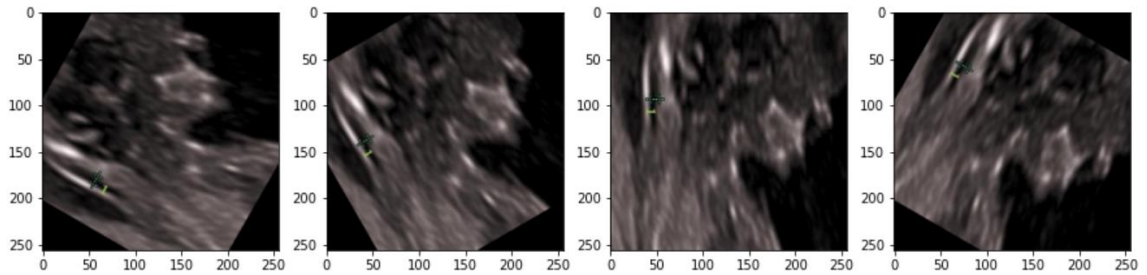


Figure (3.13) Ultrasound Image After Augmentation

Table (3.4) Post-Augmentation Sample Counts for the NHOG Dataset

| Dataset                            | Images (n)    |
|------------------------------------|---------------|
| Original total                     | 741           |
| After augmentation (training only) | 952           |
| Increase                           | +211 (+28.5%) |

Table (3.5) Summarizes the Preprocessing Steps for Second Dataset

| Step                 | Method /Tool                             | Objective                      | Output Format           |
|----------------------|------------------------------------------|--------------------------------|-------------------------|
| Dataset Integrity    | File comparison, renaming, visualization | Ensure alignment & consistency | Verified dataset        |
| Image Loading        | OpenCV                                   | Read ultrasound scans          | RGB format              |
| Resizing (Images)    | OpenCV – 256×256                         | Standardize input size         | 256×256px images        |
| Normalization        | Pixel / 255                              | Scale values to [0,1]          | Normalized image        |
| Noise Reduction      | SRAD                                     | Reduce speckle noise           | Denoised image          |
| Contrast Enhancement | CLAHE                                    | Enhance local contrast         | Contrast-enhanced image |

### **3.4 Noise in Ultrasound Images**

Ultrasound imaging, while invaluable in prenatal diagnostics, is inherently susceptible to a variety of noise types that can significantly degrade image quality and obstruct accurate clinical interpretation. These artifacts originate from both physical and electronic sources and must be addressed through tailored preprocessing techniques.

#### **3.4.1 The Influence of Ultrasound Image Noise on Deep Learning**

Ultrasound imaging is a widely used, non-invasive modality in prenatal care. However, it is inherently affected by various types of noise that degrade image quality, obscure anatomical structures, and reduce diagnostic confidence. In the context of AI, particularly Deep Learning (DL) and Machine Learning (ML) models, such noise can significantly impair model performance by introducing variability and reducing the reliability of extracted features.

Noise disrupts the spatial consistency and intensity patterns within images, which are critical for learning-based models to generalize well. Therefore, detecting and reducing noise is not only essential for human interpretation but also crucial for the success of automated classification, segmentation, or measurement tasks in medical image analysis.

#### **3.4.2 Noise Type Detection Using Deep Learning (CNN)**

To identify the dominant type of noise present in our dataset, we developed a Convolutional Neural Network (CNN)-based classifier. The network was trained on synthetic ultrasound-like images embedded with controlled levels of various noise types (speckle, Gaussian, salt-and-pepper, and quantization).

After training, the model was applied to our ultrasound dataset, and it consistently predicted speckle noise as the dominant noise type. This finding aligns with prior literature, confirming that speckle noise is characteristic of ultrasound due to coherent wave interference.

### 3.4.3 Common Types of Noise in Ultrasound Images

- **Speckle Noise**

Speckle noise is the most prevalent form of degradation in ultrasound images. It emerges from the constructive and destructive interference of reflected sound waves and is inherently tied to the physics of coherent wave imaging. As a multiplicative noise, its intensity varies with the underlying signal, producing a granular, spatially-correlated appearance that can obscure anatomical boundaries (Damerjian et al., 2014).

- **Gaussian Noise**

Introduced primarily by thermal motion and electrical components (e.g., amplifiers) within the ultrasound system, Gaussian noise is additive, uniform, and uncorrelated with the image signal. Its presence leads to subtle blurring and loss of fine contrast (Kavitha et al., 2023).

- **Salt-and-Pepper Noise**

This impulse-like noise appears as sparse, high-contrast black and white pixels, often caused by sensor faults, transmission errors, or probe interference. While less frequent, its high-intensity nature can distort critical features (Kavitha et al., 2023).

- **Quantization Noise**

Quantization artifacts arise during the analog-to-digital conversion process, where continuous signals are mapped to discrete intensity levels. This can lead to a "blocky" appearance, particularly in homogeneous regions of the image (Boncellet, 2009). Illustrative samples of the principal noise types described appear in Figure 3.14.



(a) Spackle Noise



(b) Gaussian Noise



(c) Quantization Noise



(d) Salt and pepper Noise

Table (3.14) Types of Noise in Ultrasound Images

### 3.4.4 Denoising Method Comparison and Implementation

Following noise identification, we applied and compared several noise reduction techniques, including traditional filters and deep learning-based methods, to assess their effectiveness. The methods used include in Table 3.6.

Table (3.6) Denoising Performance: CNN Autoencoder Vs. Traditional Filters

| Method                  | PSNR<br>(dB) | SSIM<br>(%) | Best For                         |
|-------------------------|--------------|-------------|----------------------------------|
| <b>BM3D</b>             | 28.27        | 93%         | Overall denoising & preservation |
| <b>Gaussian Blur</b>    | 25.68        | 84%         | General smoothing                |
| <b>Median Filter</b>    | 23.85        | 74%         | Salt-and-pepper noise            |
| <b>Average Filter</b>   | 23.49        | 73%         | Gaussian noise                   |
| <b>Bilateral Filter</b> | 22.72        | 66%         | Edge-preserving filtering        |

**Each method was evaluated using two metrics:**

- PSNR (Peak Signal-to-Noise Ratio): A measure of image quality post-denoising.
- SSIM (Structural Similarity Index): A perceptual metric that quantifies structural integrity preservation (Horé & Ziou, 2010).

### 3.4.5 Effect of BM3D Denoising

BM3D denoising was tested with  $\sigma$  values of 10, 25, 50, and 75 to evaluate its effect on ultrasound image quality. As shown in Figure 3.15, the lowest value ( $\sigma = 10$ ) provided the best trade-off between speckle noise suppression and structural preservation, achieving the highest PSNR (38.60 dB) and SSIM (0.997). In contrast, larger  $\sigma$  values introduced progressive over-smoothing, which blurred fine anatomical details in the NT region and decreased image quality. These findings confirm that small  $\sigma$  values maintain clinical reliability, while high values compromise diagnostic clarity. Consequently,  $\sigma = 10$  was selected as the optimal parameter for preprocessing in this study.

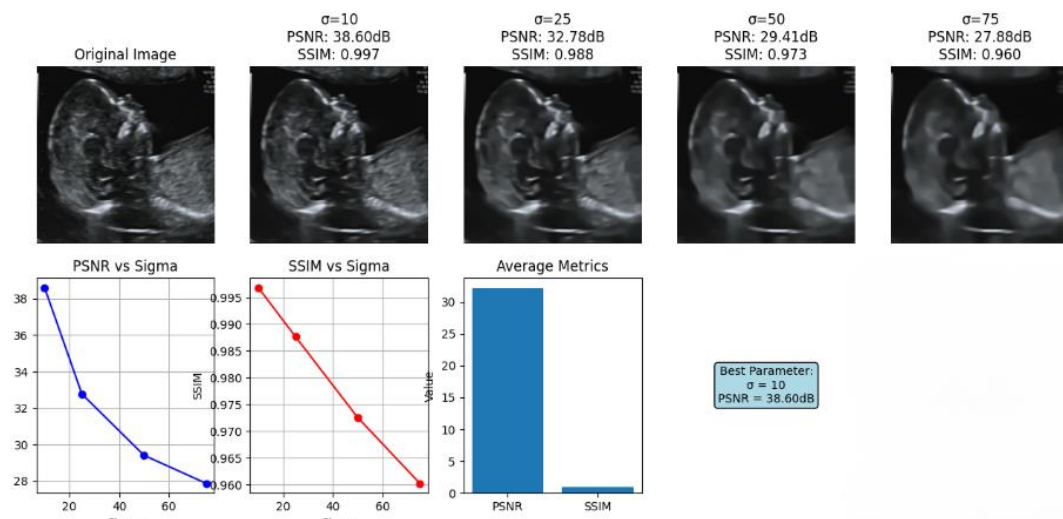


Figure (3.15) Effect of BM3D Denoising with Different  $\sigma$  Values on Ultrasound Images

### 3.5 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial phase in the data science process, serving as the foundation for understanding data distributions, identifying patterns, and detecting anomalies before conducting formal modeling. It combines both statistical methods and visual techniques to uncover meaningful insights that guide the selection of appropriate analytical approaches. This section focuses on applying EDA as a preparatory step for more advanced modeling and interpretation tasks. As illustrated in Figure 3.16. The EDA process encompasses various components including descriptive statistics, univariate and bivariate analysis, multivariate relationships, and the detection of missing data or outliers. This structured approach allows for a comprehensive examination of the data's integrity and underlying structure.

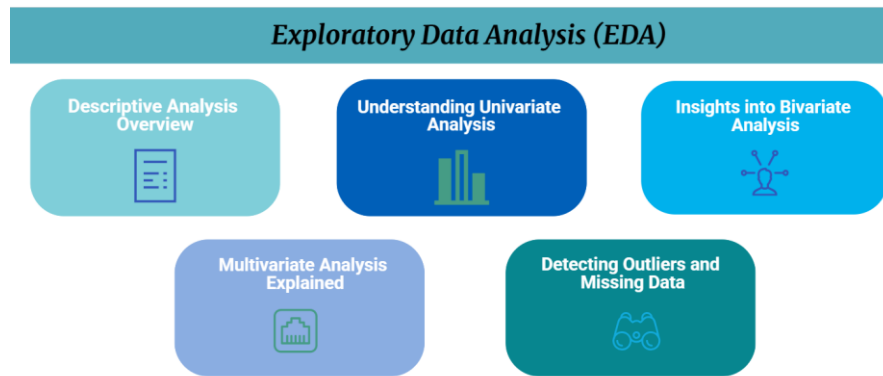


Figure (3.16) Visual Guide to the EDA Process

#### 3.5.1 Original Dataset Exploratory

In this research, two distinct fetal ultrasound datasets were utilized to perform comprehensive exploratory analyses: the Fetus Framework dataset and the NHOG Vietnam dataset.

The first dataset, the Fetus Framework, underwent an initial exploratory data analysis (EDA) to examine the distribution of ultrasound image views. Images were categorized into "standard" and "non-standard" based on established ultrasound image quality criteria. This classification enabled a clear evaluation of the dataset's suitability for subsequent analytical and modeling tasks.

For the second dataset, NHOG Vietnam, due to the absence of standardized metadata or detailed medical history accompanying the images, it was essential to systematically extract relevant clinical information directly from each image. This structured extraction process was conducted using advanced Python scripting, particularly leveraging image processing techniques with libraries such as OpenCV and Tesseract OCR. The detailed metadata extracted includes key variables: image names, NT thickness measurements (in millimeters), gestational age (expressed in weeks), Crown-Rump Length (CRL, in millimeters), and NT status (categorized as either "Normal" or "High Risk"). A detailed snapshot of this structured metadata is shown in Figure 3.17.

Post-extraction, data validation and standardization were meticulously performed to ensure the reliability and consistency of the clinical parameters.

| ID Image View   | Image Name         | NT Thickness_mm | Scan Date  | Gestational_Age_weeks | Crown_Rump_Length_mm | NT_Status |
|-----------------|--------------------|-----------------|------------|-----------------------|----------------------|-----------|
| 1 Standard      | A_000010_10_tr.png | 1.3             | 20-01-2020 | 13+1                  | 72                   | Normal    |
| 2 Non-standard  | A_000010_11_tr.png | 0.98            | 17-02-2020 | 12+3                  | 64                   | Normal    |
| 3 Standard      | A_000010_13_tr.png | 0.95            | 17-02-2020 | 11+6                  | 58                   | Normal    |
| 4 Standard      | A_000010_14_tr.png | 2.3             | 18-02-2020 | 11+6                  | 70                   | Normal    |
| 5 Non-standard  | A_000010_15_tr.png | 2.2             | 18-02-2021 | 13+0                  | 63                   | Normal    |
| 6 Non-standard  | A_000010_17_tr.png | 2.7             | 21-02-2020 | 12+2                  | 57                   | Normal    |
| 7 Standard      | A_000010_18_tr.png | 0.7             | 21-02-2021 | 11+5                  | 66                   | Normal    |
| 8 Standard      | A_000010_19_tr.png | 1.6             | 25-02-2020 | 12+6                  | 74                   | Normal    |
| 9 Standard      | A_000010_20_tr.png | 1.5             | 25-02-2020 | 13+2                  | 66                   | Normal    |
| 10 Standard     | A_000010_21_tr.png | 1.5             | 05-03-2020 | 12+6                  | 66                   | Normal    |
| 11 Non-standard | A_000010_22_tr.png | 1.5             | 05-03-2021 | 12+6                  | 74                   | Normal    |
| 12 Non-standard | A_000010_23_tr.png | 1.2             | 12-03-2020 | 13+2                  | 70                   | Normal    |
| 13 Non-standard | A_000010_24_tr.png | 1.2             | 17-03-2020 | 13+0                  | 64                   | Normal    |
| 14 Standard     | A_000010_25_tr.png | 1.5             | 25-02-2021 | 12+1                  | 62                   | Normal    |
| 15 Non-standard | A_000010_26_tr.png | 1.6             | 05-03-2022 | 11+6                  | 58                   | Normal    |
|                 | A_000010_27_tr.png | 0.7             | 05-03-2023 | 12+4                  | 65                   | Normal    |
|                 | A_000010_28_tr.png | 1.33            | 12-03-2021 | 13+1                  | 72                   | Normal    |
|                 | A_000010_29_tr.png | 1.4             | 17-03-2021 | 12+5                  | 67                   | Normal    |
|                 | A_000010_2_tr.png  | 1.4             | 25-02-2022 | 13+0                  | 69                   | Normal    |

Figure (3.17) Snapshot from the Metadata Sheet for the Fetus Framework Dataset and NHOG Vietnam Dataset

### 3.5.2 Metadata Features

The following tables, specifically Table 3.7 and Table 3.8, provide a comprehensive overview of the metadata features, detailing each column's data type and corresponding values.

Table (3.7) Metadata Features for Frist Dataset-Fetus Framework Dataset

| Feature Name    | Data Type   | Description                                                                                          |
|-----------------|-------------|------------------------------------------------------------------------------------------------------|
| <b>ID Image</b> | Numerical   | Unique filename identifying each ultrasound image                                                    |
| <b>View</b>     | Categorical | Categorization of ultrasound images based on established quality criteria (standard or non-standard) |

Table (3.8) Metadata Features for second dataset-NHOG Vietnam Dataset

| Feature Name                 | Data Type   | Description                                               |
|------------------------------|-------------|-----------------------------------------------------------|
| <b>Image Name</b>            | Categorical | Unique filename identifying each ultrasound image         |
| <b>NT Thickness mm</b>       | Numerical   | Nuchal Translucency thickness measured in millimeters     |
| <b>Scan Date</b>             | Date        | Date when the ultrasound scan was performed               |
| <b>Gestational Age weeks</b> | Categorical | Gestational age expressed in weeks and days (weeks+days)  |
| <b>Crown Rump Length mm</b>  | Numerical   | Crown-Rump Length measurement of the fetus in millimeters |

### 3.5.3 Statistical Analysis for The Original Datasets

This section provides a statistical overview of the original fetal ultrasound datasets used in this study. Descriptive statistics were calculated to summarize key clinical parameters. These measures help in understanding the distribution, central tendency, and variability of the data, offering valuable insights for subsequent analysis and model development.

#### 3.5.3.1 Frist Dataset-Fetus Framework

The Fetus Framework dataset consists of a total of 1,684 ultrasound images, classified into two categories based on established quality criteria: "Standard" and "Non-Standard." The analysis revealed that 1,197 images (71.1%) were categorized as "Standard," indicating optimal quality suitable for detailed analytical processing. Conversely, the remaining 487 images (28.9%) were identified as "Non-Standard," reflecting suboptimal quality which may require additional preprocessing or exclusion from certain analyses. This distribution underscores the dataset's overall suitability for subsequent comprehensive analyses and modeling tasks, with a significant majority of images meeting quality standards, as summarized in Table 3.9.

Table (3.9) Distribution Fetus Framework for Dataset

| View         | Count | Percentage |
|--------------|-------|------------|
| Standard     | 1197  | 71.1 %     |
| Non-Standard | 487   | 28.9 %     |

### 3.5.3.2 Second Dataset-NHOG

The statistical summary for the NHOG Vietnam dataset presents descriptive insights into two key clinical parameters: Crown-Rump Length (CRL), and Gestational Age, show in Table 3.10. The dataset comprises 641 observations for each variable. NT thickness measurements exhibit a mean of approximately 1.82 mm, with moderate variability indicated by a standard deviation of 1.16 mm, and range broadly from 0.20 mm to 9.50 mm. Percentile values further highlight this distribution, with the median NT thickness at 1.40 mm, reflecting a concentration towards lower measurements. Crown-Rump Length (CRL) averages about 65.62 mm, with a standard deviation of 5.98 mm, suggesting typical biological variability in fetal size. The CRL measurements range from 45.00 mm to 83.00 mm, aligning with expected fetal growth in the analyzed gestational period. Gestational age predominantly centers around the late first trimester, averaging 12.65 weeks with low variability (standard deviation: 0.46 weeks), and spans from approximately 11.57 to 13.43 weeks, confirming the dataset's consistency with standard first-trimester assessment timelines. Overall, these statistics underscore the dataset's robustness and suitability for subsequent clinical research and predictive modeling efforts.

Table (3.10) Descriptive Statistics of NT Thickness, Crown-Rump Length (CRL), and Gestational Age in the NHOG Vietnam Dataset

| Statistical Measure | Gestational Age (weeks) | CRL (mm) |
|---------------------|-------------------------|----------|
| Count               | 641                     | 641      |
| Standard Deviation  | 0.46                    | 5.98     |
| Mean                | 12.65                   | 65.62    |

|                     |       |       |
|---------------------|-------|-------|
| <b>Minimum</b>      | 11.57 | 45.00 |
| <b>Maximum</b>      | 13.43 | 83.00 |
| <b>25% Quartile</b> | 12.29 | 62.00 |
| <b>Median</b>       | 12.71 | 65.00 |
| <b>75% Quartile</b> | 13.00 | 70.00 |

For NT thickness measurements (mm), the dataset exhibited a minimum value of 0.20 mm and a maximum of 9.50 mm, indicating a wide range of clinical observations. The mean NT thickness was calculated as 1.82 mm, with a median of 1.40 mm, suggesting a moderate central tendency with slight right skewness (skewness = 2.71), indicative of a distribution with a notable frequency of lower values. The mode was identified at 1.20 mm, highlighting the most frequently occurring measurement. Additionally, the variance and standard deviation values were computed as 1.35 and 1.16 mm, respectively, reflecting moderate variability across the dataset. As summarized in Table 3.11.

Table (3.11) Descriptive Statistics of NT Thickness in the NHOG Vietnam Dataset

| <b>Statistical Measure</b>    | <b>NT Thickness (mm)</b> |
|-------------------------------|--------------------------|
| <b>Total number of values</b> | 641                      |
| <b>Minimum</b>                | 0.20                     |
| <b>Maximum</b>                | 9.50                     |
| <b>Mean</b>                   | 1.82                     |
| <b>Median</b>                 | 1.40                     |
| <b>Mode</b>                   | 1.20                     |
| <b>Variance</b>               | 1.35                     |
| <b>Standard Deviation</b>     | 1.16                     |
| <b>Skewness</b>               | 2.71                     |

### 3.5.4 Data Visualization for The Original Datasets

This section presents a series of visualizations designed to summarize and explore the distribution characteristics of the original fetal ultrasound datasets. By visually examining the composition and quality of the data, these plots provide intuitive insights into dataset balance, variability, and structural patterns. Such visual analyses serve as a critical foundation for evaluating the data sets' readiness and appropriateness for subsequent machine learning and deep learning applications.

This pie chart in Figure 3.18 illustrates the distribution of ultrasound image quality views within the Fetus Framework dataset. A total of 1,684 images were evaluated and categorized into two distinct groups: "Standard" and "Non-standard." The chart reveals that 71.1% of the images (1,197 instances) were classified as Standard, while 28.9% (487 instances) were considered Non-standard. This significant proportion of Standard views suggests that the dataset offers a strong foundation of high-quality images suitable for clinical analysis and algorithmic training. The presence of Non-standard images, however, also presents opportunities for augmentation, filtering, or use in robustness testing for models.

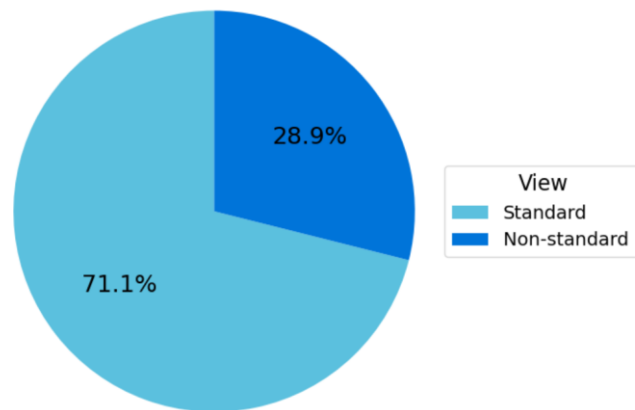


Figure (3.18) Distribution of Standard vs. Non-Standard Views in the Fetus Framework Dataset

The histogram in Figure 3.19 visualizes the distribution of Crown–Rump Length (CRL) values measured in millimeters across the NHOG Vietnam dataset. The distribution is unimodal and approximately Gaussian, with a slight positive skew toward higher values. The peak frequency occurs around 65 mm, which

aligns with the mean and median CRL values reported in the descriptive statistics. This concentration suggests a strong consistency in fetal size measurements during the targeted gestational period (approximately 11.5 to 13.5 weeks). The distribution also highlights a well-covered range of CRL values (45 mm to 83 mm), supporting the dataset's reliability for modeling tasks that require representative sampling across early fetal development stages. The Kernel Density Estimate (KDE) overlay provides a smoothed curve emphasizing the central tendency and spread of the data, reinforcing its suitability for clinical and computational analysis.

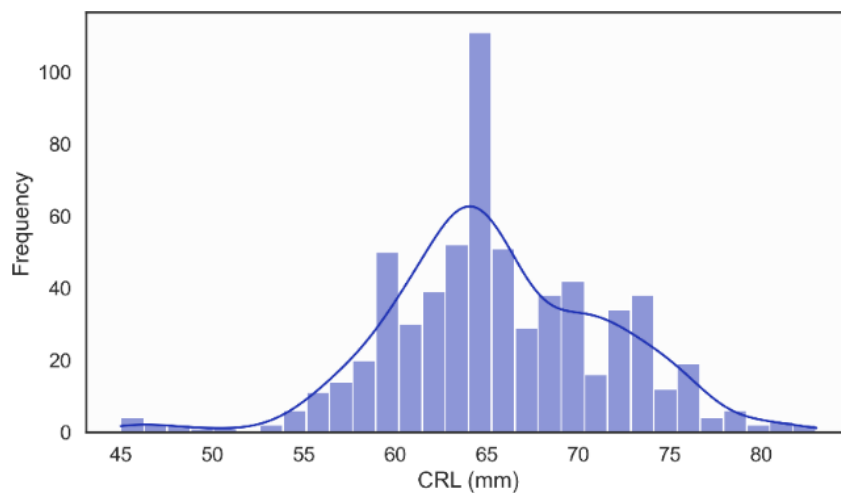


Figure (3.19) Distribution of Crown Rump Length (CRL)

Figure 3.20 presents a histogram illustrating the distribution of Nuchal Translucency (NT) thickness (in millimeters) across the NHOG Vietnam dataset. The distribution is right-skewed, with the majority of NT measurements concentrated between 1.0 mm and 2.0 mm. The peak frequency occurs near 1.2 mm, aligning with both the mode and median reported in the descriptive statistics. This skewness reflects a clinical pattern where most cases exhibit low NT thickness, while higher NT values are less frequent but present in the dataset, extending up to 9.5 mm. The Kernel Density Estimate (KDE) overlay accentuates this trend, providing a smoothed representation of the distribution and highlighting the predominance of typical NT values. Such distribution characteristics are crucial for downstream analyses, ensuring that the dataset

captures both normal and abnormal clinical presentations for robust predictive modeling and clinical research.

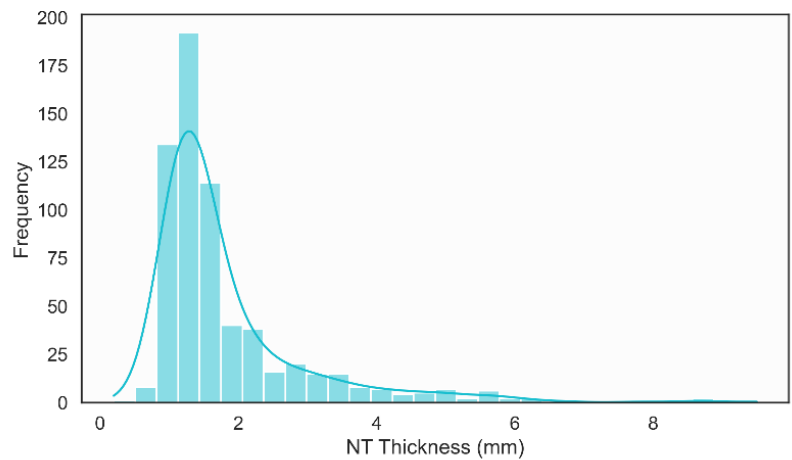


Figure (3.20) Distribution of Nuchal Translucency (NT) Thickness

Figure 3.21 displays the distribution of gestational age (in weeks) among cases in the NHOG Vietnam dataset. The histogram reveals that most ultrasound measurements are clustered between 12 and 13 weeks, with the highest frequency centered just above 12.25 weeks. The overall distribution is relatively symmetric, with a slight central peak and modest spread, as supported by the low standard deviation observed in the descriptive statistics. This pattern reflects targeted data collection during the late first trimester, ensuring consistency in gestational timing across samples. Such distribution confirms the dataset’s alignment with clinical protocols for first-trimester fetal assessment and underscores its appropriateness for research focused on early prenatal screening and risk prediction.

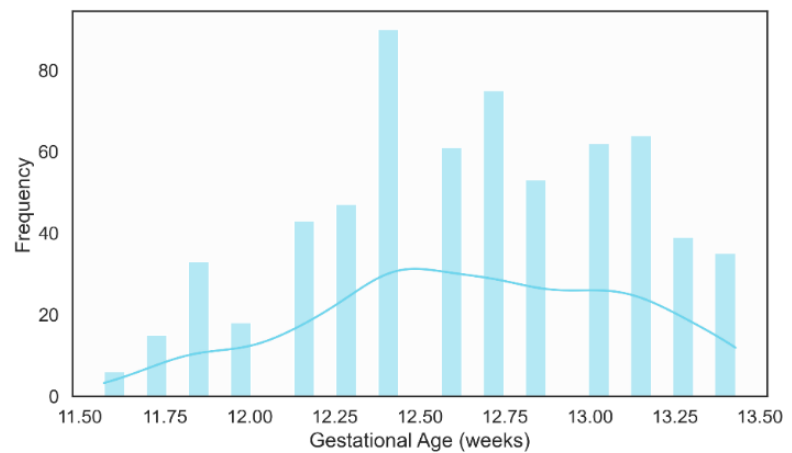


Figure (3.21) Distribution of Gestational Age (weeks)

Figure 3.22 displays a violin plot comparing the distribution of Nuchal Translucency (NT) thickness across two NT status groups: Normal and Increased. The plot highlights that NT thickness measurements in the “Normal” group are tightly clustered around lower values (primarily between 1.0 mm and 2.0 mm), with a relatively narrow distribution and low variance. In contrast, the “Increased” group demonstrates a wider spread and higher NT thickness values, with some measurements extending beyond 9 mm. This distinct separation between the two groups underscores the clinical utility of NT status classification, as increased NT thickness is a well-established marker for chromosomal abnormalities and other fetal conditions. The violin plot effectively visualizes both the central tendency and the spread within each group, reinforcing the diagnostic importance of NT measurement in prenatal screening and supporting robust stratification for further analysis or predictive modeling.

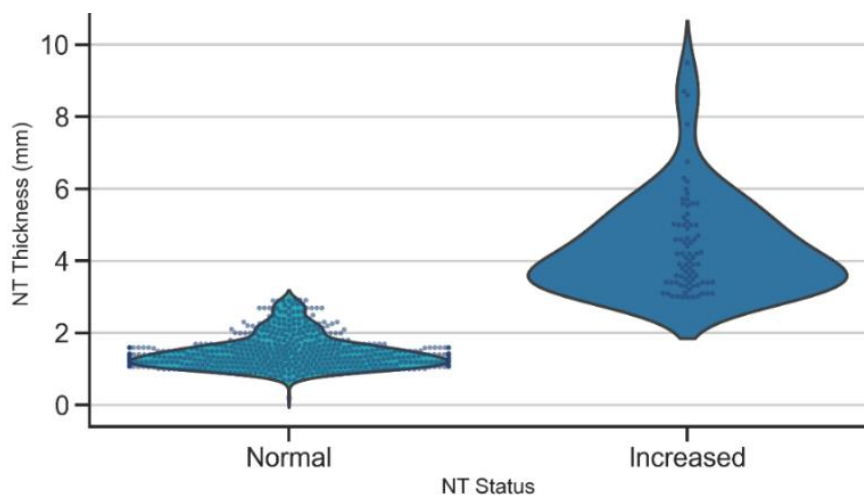


Figure (3.22) Distribution of NT Thickness by NT Status

Figure 3.23 displays scatter plot illustrates the relationship between Crown Rump Length (CRL) and Nuchal Translucency (NT) Thickness, stratified by NT Status, offering valuable insights for fetal assessment. The x-axis represents CRL in millimeters, ranging from approximately 45 mm to 80 mm, while the y-axis displays NT Thickness in millimeters. The data points are clearly differentiated by NT status: 'Normal' (light blue) and 'Increased' (dark blue). A critical observation is the distinct separation of the 'Increased' NT status group, which consistently exhibits higher NT thickness values across the entire range of CRLs.

While there is a general increase in NT thickness with increasing CRL for both groups, the magnitude of NT thickness in the 'Increased' group is substantially greater, often exceeding 3 mm, a clinically significant threshold. This visual evidence reinforces the utility of NT thickness as a robust indicator for identifying fetuses at higher risk for chromosomal anomalies or structural defects, even as CRL progresses. The scatter suggests that while CRL is a proxy for gestational age, the absolute NT measurement remains a powerful independent predictor, warranting further investigation into the specific biological mechanisms underlying these distinct NT phenotypes at varying stages of early fetal development.

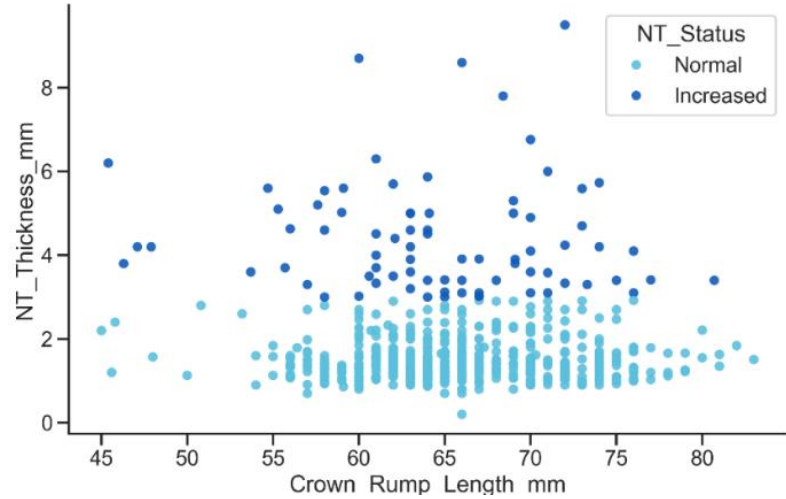


Figure (3.23) Relationship between CRL and NT Thickness by NT Status

This scatter plot in Figure 3.24, depicting Nuchal Translucency (NT) thickness across gestational ages, serves as a compelling visual representation of a key biomarker in prenatal screening. From a research standpoint, the clear stratification of 'Normal' versus 'Increased' NT status, particularly the distinct clustering of higher values in the 'Increased' cohort, highlights the robust discriminatory power of this measurement. The observed slight positive correlation between gestational age and NT thickness in both groups, while present, is overshadowed by the profound difference in the absolute NT values between the two status categories. This suggests that while fetal development naturally influences NT, the primary driver for an 'Increased' classification is

likely an underlying pathological process rather than simply maturational changes. Future research could leverage such visual patterns to refine predictive models, investigate the specific etiologies contributing to the 'Increased' NT phenotype at different gestational time points, and explore the potential for integrating additional biomarkers to enhance diagnostic precision beyond the current visual assessment.

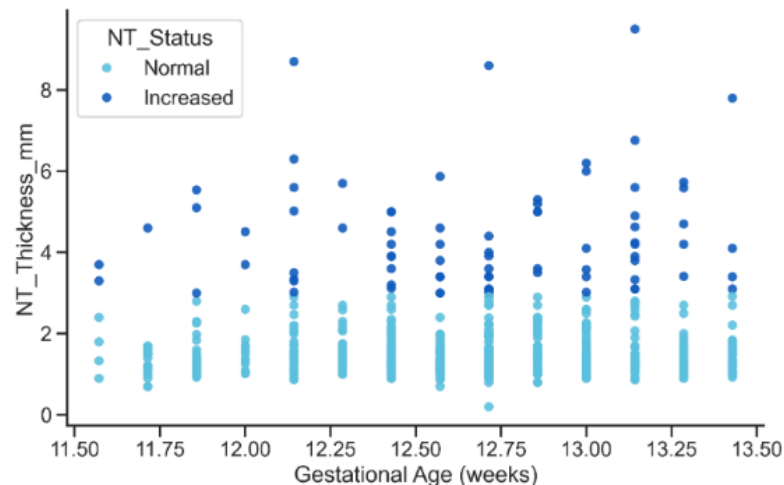


Figure (3.24) Relationship between GA and NT Thickness by NT Status

This scatter plot in Figure 3.25 meticulously illustrates the relationship between Crown Rump Length (CRL) and Gestational Age, stratified by Nuchal Translucency (NT) Status, offering critical insights into fetal growth dynamics within the context of prenatal screening.

From a research perspective, the plot primarily confirms the well-established strong positive correlation between CRL and Gestational Age, as evidenced by the clear upward trend of data points for both NT status groups. This reinforces CRL as a highly reliable biometric parameter for accurate gestational dating. Interestingly, while both 'Normal' and 'Increased' NT status groups generally follow this growth trajectory, there appears to be a subtle, yet discernible, pattern where fetuses with 'Increased' NT status might exhibit a slightly wider dispersion or, in some instances, potentially fall at the lower end of CRL for a given gestational age, particularly in the later weeks of the observed window. Conversely, the 'Normal' NT status group shows a more tightly clustered distribution, consistent with expected growth curves. This observation warrants

further investigation into whether an 'Increased' NT, beyond its direct association with chromosomal anomalies, might also be subtly linked to variations in fetal growth patterns, even within the clinically accepted range for gestational age. Such research could explore the underlying physiological mechanisms contributing to these subtle differences and their potential implications for long-term developmental outcomes, thereby enriching our understanding of the complex interplay between fetal growth, NT development, and overall fetal health.

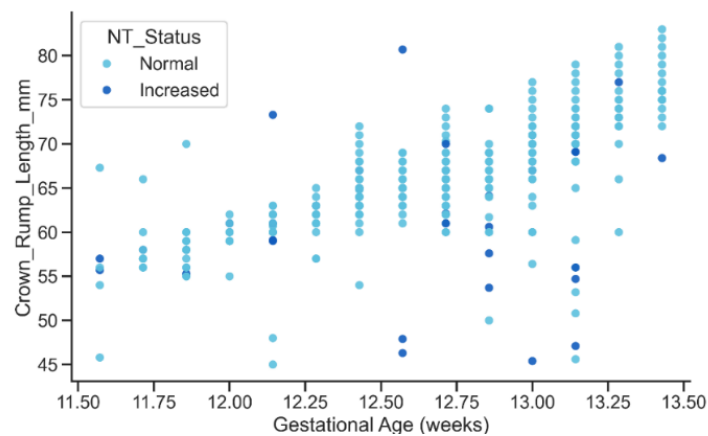


Figure (3.25) Relationship between GA and CRL by NT Status

This heat map in Figure 3.26 to explore the relationships between key clinical measurements in the dataset, a Pearson correlation matrix was computed, including Nuchal Translucency thickness (NT\_Thickness\_mm), Gestational Age (weeks), and Crown-Rump Length (CRL in mm). The analysis revealed a strong positive correlation ( $r = 0.76$ ) between gestational age and CRL, which aligns with established clinical understanding that fetal length increases progressively with gestational development. In contrast, NT thickness exhibited a very weak positive correlation with gestational age ( $r = 0.044$ ) and a weak negative correlation with CRL ( $r = -0.062$ ), suggesting that NT measurement does not significantly co-vary with fetal age or size in this dataset. These findings highlight the relative independence of NT thickness as a biometric marker and support its use in early screening protocols without being overly influenced by gestational age or fetal length.

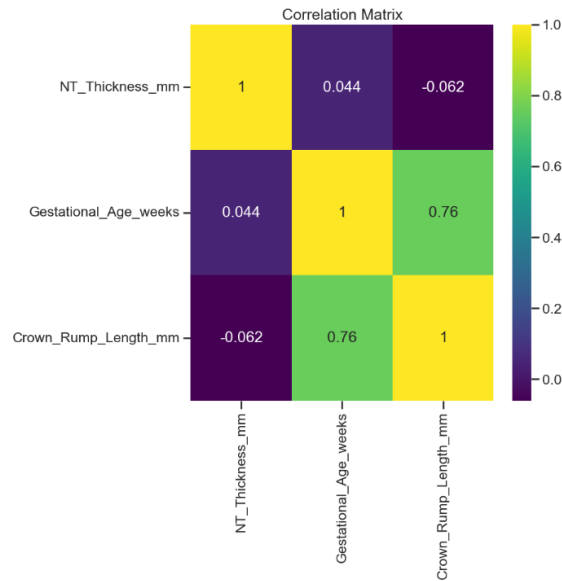


Figure (3.26) Heatmap for Relationships between Key Clinical Measurements in the Dataset

### 3.5.5 Qualitative Data Source and Description

To complement the quantitative analysis of ultrasound imaging data, a qualitative dataset was compiled from user-generated posts on Reddit, focusing on discussions surrounding NT scans during pregnancy. This dataset includes 30+ posts and comments from individuals who shared their experiences, emotions, and medical outcomes related to NT testing, such as missed scans, increased measurements, genetic testing decisions, and psychological impacts.

Each data point in this dataset comprises:

- **Post Title and Body:** The original Reddit post.
- **Comment:** A representative or supportive community response.
- **Metadata:** Including username, year, and categorization tags.
  - **Categorization Fields:**
    - **Category** (e.g., Support and Reassurance, Emotional Impact)
      - **Sentiment** (Positive, Mixed, or Negative)
      - **Key Quotes** (Thematic insights)
      - **Missed NT Scan** (Tagged for contextual analysis)

This dataset offers rich, real-world insights into parental concerns, emotional well-being, and public perception of prenatal screening procedures, as summarized in Figure 3.27. It is particularly valuable for understanding how NT results influence decision-making, psychological health, and trust in medical advice.

| Post Title           | Post Body                              | Comment                                          | Username      | Date | Category         | Sentiment | Key Quotes                         | Missed NT Scan           |
|----------------------|----------------------------------------|--------------------------------------------------|---------------|------|------------------|-----------|------------------------------------|--------------------------|
| Feeling upset I      | I was supposed to get a nuchal translu | We are doing NT, however I read today            | External_C    | 2023 | Missed NT Scan   | Mixed     | Missed NT Scan - Mixed             | High NT Measurement      |
| Feeling upset I      | I was supposed to get a nuchal translu | Tried to do it. Baby wouldn't cooperate          | External_C    | 2023 | NIPT vs. NT Scan | Neutral   | Testing - Nuchal Translucency,     | Emotional Impact         |
| Feeling upset I      | I was supposed to get a nuchal translu | We did not do the NT; in fact, my doct           | External_C    | 2023 | NIPT vs. NT Scan | Positive  | Testing - NIPT, Testing - AFP,     | Decision-Making          |
| Feeling upset I      | I was supposed to get a nuchal translu | The NT is also an ultrasound that look           | External_C    | 2023 | NIPT vs. NT Scan | Positive  | Testing - NT Scan, Testing - NIPT, | NIPT vs. NT Scan         |
| Feeling upset I      | I was supposed to get a nuchal translu | This is true, but it's not helpful to OP         | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          | Termination Decisions    |
| Feeling upset I      | I was supposed to get a nuchal translu | I'm pretty sure the nuchal fold doesn't          | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          | Support and Reassurance  |
| Feeling upset I      | I was supposed to get a nuchal translu | This is actually incorrect. NIPT tells           | yo External_C | 2023 | Support and      | Positive  | Support and Reassurance -          | Correction/Clarification |
| Feeling upset I      | I was supposed to get a nuchal translu | I wasn't aware NT was useful for anyth           | External_C    | 2023 | Support and      | Neutral   | Support and Reassurance - Neutral  | Personal Experiences     |
| Feeling upset I      | I was supposed to get a nuchal translu | You're totally correct for trisomy 21 (D         | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          |                          |
| Feeling upset I      | I was supposed to get a nuchal translu | Just to be a bit of stickler for correctn        | External_C    | 2023 | Support and      | Neutral   | Support and Reassurance - Neutral  |                          |
| Feeling upset I      | I was supposed to get a nuchal translu | Yea but main concern is checking the             | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          |                          |
| Feeling upset I      | I was supposed to get a nuchal translu | This government site has a good comp             | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          |                          |
| Feeling upset I      | I was supposed to get a nuchal translu | I'm beginning to sound like a broken re          | External_C    | 2023 | Support and      | Neutral   | Support and Reassurance - Neutral  | Positive                 |
| Feeling upset I      | I was supposed to get a nuchal translu | This is super helpful! I'm going to shar         | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          | Negative                 |
| Feeling upset I      | I was supposed to get a nuchal translu | I'm glad to see others say this. I find it       | External_C    | 2023 | Support and      | Neutral   | Support and Reassurance - Neutral  | Neutral                  |
| Feeling upset I      | I was supposed to get a nuchal translu | I did the NT but baby wouldn't stop mo           | External_C    | 2023 | Support and      | Positive  | Support and Reassurance -          | Mixed                    |
| Nervous for nuchal   | I am 12W5D and have my nuchal tran     | 34 is not that old. And an increased N           | curiousbea    | 2025 | Support and      | Positive  | Support and Reassurance -          |                          |
| Positive experiences | the scan went great, our baby looks a  | I don't have experience of this yet, but swinva4 |               | 2024 | Support and      | Neutral   | Support and Reassurance - Neutral  |                          |
| Positive experiences | the scan went great, our baby looks a  | On re-read I guess I missed that you diswinva4   |               | 2024 | Missed NT Scan   | Neutral   | Support and Reassurance - Neutral  |                          |

Figure (3.27) A Snapshot of the Qualitative Dataset as Collected and Structured

### 3.5.5.1 Insights from Analysis of Nuchal Translucency (NT)

#### • Emotional and Psychological Impact of Abnormal NT Results

The data reveals that abnormal or high NT measurements (e.g.,  $>3$  mm) are a significant source of emotional distress for expectant parents. Many users describe feelings of anxiety, fear, and uncertainty upon receiving such results, particularly when combined with inconclusive or conflicting information from healthcare providers. The emotional toll is exacerbated by the waiting period for further diagnostic tests (e.g., NIPT, CVS, or amniocentesis), during which parents often experience heightened stress and difficulty coping with daily life. This aligns with existing literature on the psychological impact of uncertain prenatal diagnoses (e.g., Lalor et al., 2009).

#### • NT Scan vs. NIPT: Complementary but Distinct Roles

A recurring theme in the data is the comparison between NT scans and NIPT. While NT scans measure the fluid at the back of the fetus's neck and can indicate structural abnormalities or chromosomal issues, NIPT is a blood test that screens for common chromosomal abnormalities (e.g., Trisomy 21, 18, 13) with higher accuracy. However, NIPT does not detect structural abnormalities, which underscores the complementary nature of these tests. Some user's express

confusion about whether both tests are necessary, while others rely solely on NIPT due to its perceived reliability. This highlights the need for clear communication from healthcare providers about the distinct purposes and limitations of each test.

- **High NT Measurements: Outcomes and Decision-Making**

The dataset includes a range of outcomes associated with high NT measurements. While some users report positive outcomes (e.g., healthy babies born after further testing), others share heartbreaking stories of termination for medical reasons (TFMR) due to confirmed chromosomal abnormalities or severe health issues. The uncertainty surrounding high NT measurements often leads to difficult decision-making, particularly when invasive diagnostic tests (e.g., CVS or amniocentesis) are recommended.

- **Support and Reassurance in Online Communities**

A notable finding is the role of online communities in providing emotional support and reassurance to individuals facing uncertain prenatal diagnoses. Users frequently share positive stories of health outcomes despite high NT measurements, offering hope to others in similar situations. Additionally, users emphasize the importance of avoiding over-reliance on internet research (e.g., "Dr. Google"), which can exacerbate anxiety. This aligns with research on the benefits of peer support in coping with prenatal diagnoses (Chojenta et al., 2014).

- **Misunderstandings and Misinformation**

The data also reveals instances of misunderstandings and misinformation about NT scans and NIPT. For example, some users mistakenly believe that NIPT can detect structural abnormalities, while others are unaware of the optimal timing for NT scans (11-13 weeks). These misunderstandings can lead to unnecessary anxiety or incomplete prenatal screening. This underscores the importance of patient education and clear communication from healthcare providers.

- **Financial and Insurance Barriers**

Several users' express concerns about the cost of additional testing (e.g., NIPT, CVS, amniocentesis), particularly when insurance does not cover these

procedures. Financial barriers can limit access to comprehensive prenatal screening, potentially delaying or preventing the detection of fetal abnormalities.

- **Positive Outcomes and Resilience**

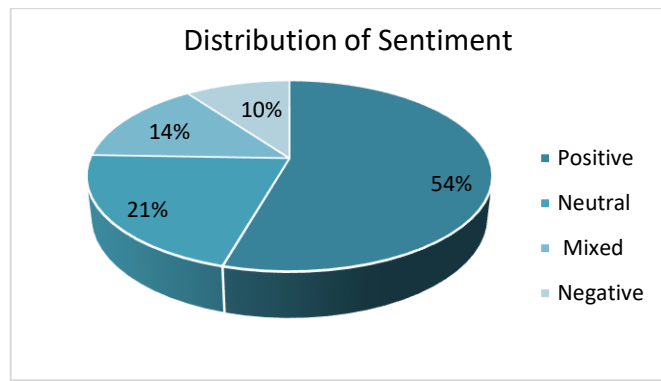
Despite the challenges, many users share positive outcomes and express gratitude for the support they received from online communities and medical professionals. These stories of resilience and hope serve as a reminder that high NT measurements do not always indicate serious health issues. However, the emotional journey of navigating uncertain prenatal diagnoses remains a significant burden for many expectant parents.

### **Implications for Research and Practice**

- **Patient Education:** There is a need for clear and accessible information about the purposes, limitations, and timing of NT scans and NIPT. Healthcare providers should prioritize patient education to reduce misunderstandings and anxiety.
- **Emotional Support:** The emotional toll of uncertain prenatal diagnoses highlights the importance of psychological support for expectant parents. Online communities can play a valuable role in providing peer support, but healthcare systems should also integrate mental health services into prenatal care.
- **Equitable Access:** Financial and insurance barriers to prenatal testing must be addressed to ensure equitable access to comprehensive screening and diagnostic services.

### **Conclusion**

The analysis of user-generated data on NT scans and prenatal testing reveals a complex interplay of medical, emotional, and social factors, as illustrated in Figure 3.28. While high NT measurements can indicate serious health issues, they are not always predictive of negative outcomes. The experiences shared in this dataset underscore the importance of clear communication, emotional support, and equitable access to prenatal testing.



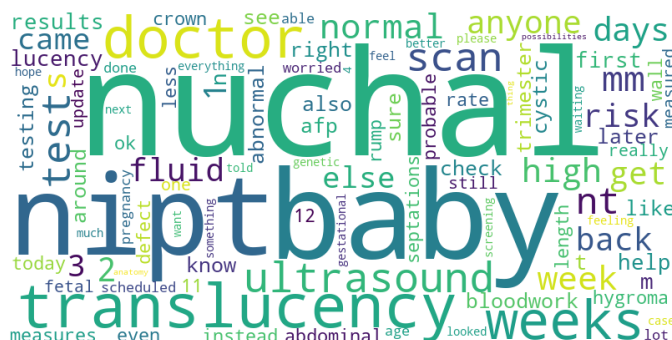
### 3.5.5.2 Visualizing Key Themes: Word Frequency Analysis

To better understand the dominant concerns and recurring terms in the collected Reddit dataset, a word frequency analysis was conducted, and the results are presented as a word cloud in Figure 3.29. This visual tool highlights the most frequently mentioned words.

As shown in the figure 3.29, prominent terms include:

- **"nuchal"**, **"translucency"**, **"baby"**, and **"nt"** – central to the scan topic.
- **"nipt"**, **"ultrasound"**, and **"fluid"** – indicating diagnostic procedures and clinical concerns.
- **"doctor"**, **"risk"**, **"weeks"**, and **"results"** – illustrating the importance of timing, medical advice, and anxiety about outcomes.

This visualization confirms that discussions are largely centered on test results, medical guidance, fetal health, and the emotional weight of screening decisions.



### 3.6 Proposed Models

This section presents the main steps used to build the deep learning pipeline for analyzing NT in first-trimester ultrasound images. After preparing and cleaning the data, the system performs three main tasks checking the quality of the ultrasound image, detecting the NT region and measuring its thickness, and classifying the result as normal or high risk.

Each step is carried out using deep learning models. The proposed workflow is illustrated in Figure 3.30 and will be explained in more detail in the following sections. The previous sections already described the data collection, EDA and preprocessing phases. This section focuses on the proposed models.

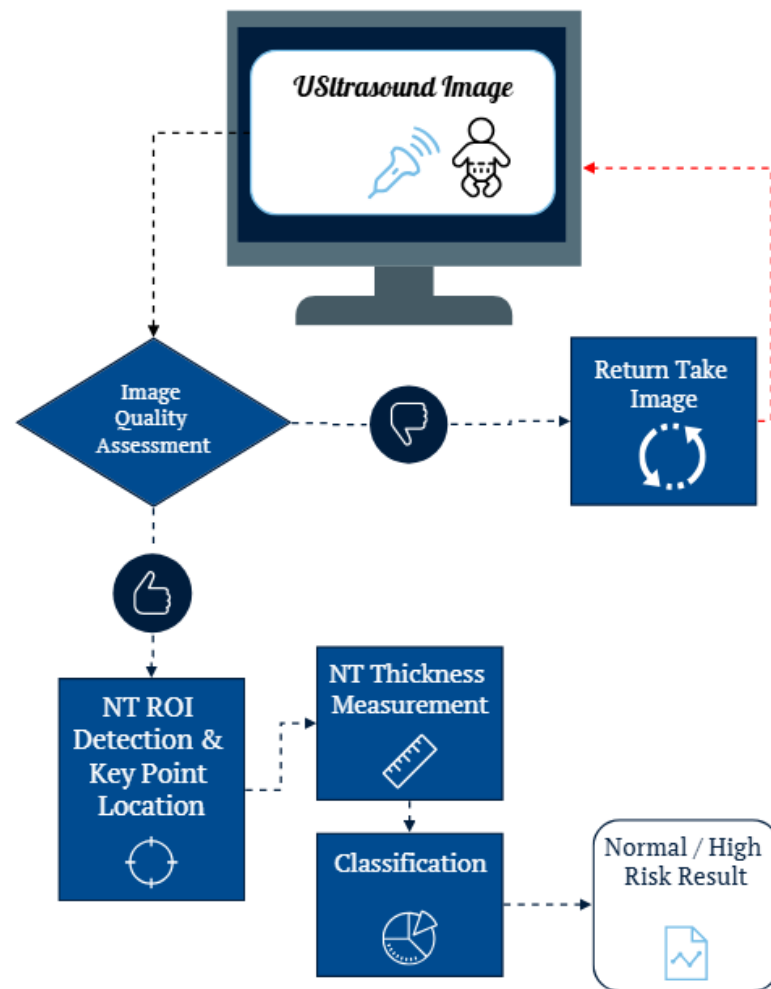


Figure (3.30) The Proposed Workflow for Automated NT Measurement and Classification

### 3.7 Data Split

Data splitting is a critical step in building robust and generalizable machine learning models. By dividing the dataset into distinct subsets, we ensure that model performance evaluates fairly and that the training process does not lead to overfitting.

In this thesis, two datasets split. The first dataset Fetus Framework, which contains 1,684 ultrasound images labeled as standard or non-standard, splits using a structure of 70% for training, 15% for validation, and 15% for testing. The second dataset NHOG, which includes 741 ultrasound images, splits using an 80% for training and 20% for testing.

The training set fits the models and learns relevant features from the data. The validation set, used only with the Fetus Framework dataset, helps tune model hyperparameters and monitor performance during training. The testing set keeps separate for the final evaluation to assess how well the model performs on unseen data.

This split ratio selects based on best practices for medical image modeling, especially when the dataset size is limited but balanced. It ensures enough data is available for learning while still providing reliable metrics during evaluation. The general concept of this data splitting strategy illustrates in Figure 3.31.

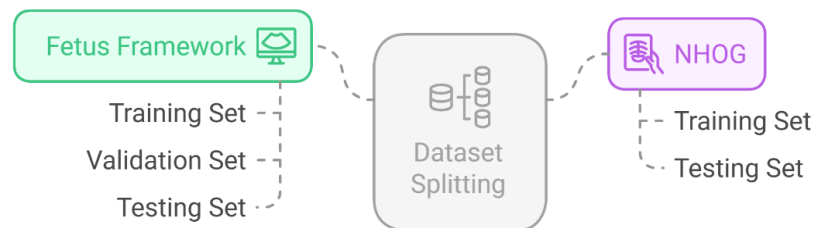


Figure (3.31) Dataset Splitting

### 3.8 Selected Models

In this thesis, a transfer learning approach was followed and Deep Learning pre-trained models were used. Transfer learning is a machine-learning concept that utilizes gained knowledge from a task to enhance performance in another different but related task. The chosen models are DenseNet121 (Huang et al.,

2017), VGG16 (Jain et al., 2023), VGG19 (Hindarto et al., 2023), ResNet18 (Ramzan et al., 2019), ResNet50 (A. Zhang et al., 2020), EfficientNetB7 (Tan & Le, 2019), Inceptionv3 (Szegedy et al., 2015) and U-Net (Zhou et al., 2018). These models were selected based on their performance and evaluation results in prior studies for similar tasks. The following subsections will provide an overview of each model.

### **3.8.1 DenseNet121**

DenseNet121 is a Convolutional Neural Network (CNN) architecture that was introduced by Gao Huang et al. in their 2017 paper, "Densely Connected Convolutional Networks." It is a member of the DenseNet family, which is known for its dense connectivity pattern that improves information flow and parameter efficiency. The model has demonstrated impressive performance on various computer vision tasks, including image classification (Raju et al., 2023).

#### **3.8.1.1 Architectural Design for DenseNet121**

The architectural design of the Dense Convolutional Network (DenseNet) is based on the idea of connecting every layer to every other layer in a feed-forward manner. This is a departure from traditional convolutional networks, which have  $L$  connections for  $L$  layers, as DenseNets have  $L(L+1)/2$  direct connections.

- **Dense Connectivity:** In a DenseNet, each layer receives the feature maps of all preceding layers as input and passes its own feature maps to all subsequent layers.
- **Concatenation:** Instead of combining features through summation, as is done in ResNets, DenseNets combine them by concatenating the feature maps.
- **Dense Blocks and Transition Layers:** To handle changes in feature map sizes, the network is divided into multiple dense blocks. The layers between these blocks are called transition layers, which perform convolution and pooling operations to down-sample the feature maps.
- **Growth Rate (k):** Each layer in a dense block adds a small, fixed number of feature maps, referred to as the growth rate (k). The document shows that a small growth rate is sufficient to achieve state-of-the-art results and that it regulates how much

new information each layer contributes to the network's "collective knowledge" (Huang et al., 2017).

Figure 3.32 illustrates the overall DenseNet architecture, showing the dense connectivity pattern, dense blocks, and transition layers that enable efficient feature reuse and enhanced representational power.

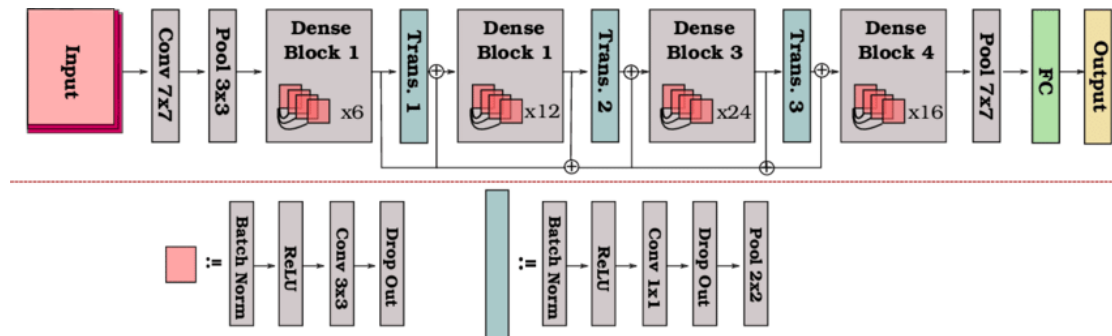


Figure (3.32) Architectural Design of DenseNet121 (Radwan, 2019)

### 3.8.2 VGG16

VGG, or the Visual Geometry Group network, is a deep learning model built on a Convolutional Neural Network (CNN) architecture. It is widely used in image processing tasks like object detection and image classification. The model consists of multiple layers, including convolutional, max pooling, and fully connected layers. The VGG16 model, for example, is composed of 13 convolutional layers and 3 fully connected layers. The following Figure 3.33 is VGG16 Structure (Jain et al., 2023).

#### 3.8.2.1 Architectural Design for VGG16

- **Input Layer:** This layer accepts an input image with dimensions of 224x224 pixels.
- **Convolutional Layers:** VGG16 contains 13 convolutional layers that use 3x3 filters. The number of filters progressively increases in different layer sets, beginning with 64, then 128, 256, and finally 512 filters.
- **Max Pooling Layers:** These layers follow each group of convolutional layers and are used to reduce the spatial dimensions of the feature maps.

- **Fully Connected Layers:** The model has three fully connected layers. The first two layers have an output size of 4096 and utilize the ReLU activation function. The final layer has an output size of 1000 and uses the softmax activation function to produce classification outputs (Jain et al., 2023).

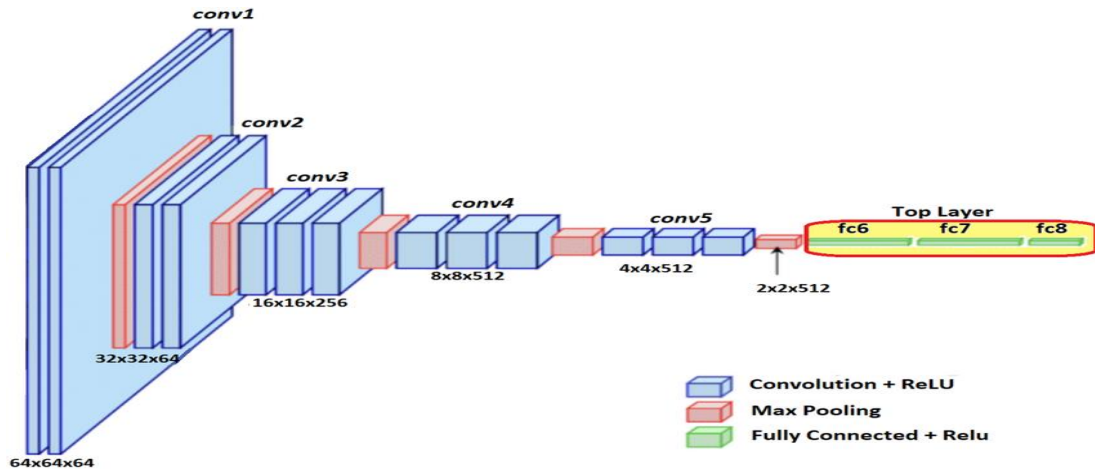


Figure (3.33) VGG16 Architectural (Hajamohideen et al., 2023)

### 3.8.3 VGG19

VGG19 is a deep Convolutional Neural Network (CNN) architecture developed by the Visual Geometry Group at the University of Oxford. This model is notable for its architectural simplicity and its effectiveness in large-scale image recognition tasks. It was a pivotal model that demonstrated the benefits of increased network depth for achieving superior performance.

#### 3.8.3.1 Architectural Design for VGG19

The VGG19 architecture is composed of 19 layers with learnable weights, specifically 16 convolutional layers and 3 fully connected layers. The network is designed to accept a fixed-size input of 224x224-pixel RGB images. The core design principles are based on a consistent, uniform structure:

- **Small Convolutional Filters:** The architecture exclusively employs small 3x3 convolutional filters with a stride of 1. This strategic choice allows the network to capture intricate, multi-scale features while keeping the number of parameters manageable compared to models using larger filters.
- **Max-Pooling Layers:** The network is organized into five blocks of convolutional layers, with each block concluded by a max-pooling layer with a 2x2 filter and a

stride of 2. This down-sampling process progressively reduces the spatial dimensions of the feature maps, which aids in computational efficiency and enhances the model's ability to learn robust features.

- **ReLU Activation:** The Rectified Linear Unit (ReLU) is applied as the activation function after each convolutional layer to introduce non-linearity, which is essential for learning complex patterns.
- **Deep and Uniform Structure:** The repetitive stacking of small convolutional filters creates a very deep architecture. This depth is critical as it enables the network to learn a rich hierarchy of features, from basic edges in early layers to complex object representations in deeper layers.
- **Fully Connected Classifier:** The convolutional feature extraction blocks are followed by three fully connected layers, which serve as the final classifier. The first two fully connected layers each have 4096 neurons, and the final layer contains 1000 neurons, corresponding to the 1000 classes of the ImageNet challenge. A softmax activation is applied to this final layer to produce class probabilities, As illustrated in Figure 3.34 (Hindarto et al., 2023).

VGG19's profound depth and architectural elegance solidified the notion that increasing the number of layers in a CNN can significantly boost performance, an insight that has influenced the design of subsequent advanced architectures.

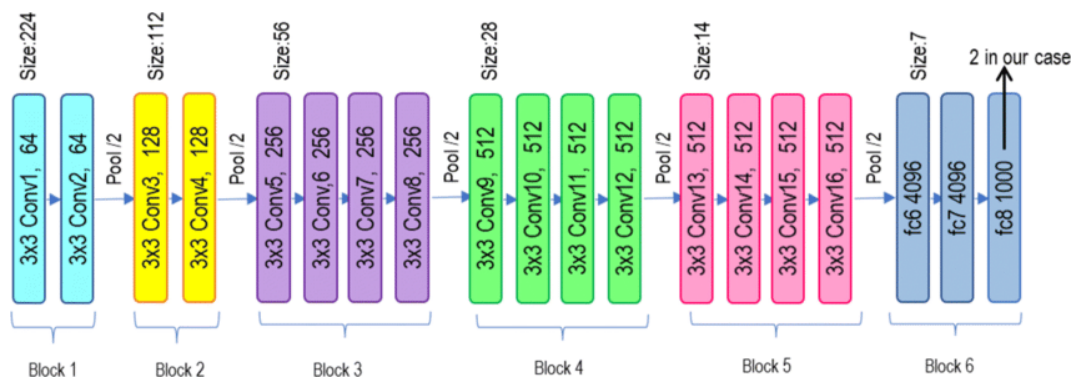


Figure (3.34) VGG19 Architectural (Khattar & Quadri, 2022)

### 3.8.4 ResNet-18

ResNet-18 is a foundational deep learning architecture from the family of Deep Residual Networks (ResNets), introduced by He et al. (2016). Its primary innovation lies in the use of residual blocks. The fundamental problem it addresses

is the degradation problem, where simply adding more layers to a deep Convolutional Neural Network (CNN) can lead to a decrease in performance due to optimization difficulties, not overfitting.

### 3.8.4.1 Architectural Design for ResNet-18

ResNet-18 is a relatively lightweight model with 18 trainable layers, making it a good choice for tasks where computational resources are limited or for use as a backbone in more complex architectures. Its structure is composed of:

- Initial Layers: A large 7x7 convolutional layer with a stride of 2, followed by a max-pooling layer.
- Four Stages: The network is divided into four main stages, each consisting of two basic residual blocks. Each stage progressively increases the number of feature maps while halving the spatial dimensions.
  - Stage 1: 2 blocks, 64 filters.
  - Stage 2: 2 blocks, 128 filters.
  - Stage 3: 2 blocks, 256 filters.
  - Stage 4: 2 blocks, 512 filters.
- Final Layers: A global average pooling layer and a final fully connected layer for classification.

The basic residual block in ResNet-18 consists of two 3x3 convolutional layers, each followed by Batch Normalization and a Rectified Linear Unit (ReLU) activation function, before the skip connection is added, As shown in Figure 3.35 (Ramzan et al., 2019).

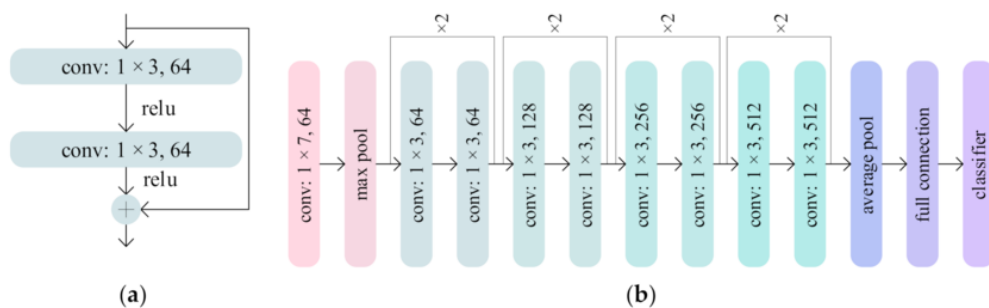


Figure (3.35) ResNet-18 Architectural (Y. Zhao et al., 2022)

### **3.8.5 Resnet 50**

The ResNet-50 architecture, a key advancement in the field of deep learning, was introduced to solve the critical challenges associated with training very deep Convolutional Neural Networks (CNNs). As networks grew deeper to enhance representational capacity, they faced two primary issues: the degradation problem and the vanishing/exploding gradient problem. The degradation problem refers to the phenomenon where a deeper network exhibits higher training error than its shallower counterpart, indicating an optimization difficulty rather than overfitting. The vanishing/exploding gradient problem, on the other hand, is a numerical instability issue during backpropagation, where gradients become infinitesimally small or excessively large, respectively, making it difficult for the network to learn. ResNet-50 mitigates these problems through its core innovation: the residual block and skip connections.

#### **3.8.5.1 Architectural Design for Resnet 50**

The ResNet-50 architecture is a sophisticated deep learning model designed to overcome the training limitations of very deep neural networks. Its primary innovation is the use of residual blocks that employ skip connections. This structure allows the network to learn a residual function, rather than an unconstrained mapping, ensuring that adding more layers does not degrade performance.

The network's structure is modular, consisting of 50 layers arranged in a hierarchical fashion. It begins with an initial convolutional and max-pooling stage for downsampling the input image. This is followed by four distinct stages, each comprising a series of bottleneck residual blocks. This specific block design uses 1x1 convolutions to efficiently reduce and then restore dimensionality, which significantly lowers the number of parameters and computational complexity. The skip connections within these blocks serve as direct paths for gradients, preventing them from vanishing during backpropagation and ensuring stable training. The final part of the network includes a global average pooling layer and a fully connected layer for classification, As shown in Figure 3.36 (A. Zhang et al., 2020).

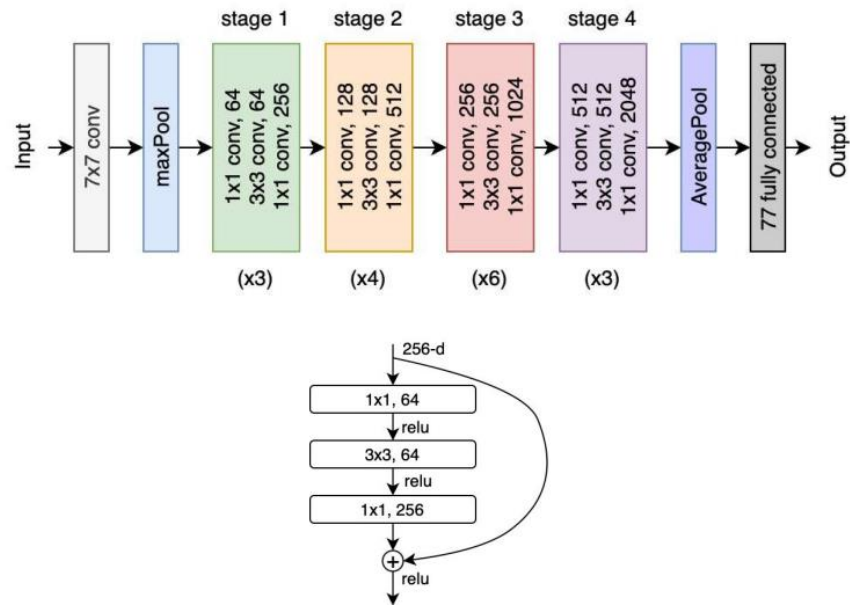


Figure (3.36) ResNet-50 Architectural (FATURRAHMAN et al., 2023)

### 3.8.6 EfficientNetB7

The Efficient Nets are a family of eight CNN models, from EfficientNetB0 to EfficientNetB7, which are known for their superior efficiency and accuracy compared to prior models. They were developed by using a novel scaling method called compound scaling. This method uniformly scales the network's depth (number of layers), width (number of channels), and resolution (image size) using a single compound coefficient. This approach allows for enhanced performance with fewer parameters (Tan & Le, 2019).

#### 3.8.6.1 Architectural Design for EfficientNetB7

The base architecture for these models is built upon the MBConv block, which is a combination of a squeeze-and-excitation (SE) block and an inverted residual block. The overall architecture of EfficientNetB7, which achieved the highest top-1 accuracy on ImageNet among the family, consists of seven such blocks as shown in Figure 3.37, with the number of filters varying between layers (Tan & Le, 2019).

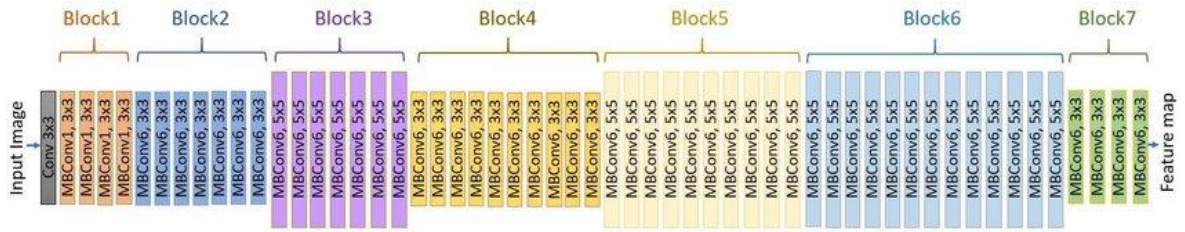


Figure (3.37) EfficientNetB7 Architectural (Baheti et al. 2020)

### 3.8.7 Inception V3

Inception v3 is a deep convolutional neural network architecture developed by Google as an evolution of the Inception family of models. It's designed to improve upon the previous versions (Inception v1 and v2) by introducing a more computationally efficient and scalable design. The central concept behind Inception v3 is to **factorize convolutions** into smaller, more manageable operations. This reduces the number of parameters and computational cost without sacrificing network performance.

The key factorization techniques include:

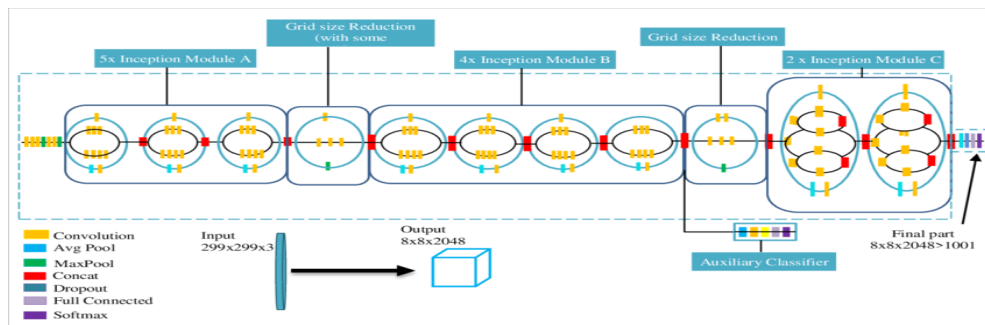
- Factorizing a 7x7 convolution into a sequence of a 1x7 and a 7x1 convolution. This significantly reduces the number of parameters. For example, a 3x3 convolution with 128 filters requires  $3 \times 3 \times 128 \times 128$  parameters, while a 1x3 followed by a 3x1 convolution requires only  $1 \times 3 \times 128 \times 128 + 3 \times 1 \times 128 \times 128$  parameters, which is a substantial saving.
- Factorizing an nxn convolution into a 1xn followed by an nx1 convolution. This approach is known as asymmetric convolution.
- Factorizing larger convolutions into smaller ones. For example, a 3x3 convolution can be replaced by a 1x1 convolution followed by a 3x3 convolution.

These factorizations are applied extensively within the network's core building blocks, known as Inception modules.

#### 3.8.7.1 Architectural Design for Inception V3

Inception v3's architecture is a composite of several interconnected Inception modules, as shown in Figure 3.38 The network consists of a stem (the initial layers), followed by a sequence of different Inception modules and a classifier.

- **Stem:** The initial part of the network, which handles the first few layers of feature extraction.
- **Inception Modules:** The network is built from multiple types of these modules (Type A, B, C, D, E). Each module processes the input feature map in parallel through a series of different convolutional layers (e.g., 1x1, 3x3, 5x5) and pooling operations before concatenating their outputs. This allows the network to capture patterns at various scales simultaneously.
- **Auxiliary Classifiers:** A notable feature of Inception v3 is the use of auxiliary classifiers. These are small, secondary classifiers attached to intermediate layers of the network. They are trained alongside the main classifier and help combat the vanishing gradient problem by providing additional supervision at lower layers. This encourages earlier layers to produce more meaningful representations, which is crucial in very deep networks.
- **Batch Normalization:** The model makes extensive use of batch normalization layers, which normalize the activations of the previous layer. This helps in regularizing the network and accelerating the training process by reducing internal covariate shift (Szegedy et al., 2015).



Figure(3.38) Inception V3 Architectural (Iparraguirre-Villanueva et al., 2022)

### 3.8.8 U-Net

U-Net is a fully convolutional encoder–decoder architecture introduced by Ronneberger et al. to perform pixel-wise image segmentation in biomedical images (Ronneberger et al., 2015). Its defining feature is a symmetric “U” shape composed of a contracting (encoder) path that progressively downsamples to learn

rich semantic features and an expansive (decoder) path that upsamples to recover spatial resolution, as shown in Figure 3.39. Crucially, skip connections concatenate feature maps from encoder layers to their decoder counterparts, restoring fine spatial detail lost during downsampling and producing segmentation maps that are spatially aligned with the input image. This design directly addresses the localization limitations of conventional CNNs and generic encoder–decoder models by enforcing per-pixel outputs at (or near) the original resolution while preserving contextual semantics. Since its introduction, U-Net has become a standard baseline in medical image analysis and has inspired numerous variants (e.g., attention mechanisms, deeper backbones, and task-specific adaptations) due to its strong performance with limited annotated data (Çiçek et al., 2016)(Zhou et al., 2018).

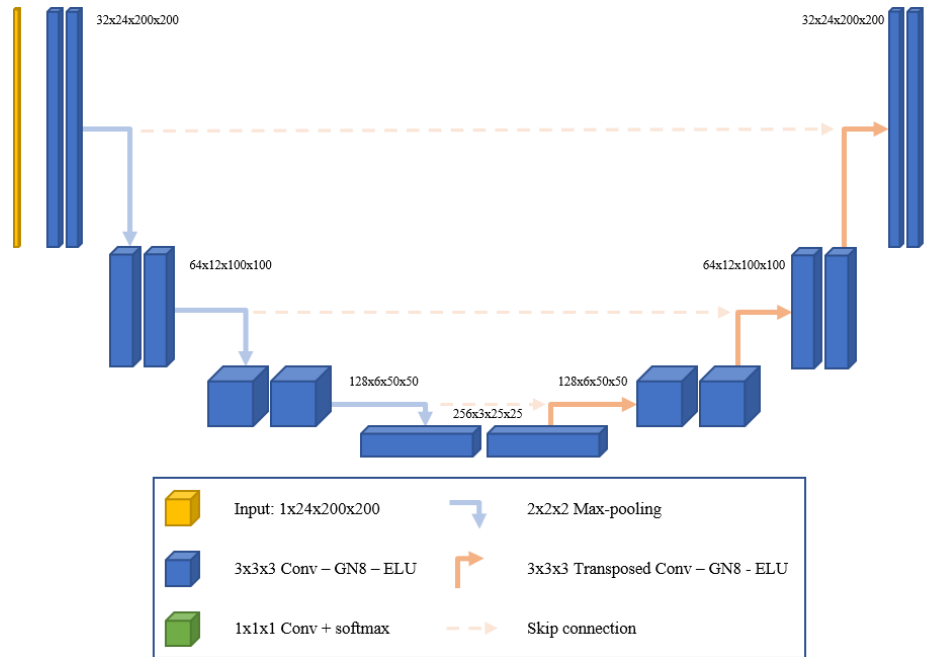


Figure (3.39) The U-Net Architecture (Schmidt, 2020)

### 3.8.8.1 Architectural Design for U-Net

The U-Net is composed of two main components: a contracting path and an expanding path.

- **Contracting path:** systematically reduces spatial dimensions while extracting hierarchical feature representations and capturing contextual information about the input image.

- **Expanding path:** reconstructs spatial resolution through upsampling operations and generates precise segmentation maps by integrating learned patterns from the contracting path with preserved spatial details.

### **The Contracting (Encoder)**

- **Encoder Stage 1**

1. An input image with dimensions  $572 \times 572$  is fed into the U-Net. This input image consists of a single grayscale channel, typical of biomedical imaging modalities such as electron microscopy.
2. Two  $3 \times 3$  convolutional layers (unpadded) are applied to the input image, each followed by a ReLU activation function. The number of feature channels is increased to 64 to capture higher-level feature representations and establish the foundational feature space.
3. A  $2 \times 2$  max pooling layer with stride 2 is applied, downsampling the feature map to half its spatial dimensions, resulting in  $284 \times 284$  resolution.

- **Encoder Stage 2**

1. Following the established pattern from Block 1, two  $3 \times 3$  convolutional layers (unpadded) are applied to the output of Block 1, each followed by a ReLU activation function. At each successive block, the number of feature channels is doubled, increasing to 128 to capture more complex and abstract feature representations.
2. A  $2 \times 2$  max pooling layer is applied to the resulting feature map, reducing the spatial dimensions by half to  $140 \times 140$ .

- **Encoder Stage 3**

The architectural procedure established in Block 1 and Block 2 is replicated in Block 3. Two  $3 \times 3$  unpadded convolutions with ReLU activations double the channels to 256, followed by  $2 \times 2$  max pooling that reduces spatial dimensions to  $68 \times 68$ .

- **Encoder Stage 4**

Same architectural pattern as Block 3. The feature channels are doubled to 512 through dual convolutions with ReLU activations, and spatial dimensions are reduced to  $32 \times 32$  via max pooling.

- **Encoder Stage 5**

1. In the final block of the contracting path, the number of feature channels reaches 1024 after being doubled at each successive block, representing the deepest and most abstract feature representation.
2. This block contains two  $3 \times 3$  convolutional layers (unpadded), each followed by a ReLU activation function. However, for architectural symmetry purposes, only one convolution is included in this block, with the second convolution incorporated into the expanding path to maintain balanced processing between encoder and decoder components, as shown in Figure 3.40.

Following the extraction of high-level features in the contracting path, the learned representation enters the expanding path, where transposed convolutions and skip connections progressively restore spatial resolution to produce a pixel-aligned segmentation map (Ronneberger et al., 2015).

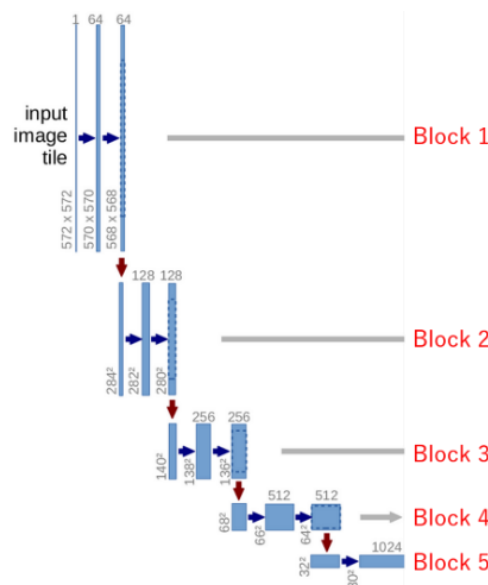


Figure (3.40) The Contracting Path of the U-Net (Oğuz & Ertuğrul, 2023)

## **Expanding Path (Decoder)**

The expanding path utilizes both standard convolution and transposed convolution (up-convolution) operations to integrate learned features and progressively upsample the feature map until generating a high-resolution segmentation map.

### **The Expanding Path of the U-Net**

- **Decoder Stage 5**

1. Continuing from the contracting path, a second  $3\times 3$  convolution (unpadded) is applied with a ReLU activation function, completing the dual-convolution pattern established throughout the architecture.
2. A  $2\times 2$  transposed convolution (up-convolution) layer is applied, upsampling the spatial dimensions twofold from  $28\times 28$  to  $56\times 56$  while simultaneously halving the number of channels to 512.

- **Decoder Stage 4**

1. Utilizing skip connections, the corresponding feature map from the contracting path is concatenated with the upsampled features, doubling the feature channels to 1024. This concatenation requires center cropping of the contracting path features to match the expanding path's spatial dimensions due to boundary pixel loss from unpadded convolutions.
2. Two  $3\times 3$  convolutional layers (unpadded) are applied, each followed by ReLU activations, reducing the channels to 512 while learning to integrate multi-scale feature information.
3. Subsequently, a  $2\times 2$  transposed convolution (up-convolution) layer is applied, upsampling the spatial dimensions twofold and halving the number of channels to 256.

- **Decoder Stage 3**

The processing procedure established in Block 5 and Block 4 is replicated in Block 3. Skip connection concatenation creates 512 channels, dual convolutions reduce to 256 channels, and up-convolution produces  $200\times 200\times 128$  feature maps.

- **Decoder Stage 2**

Same architectural pattern as Block 3. Skip connection concatenation yields 256 channels, dual convolutions reduce to 128 channels, and up-convolution generates  $392 \times 392 \times 64$  feature maps.

- **Decoder Stage 1**

1. In the final block of the expanding path, there are 128 channels after concatenating the skip connection from the corresponding contracting path layer with the upsampled features.
2. Two  $3 \times 3$  convolutional layers (unpadded) are applied to the feature map, with ReLU activations between them, reducing the number of feature channels to 64 while refining the integrated multi-scale representation.
3. Finally, a  $1 \times 1$  convolutional layer followed by an activation function (sigmoid for binary classification) maps the 64-channel feature space to the desired number of output classes. In this configuration, 2 classes are produced, as binary classification is commonly employed in medical imaging applications for distinguishing between target structures and background, as shown in Figure 3.41.

After progressive upsampling and feature integration throughout the expanding path, a high-resolution segmentation map is generated, with each pixel classified individually according to the learned feature representations and spatial relationships (Ronneberger et al., 2015).

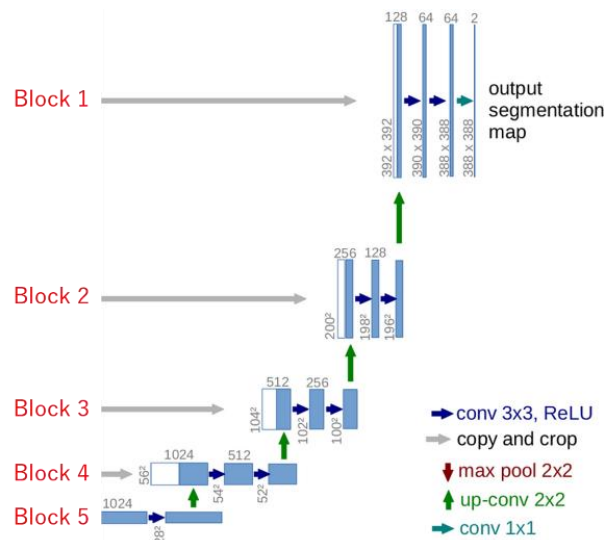


Figure (3.41) The Expanding Path of the U-Net (Oğuz & Ertuğrul, 2023)

### 3.9 Segmentation

Image segmentation is a fundamental process in computer vision and medical image analysis, involving the partitioning of an image into distinct, meaningful, and non-overlapping regions to facilitate object recognition and scene understanding (Yu et al., 2023).

#### 3.9.1 Artificial Neural Network (ANN)-Based Segmentation

ANN-based methods emulate cognitive decision-making processes to perform feature extraction and segmentation. These models consist of interconnected neurons (nodes) with adjustable weights and can learn complex mappings. (Kaur & Kaur, 2014).

#### 3.9.2 Types of Image Segmentation

There are many types of image segmentation techniques, each serving different analytical purposes. In this section, the focus is placed on instance segmentation, as it provides the precision needed to isolate the nuchal translucency region accurately.

- **Instance segmentation**

Instance segmentation is an advanced task in computer vision that performs pixel-level classification while distinguishing between individual object instances, even within the same semantic class. It enables precise delineation and identification of each object in a scene, combining the strengths of object detection and semantic segmentation to produce fine-grained, instance-aware representations of visual data (Tian et al., 2021).

### 3.10 Model Parameters

Model parameters are internal to the model and are learned directly from the training data as depicted in the provided Figure 3.42. The most common examples are weights and biases for Neural Networks.

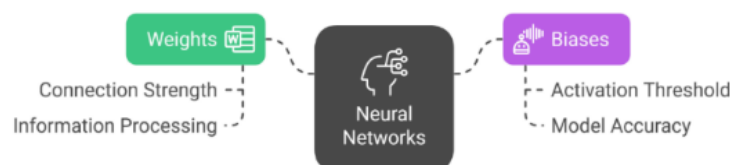


Figure (3.42) Model Parameters

### 3.11 Models hyperparameters

Hyperparameters are the external to the model and are configured manually before the training process begins. These parameters control the learning process itself and are adjusted to develop an optimal model. There are certain types of hyperparameters such as learning rate, data split ratio, number of epochs, number of branches, optimizer, etc. Show in Figure 3.43.

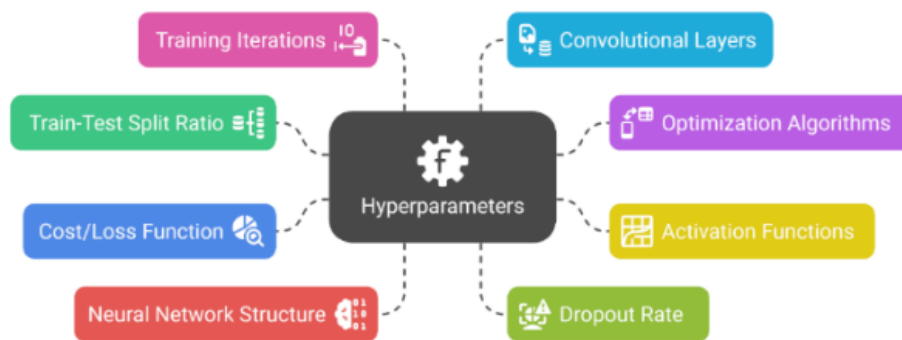


Figure (3.43) Model Hyperparameters

The process of hyperparameter tuning is crucial because selecting the wrong values can make a model ineffective and inaccurate. Therefore, it is essential to adjust these parameters according to the specific dataset being used to ensure the model provides a valid and effective prediction (Prajapati and Singh, 2023).

### 3.12 Regularization

Regularization is a technique used during training to improve a model's generalizability by preventing both overfitting—when the model fits the training data too closely, including noise, and thus performs poorly on unseen data—and underfitting—when the model is too simple or insufficiently trained to capture the true relationship between inputs and outputs; in practice, regularization often operates by adding a penalty to model parameters (e.g., weight decay) or otherwise constraining learning so the model finds a simpler, more robust solution. see Figure 3.44 for an illustrative best-fit line under regularization (Mohr et al., 2022).

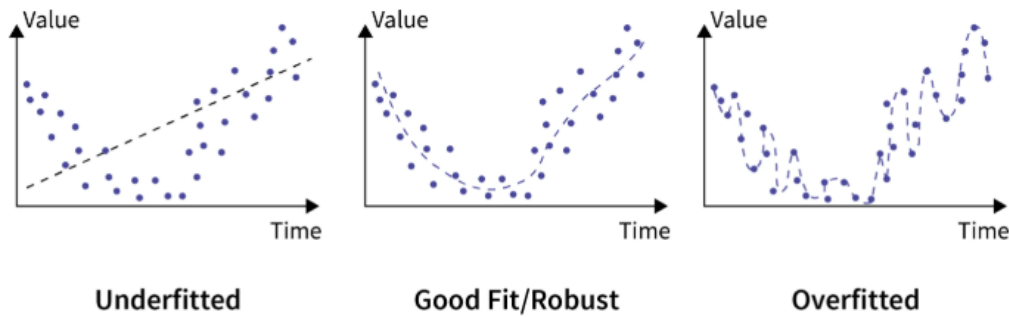


Figure (3.44) Regularization (Hu, 2024)

### ➤ Underfitting

Occurs when a machine learning model is trained with an insufficient amount of data or features, causing it to perform poorly and produce inaccurate results. This issue is typically characterized by high bias and low variance during the training process (Prajapati & Singh, 2023).

### ➤ Overfitting

Occurs when a machine learning model, despite achieving high accuracy on its training dataset, fails to generalize to new, unseen data, resulting in inaccurate predictions. This phenomenon arises when the model learns to fit every data point in the training set, including noise and irrelevant information (Prajapati & Singh, 2023).

### ➤ Good fitting

Occurs when machine learning model that has an optimal balance between bias and variance. This allows it to accurately capture the patterns in the training data without being overly influenced by noise, which is key to its ability to generalize and make accurate predictions on new, unseen data.

### ➤ Bias

Quantifies the error stemming from a model's simplifying assumptions, which can cause it to miss the true underlying relationship between features and the target variable. A model with high bias is too simple for the data and is said to be "underfitting," as it systematically makes inaccurate predictions (Doroudi, 2020).

### ➤ Variance

Measures the model's sensitivity to fluctuations in the training data. A model with high variance is overly complex and learns random noise in the data rather than the true pattern. This leads to "overfitting," where the model performs exceptionally well on the training set but poorly on unseen data.

The goal is to achieve a balance between these two errors. As model complexity increases, bias typically decreases while variance increases. This relationship is depicted in the bias-variance tradeoff, see the Figure 3.45 variance trade-off of classical Machine Learning and Deep Learning (Doroudi, 2020).

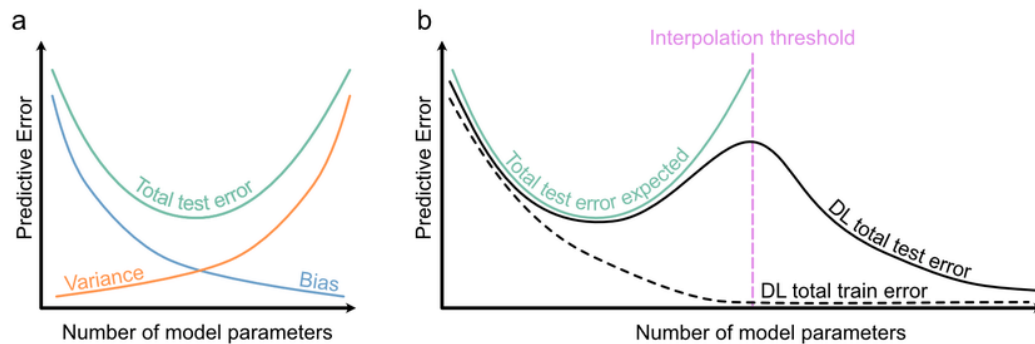


Figure (3.45) Bias-Variance (Pichler & Hartig, 2022)

There are two primary types of regularization: **Lasso** (L1) and **Ridge** (L2) As Figure 3.46.

- **L1 regularization**, or Lasso, works by adding a penalty term to the loss function equal to the absolute value of the sum of the coefficients. This technique is designed to minimize high-value, correlated coefficients by penalizing them, and it can even reduce some coefficients to zero, effectively performing feature selection.
- **L2 regularization**, or Ridge, adds the squared magnitude of the coefficients as a penalty to the loss function. It aims to reduce the size of the coefficients without forcing them to zero. This thesis uses **L2 regularization** on the selected models due to its effectiveness as presented in state-of-the-art research.

In general, regularization methods are used to prevent overfitting or underfitting, balance bias and variance, and control model complexity to improve generalization. Equations 1 and 2 illustrate the formulas for L1 and L2, respectively:

- **L1 Regularization:**

$$\text{Modified Loss function} = \text{Loss function} + \lambda \sum_{i=1}^n |W_i| \quad (1)$$

Where:

- $\lambda$  is the regularization parameter, controlling the strength of regularization.
- $w_i$  represents the coefficients of the model.

- **L2 Regularization:**

$$\text{Modified Loss function} = \text{Loss function} + \lambda \sum_{i=1}^n W_i^2 \quad (2)$$

Where:

- $\lambda$  is the regularization parameter, controlling the strength of regularization.
- $w_i$  represents the coefficients of the model.

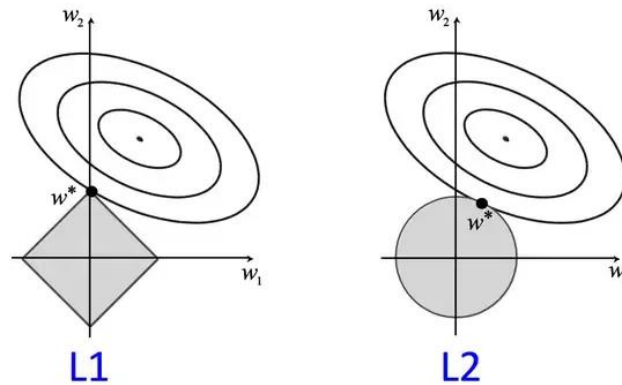


Figure (3.46) Types of Regularization in Machine Learning (Koh, 2024)

### 3.13 Activation Functions

An activation function is a mathematical function used in a neural network that determines the output of a neuron. Its primary role is to introduce non-linearity into the network, allowing it to learn and model complex, non-linear

relationships in data. Without activation functions, a neural network would only be able to perform linear transformations, regardless of how many layers it has (Goodfellow et al., 2016).

### 3.13.1 Purpose and Importance

The core purpose of an activation function is to transform the weighted sum of inputs from a neuron into a usable output value. This transformation is what makes a neural network so powerful.

- **Non-linearity:** This is the most crucial role. The composition of multiple linear functions is still a linear function. By introducing a non-linear activation function, a neural network can approximate any continuous function, a property known as the Universal Approximation Theorem (Hornik et al., 1989). This allows the network to learn intricate patterns in the data that simple linear models cannot.
- **Decision Making:** An activation function acts as a "gate," determining whether a neuron should "fire" or not. It takes the input signal and decides the strength of the signal that should be passed on to the next layer.

### 3.13.2 Common Types of Activation Functions

There are several types of activation functions, each with unique characteristics that make them suitable for different tasks (Glorot & Bengio, 2010).

- **Sigmoid**

The sigmoid function maps any real-valued input to a value between 0 and 1. It is often used in the output layer for binary classification problems where the output represents a probability (Goodfellow et al., 2016). Figure 3.47 and Equation (3) illustrate the ReLU function graphically and mathematically. (Nair & Hinton, 2010).

**Formula:**

$$\text{Sigmoid function } y = \frac{1}{1+e^{-x}} \quad (3)$$

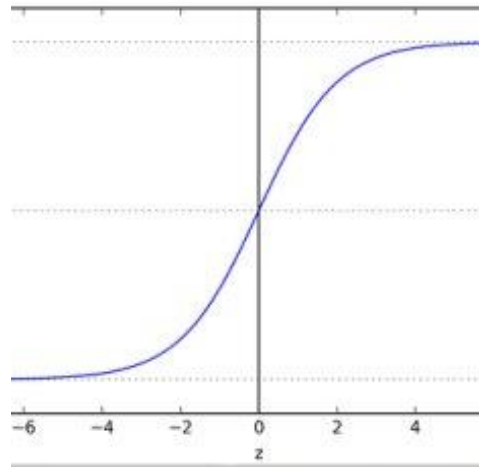


Figure (3.47) Sigmoid Activation Function (Pillalamarry et al., 2021)

- **ReLU (Rectified Linear Unit)**

The Rectified linear unit. It's commonly used in classification tasks. In this function, the neurons are not activated all at once. That is, the output will be equal to the input if it is positive, otherwise, it converts the output to zero so the neurons will not get activated. Figure 3.48 and Equation (4) illustrate the ReLU function graphically and mathematically. (Nair & Hinton, 2010).

**Formula:**

$$\text{ReLU function } f(x) = \max(0, x) \quad (4)$$

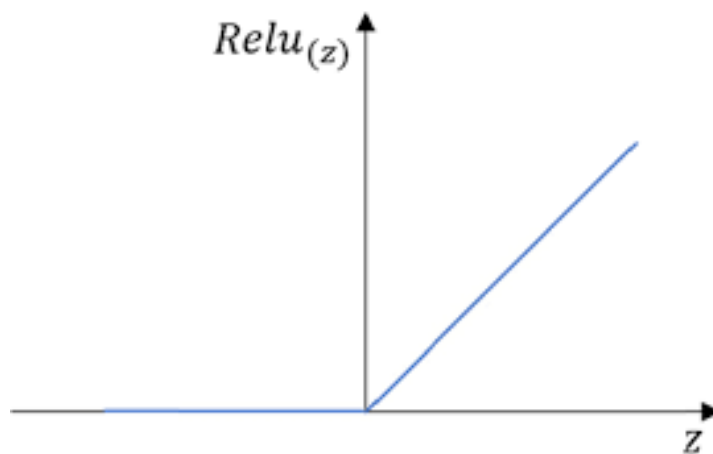


Figure (3.48) ReLU Activation Function (J. Liu & Zou, 2022)

- **Softmax**

Softmax is typically used in the output layer of a network for multi-class classification problems. It takes a vector of raw scores (logits) and converts them into a probability distribution where the sum of all probabilities is 1. Figure 3.49 and Equation (5) illustrate the softmax function graphically and mathematically. (Goodfellow et al., 2016).

**Formula:**

$$\text{Softmax function} = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (5)$$

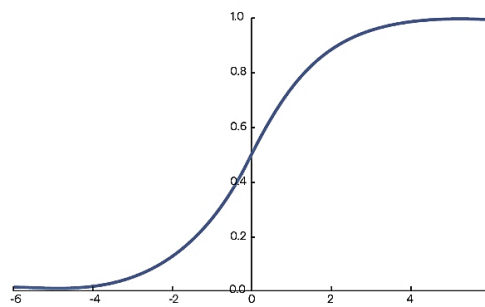


Figure (3.49) SoftMax Activation Function (Ting, 2023)

### 3.14 Optimizer

An optimizer is an algorithm used to update the weights and biases of a neural network in order to minimize the loss function. The goal is to find the optimal set of parameters that results in the most accurate model.

#### 3.14.1 Optimizers Work

In a neural network, the optimizer's job is to figure out how to adjust the network's parameters (weights and biases) during training. This process is driven by the loss function, which measures how far the network's predictions are from the actual target values.

The optimizer uses the loss function to calculate the gradient, which is a vector of partial derivatives. The gradient indicates the direction and magnitude of the steepest increase in the loss function. The optimizer then adjusts the parameters in the opposite direction of the gradient to decrease the loss. This process is known as backpropagation (Goodfellow et al., 2016).

This can be visualized as a hiker in a hilly landscape (the loss function). The hiker's goal is to find the lowest point (the minimum loss). At each step, they look at the slope (the gradient) and take a step in the steepest downward direction. The optimizer controls the size of these steps, known as the learning rate.

### 3.14.2 Common Types of Optimizers

Optimizers have evolved to address various challenges in training, such as the speed of convergence and the risk of getting stuck in local minima.

- **Gradient Descent (GD):** This is the most basic optimizer. It calculates the gradient using the entire dataset. While it guarantees finding the minimum, it can be very slow for large datasets (Goodfellow et al., 2016).
- **Stochastic Gradient Descent (SGD):** To speed things up, SGD calculates the gradient using only a single randomly selected data point at a time. This makes the updates much faster but also more noisy and less stable (Bottou, 2010).
- **Mini-batch Gradient Descent:** This is a compromise between GD and SGD. It calculates the gradient using a small, randomly selected subset of the data (a "mini-batch"). This method is widely used because it combines the speed of SGD with the stability of GD (Goodfellow et al., 2016).
- **Adam (Adaptive Moment Estimation):** Adam is a very popular and effective optimizer. It adapts the learning rate for each parameter individually based on past gradients. It combines the benefits of other adaptive learning rate optimizers, making it fast, efficient, and robust. (Kingma & Ba, 2014).

### 3.15 Epoch

In the context of artificial neural networks, an epoch is defined as the full presentation of the entire set of training vectors to the network, after which the network's weights and biases are updated. The number of epochs is a critical parameter because choosing an optimal value is essential to prevent both undertraining and overtraining of the network. If the number of epochs is too low, the network will not be fully trained, and if it is too high, it can lead to overfitting, where the network performs well on the training data but poorly on new data (Sinha et al., 2010).

### 3.16 Batch Size

Batch size refers to the number of training samples processed simultaneously before the model's weights are updated. It is considered a key hyperparameter that directly influences the learning rate, model accuracy, and overall performance. Numerous studies have explored the impact of batch size on aspects such as generalization, training stability, and computational efficiency. While many researchers suggest using small batch sizes, typically ranging from 2 to 64, some studies argue against this approach. In practice, batch size is often determined empirically, with values commonly selected between 2 and 128 through trial and error. Nevertheless, it remains unclear whether these commonly recommended ranges truly optimize the predictive performance of deep learning models, and there is still no definitive guideline to ensure that the selected batch size is ideal for a given task (Hwang et al., 2024).

### 3.17 Iteration

Refers to the number of times a batch of data is passed through the learning algorithm during training. In the context of neural networks, this includes both the forward pass, where inputs propagate through the network to generate predictions, and the backward pass, where gradients are computed and weights are updated. Thus, each complete cycle of processing a single batch of data constitutes one iteration. Figure 3.50 illustrate the difference between epochs, batch size and iteration (El Baba, 2021).

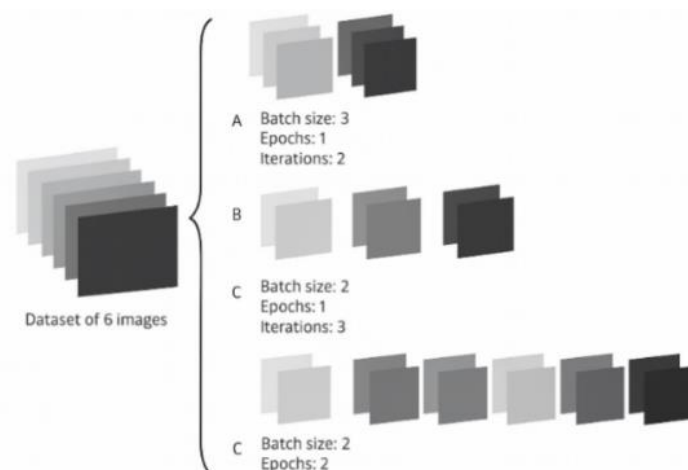


Figure (3.50) The Difference Between Epochs, Batch Size and Iteration  
(El Baba, 2021)

### 3.18 Dropout Rate

Dropout is a technique used in neural networks where individual neurons are randomly deactivated during training with a certain probability, known as the drop rate ( $p$ ). This prevents neurons from co-adapting and is traditionally used as a regularizer to prevent overfitting, where a model performs well on training data but poorly on new, unseen data. The document, however, demonstrates that dropout can also be used to mitigate underfitting.

According to the study, during the initial phases of training, dropout helps reduce the directional variance of gradients across mini-batches and aligns the mini-batch gradients more closely with the overall dataset's gradient. This effect helps the model optimize the training loss more effectively and can be leveraged to address underfitting. The paper introduces two specific approaches based on this insight:

- **Early dropout:** Dropout is applied only during the initial phase of training and then turned off. This is intended to help underfitting models fit the data better and achieve a lower final training loss.
- **Late dropout:** This is a symmetric approach where dropout is not used in the early iterations but is activated later in training. This method is proposed to help improve the generalization performance of models that are prone to overfitting.

The optimal drop rate depends heavily on the size of the model and the dataset. The paper notes that with a large amount of data or a small model, the optimal drop rate can be zero, indicating that dropout may not be necessary and could even be detrimental. The trend over the years has been a decrease in the default drop rate used in many models, largely due to the increasing size of available training datasets (Z. Liu et al., 2023). Figure 3.51 illustrate the dropout on neural network architecture.

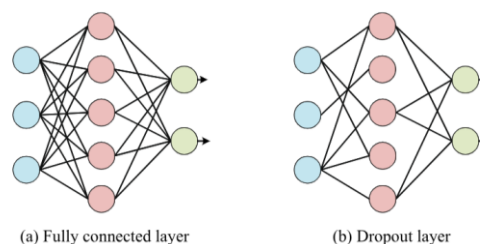


Figure (3.51) Dropout on Neural Network Architecture (X. Zhao et al., 2024)

### 3.19 Loss Function

loss function is a fundamental component of deep learning that quantifies the difference between a model's predicted outputs and the actual ground truth labels. The selection of an appropriate loss function is crucial as it directly influences a model's convergence, generalization, and overall performance across various tasks, such as computer vision, tabular data prediction, and time series forecasting (Elharrouss et al., 2025).

Loss functions are typically categorized based on the task they are used for Loss Functions for Classification and Segmentation.

#### 3.19.1 Classification Loss Functions

The document identifies several loss functions for classification, which is a task that deals with categorical data. The most common ones are variations of cross-entropy:

- Binary Cross-Entropy Loss (BCE): Used for binary classification problems and measures the dissimilarity between a predicted probability and the true class label.
- Weighted Binary Cross-Entropy (WBCE): A variation of BCE that can be used to handle class imbalance.
- Categorical Cross-Entropy Loss (CCE): Used for multi-class classification.
- Focal Loss: This loss function is designed to handle class imbalance by down-weighting the loss contribution from "easy" examples, thereby focusing training on "hard" or difficult examples.
- Hinge Loss: This function is typically used for support vector machines and aims to maximize the margin between classes (Terven et al., 2023).

### 3.19.2 Segmentation Loss Functions

Image segmentation is a more complex task that also utilizes specialized loss functions.

- **Cross-Entropy Loss for Segmentation:** A modified version of cross-entropy is applied at the pixel level to determine the similarity between predicted and true segmentation masks.
- **Intersection Over Union (IoU) Loss:** This loss, also known as the Jaccard index, measures the overlap between the predicted segmentation and the ground truth segmentation.
- **Dice Loss:** Also known as the Dice similarity coefficient, this function is used to evaluate the similarity between the predicted and ground truth segmentation masks. It is especially effective for tasks with significant class imbalance, such as medical image analysis.
- **Tversky Loss:** A generalization of the Dice loss, this function can adjust the balance between false positives and false negatives.
- **Lovász Loss:** This loss is a convex surrogate of the Jaccard index, which makes it particularly useful for optimizing the IoU metric directly (Terven et al., 2023).

### 3.20 Learning Rate

The learning rate is a hyperparameter that controls the step size of an optimizer as it adjusts a neural network's weights during training. It's crucial for efficiently navigating the loss landscape to find a minimum. The value, typically ranging from 1 to  $1 \times 10^{-7}$ , directly influences the speed and stability of convergence.

- **Low Learning Rate:** A low learning rate results in small, cautious updates. While this can lead to a stable training process and a more precise final solution, it significantly increases the time required for the model to converge.
- **High Learning Rate:** A high learning rate accelerates the learning process by taking large steps. However, it risks overshooting the minimum of the loss

function, leading to oscillatory behavior or even divergence, preventing the model from ever reaching an optimal state (Apaydin et al., 2020). Figure 3.52 illustrate three learning curves showing different learning rates.

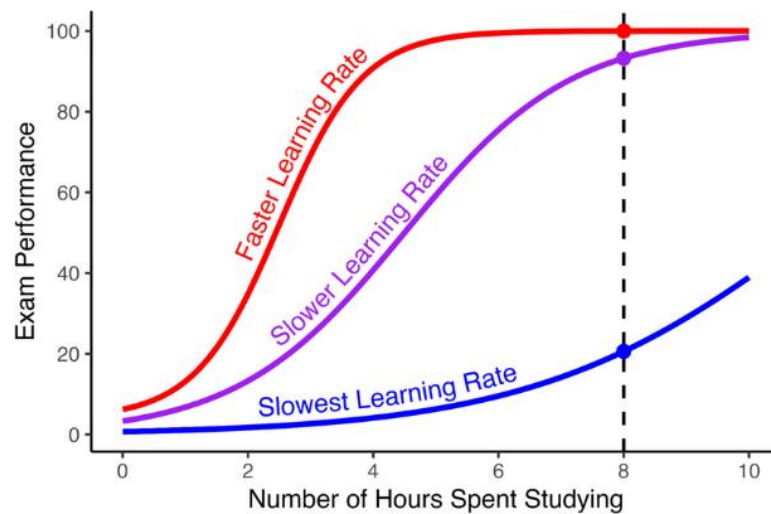


Figure (3.52) Learning Rate (Schuetze, 2024)

### 3.21 Evaluation and Performance Measures

Model performance evaluation is the final stage of the model life cycle. It involves assessing the effectiveness and performance of a trained model on a separate testing subset to ensure it meets its intended goals.

#### 3.21.1 Confusion Matrix

The confusion matrix is a fundamental tool for evaluating the performance of a classification model. It is an  $N \times N$  matrix, where 'N' is the number of target classes, and it summarizes the number of correct and incorrect predictions made by the model.

Here's how it works:

- For binary classification tasks, the confusion matrix provides a clear, summarized table of the model's predictions.
- By examining the diagonal values of the matrix, you can determine the number of accurate classifications. These values represent the cases where the model's prediction matches the actual class.
- The matrix provides insights into four key outcomes:

**TP:** True Positive: The actual value was positive and the model predicted a positive value

**FP:** False Positive: Your prediction is positive, and it is false.

**FN:** False Negative: Your prediction is negative, and result it is also false.

- **TN:** True Negative: The actual value was negative and the model predicted a negative value (Karimi, 2021). Figure 3.53 illustrate Confusion Matrix.

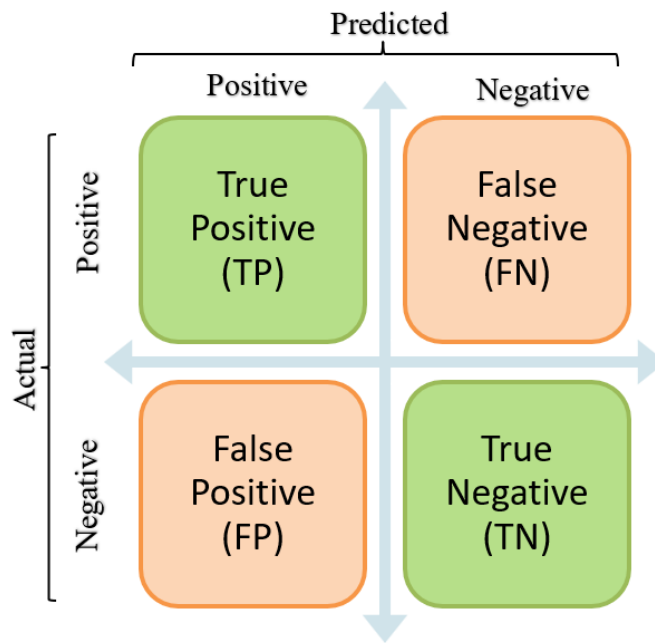


Figure (3.53) Confusion Matrix

Many performance measures are calculated from these values namely accuracy, recall, precision, F1-score, and ROC. These measures are presented in the following sections along with their formulas.

### 3.21.2 Accuracy

Accuracy measures the proportion of correct predictions made by a model across the entire dataset. While it serves as a good basic metric, it can be a poor measure of performance for unbalanced datasets (Karimi, 2021). The accuracy formula is shown in Equation (6) below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (6)$$

### 3.21.3 Recall (Sensitivity)

Recall measures the proportion of actual positive cases that a model correctly identifies. It is calculated as the number of True Positives (TP) divided by the sum of True Positives and False Negatives (FN). Moreover, it represents type-II errors (Swaminathan & Tantri, 2024) . The formula is shown in Equation (7).

$$\text{Recall (or TPR)} = \frac{TP}{TP+FN} \quad (7)$$

### 3.21.4 Precision

Precision is a measure of the model's ability to correctly predict the positive class. It represents the ratio of "true positive" predictions to all positive predictions made by the model. It is particularly valuable in situations where a Type-I error (false positive) is a major concern (Karimi, 2021). The formula is shown in Equation (8).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

### 3.21.5 F1-Score

The F1-score is a metric that combines precision and recall into a single value. It is particularly useful for models where a balance between precision and recall is equally important. The F1-score is calculated as the harmonic mean of precision and recall (Swaminathan & Tantri, 2024) . The formula is shown in Equation (9).

$$F1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

### 3.21.6 ROC – AUC

ROC stands for Receiver Operating Characteristic, and AUC is the Area Under the Curve. ROC is a plot of the True Positive Rate (TPR) against the False Positive Rate (FPR) represented in the x-axis as illustrated in Figure 3.54 below. used to evaluate a classification model's performance across all possible classification thresholds. Originally used for signal detection, this plot is now a fundamental tool in machine learning for showing how a model's performance changes as the classification boundary is adjusted. By plotting TPR and FPR at various thresholds, the ROC curve provides a comprehensive evaluation of a model's ability to distinguish between classes, which is a significant advantage

over single-threshold metrics like accuracy. An ideal ROC curve would have a TPR of 100% and an FPR of 0%, representing a perfect classifier. The curve is particularly useful for selecting an optimal threshold that balances the trade-off between maximizing true positives and minimizing false positives. The latter is calculated using Equation (10) below (Swaminathan & Tantri, 2024).

$$FPR = \frac{FP}{TN+FP} \quad (10)$$

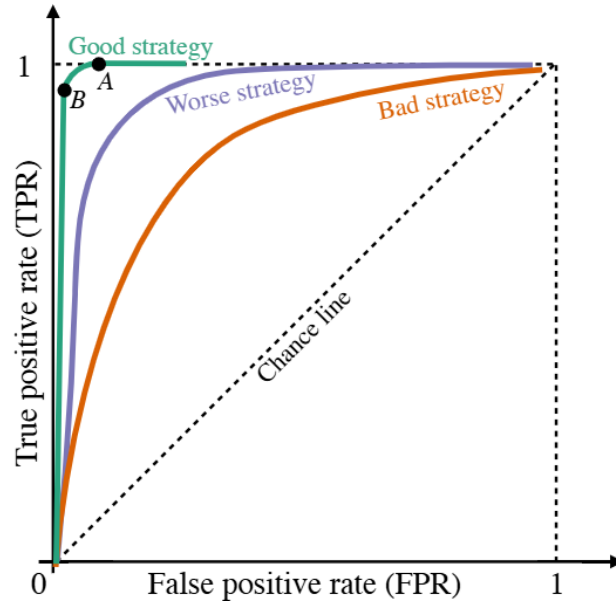


Figure (3.54) ROC Curve (Gwirtz et al., 2020)

### 3.22 Proposed Approaches

This study adopts a multi-stage deep learning pipeline for automating the analysis of Nuchal Translucency (NT) thickness from first-trimester ultrasound images. The proposed framework is structured into five main phases: image quality assessment, region segmentation, feature extraction, NT measurement, and risk classification

#### 3.22.1 First Approach: Standard and Non-Standard Ultrasound Image

Fetal ultrasound images as either **standard or non-standard**, ensuring that only clinically suitable images proceed for NT measurement. It functions as a quality control filter to enhance the reliability of downstream detection and measurement tasks.

The approach was implemented using the Fetus Framework dataset, comprising 1,684 sagittal fetal ultrasound images, with 1,197 standard and 487 non-standard images. To mitigate class imbalance, augmentation techniques such as horizontal flipping, rotation, and brightness adjustment were applied to non-standard images, resulting in a balanced dataset.

The dataset was split into 70% training, 15% validation, and 15% testing, with class-balanced distributions. As part of the preprocessing pipeline, all images were resized and normalized. Speckle noise was suppressed using (BM3D) (Lebrun, 2012).

Deep learning models trained for classification include:

- DenseNet121
- VGG16
- VGG19
- ResNet18
- ResNet50
- EfficientNetB7
- InceptionV3
- **Model Development**

Model fine-tuning involved the optimization of hyperparameters such as learning rate, optimizer type (Adam), activation function (ReLU/Softmax), and class imbalance handling. Cross Entropy was used as the loss function.

- **Evaluation Metrics**

Performance evaluated using standard classification metrics:

- Accuracy
- Precision
- Recall
- F1-score
- AUC

The complete workflow for this stage, from data preparation to classification, is illustrated in Figure 3.55.

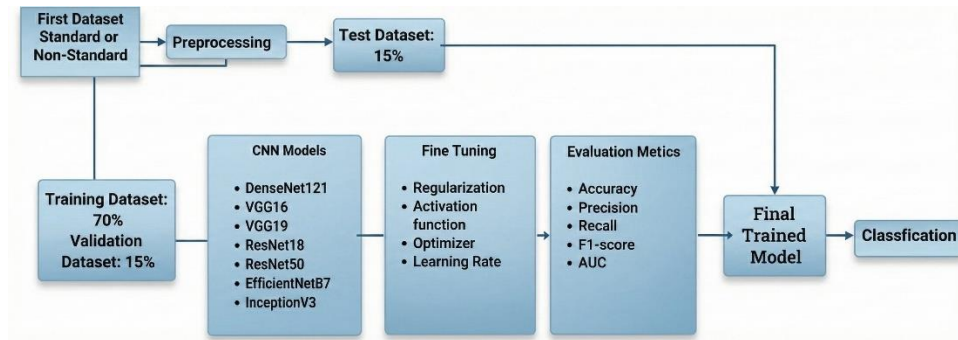


Figure (3.55) First Approach Workflow

### 3.22.2 Second Approach: NT Region Detection, Measurement, and Risk Classification

In this approach, developed a complete deep learning pipeline for automated assessment of Nuchal Translucency (NT) thickness in first-trimester fetal ultrasound images. The goal was to simulate the standard clinical procedure used by sonographers: identify the NT region, measure its thickness in millimeters, and determine whether the fetus is at high risk for chromosomal abnormalities. The method was implemented using the NHOG Vietnam dataset, which contains 741 sagittal ultrasound images with verified NT measurements and associated risk labels (normal or high-risk). These images were split into 80% for training and 20% for testing.

The pipeline consists of three main components:

- **Segmentation of the NT Region**

The first step was to segment the NT region from each image. used a DenseNet-based convolutional neural network, trained in an image-to-image fashion. For each input image, the model predicts a binary mask highlighting the NT region. DenseNet was chosen because of its efficient feature reuse, short skip connections, and strong gradient flow, which make it well-suited for medical image segmentation. The output image has the same size as the input, and each pixel is classified as either part of the NT or background. Training was supervised

using binary ground truth masks and a loss function based on the Dice coefficient to handle class imbalance.

Before training, applied data augmentation techniques to improve generalization including flips, rotations, brightness changes, and rescaling. All images were resized to 256×256 pixels and normalized using ImageNet statistics.

- **Measurement**

- 1. Contour Extraction**

In computer vision, a contour is defined as the continuous boundary that encloses an object, thereby converting pixel-level regions into geometric representations suitable for higher-level analysis. This step is particularly critical in medical imaging because it enables automated measurement by providing a mathematically robust description of anatomical structures (Suzuki & Abe, 1985).

- 2. Centroid**

In geometry, the centroid also referred to as the geometric center or center of mass is defined as the point that represents the arithmetic mean position of all points in a shape or region (Weisstein, 2014). Intuitively, it is the balancing point of a uniform object: if the object were made of a homogeneous material, the centroid would be the position at which it could be perfectly balanced.

In the context of computer vision and image processing, the centroid provides a compact descriptor of an object's spatial location within an image (Umbaugh, 2023). It is widely used as a reference point for tasks such as object alignment, shape analysis, and measurement.

- 3. Line Fitting and Measurement Axis**

In computer vision, line fitting is the process of finding the straight line that best approximates the distribution of contour points, typically based on a least-squares criterion (Duda & Hart, 1974).

#### 4. Intersection Points

A perpendicular line was drawn across the fitted axis. The two extreme intersection points between this line and the NT contour were extracted, corresponding to the maximum vertical cross-sectional boundaries of the translucency.

#### 5. Euclidean Distance in Pixels

The NT thickness in pixels was computed as the Euclidean distance between the two intersection points, using Equation (11) below.

$$\text{Formula} = d_{px} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (11)$$

#### 6. Pixel-to-Millimeter Conversion

Since the NT thickness is initially computed in pixel coordinates, a unit conversion is required to express the measurement in millimeters, which is the clinically accepted unit. This conversion was performed using calibration metadata derived from the image Dots Per Inch (DPI) property available in ultrasound image files. using Equation (12) below.

The physical size of a pixel  $S_{px}$  in millimeters can be calculated as:

$$\text{Formula} = S_{px} \frac{25.4}{DPI} \quad (12)$$

Where 25.4 corresponds to the number of millimeters in on each inch.

Thus, the final NT thickness in millimeters  $d_{mm}$  is obtain by multiplying the Euclidean distance in pixels,  $d_{px}$ , by the pixel size. using Equation (13) below.

$$\text{Formula} = d_{mm} = d_{px} \times S_{px} = d_{px} \times \frac{25.4}{DPI} \quad (13)$$

#### 7. Risk Classification

The final stage of the pipeline was dedicated to classifying each case as either normal or high-risk. This decision was made directly from the computed NT thickness expressed in millimeters. In accordance with international obstetric guidelines (Nicolaidis, 2004; Salomon et al., 2013), an NT value of 3.0 mm was used as the clinical threshold:

- **Normal case:  $NT < 3.0$  mm**
- **High-risk case:  $NT \geq 3.0$  mm**

Thus, the classification module operated on a simple yet clinically meaningful rule. Once the NT thickness was measured and converted from pixels to millimeters, the system automatically compared it against the 3.0 mm threshold. Any case exceeding or equal to this cut-off was labeled as high-risk, while values below the threshold were labeled as normal.

- **Model Development**

The classification model was developed using binary cross-entropy as the loss function, optimized with the Adam algorithm at an initial learning rate of 0.001. To mitigate overfitting and improve generalization, regularization strategies such as early stopping and dropout layers were incorporated during training.

- **Evaluation Metrics**

Model performance was assessed comprehensively using a range of evaluation metrics. Overall predictive accuracy was measured alongside sensitivity (true positive rate), specificity (true negative rate), and precision (positive predictive value). To provide a balanced assessment, the F1-score was computed as the harmonic mean of precision and sensitivity. Overlap-based measures including the Dice coefficient and Jaccard index were employed to evaluate segmentation-driven classification alignment. In addition, complementary diagnostic metrics were included: Negative Predictive Value (NPV), False Negative Rate (FNR), False Positive Rate (FPR), False Discovery Rate (FDR), and False Omission Rate (FOR), ensuring a complete analysis of both correct predictions and error tendencies.

The complete workflow for this stage, from data preparation to classification, is illustrated in Figure 3.56.

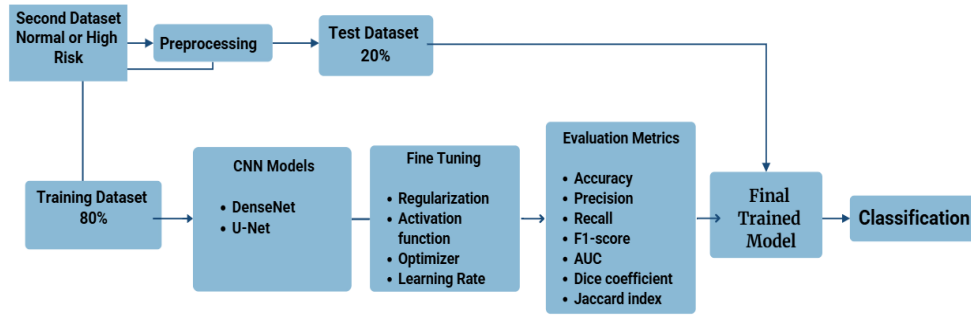


Figure (3.56) Second Approach Workflow

### 3.23 Conclusion

In summary, this chapter presented the proposed methodology and approaches developed in this thesis to analyze first-trimester fetal ultrasound images for automated NT measurement and risk classification. While previous researchers have adopted similar deep learning strategies, this study tailored the preprocessing pipeline to address the unique challenges of ultrasound imaging, such as speckle noise, low contrast, and class imbalance. Both automatic and manual preprocessing procedures were integrated to ensure dataset integrity and clinical relevance. Furthermore, a range of deep learning architectures and hyperparameter tuning strategies were applied to optimize model performance across quality assessment, segmentation, measurement, and classification tasks. Finally, multiple evaluation techniques and performance metrics were employed to rigorously assess the effectiveness of the proposed models. The subsequent chapter will present and discuss the results derived from these experiments.

### 4.1 Introduction

In this chapter, two complementary deep-learning approaches were implemented to evaluate the effectiveness of automated analysis for fetal Nuchal Translucency (NT) ultrasound images.

The first approach focused on image-quality assessment to ensure that only standard-quality NT images were used for subsequent analysis. Seven Convolutional Neural Network (CNN) architectures were explored for this task; namely DenseNet121, VGG16, VGG19, ResNet18, ResNet50, EfficientNetB7, and InceptionV3 allowing a comprehensive comparison of performance across diverse model families.

The second approach covered segmentation, measurement, and downstream risk classification. Two CNN-based models—DenseNet and U-Net—were trained to segment the Nuchal Translucency (NT), enabling precise thickness estimation and subsequent risk labeling. All experiments used the same preprocessing pipeline to ensure fairness: images were resized consistently; batch size and number of epochs were selected via validation; and optimization used Adam with a tuned learning rate and early stopping. Classification on the held-out test set is reported using accuracy, precision, recall (sensitivity), F1-score, and Area Under Curve (AUC). Segmentation quality and NT-measurement accuracy are summarized using the Dice Similarity Coefficient (DSC) and Mean Absolute Error (MAE).

This chapter presents the detailed experimental results for both approaches, demonstrating their predictive capabilities and highlighting their potential to support early detection of chromosomal abnormalities.

Table (4.1) Summarizes the Models and Their Key Training Settings for Both Datasets

| Dataset                                  | Model / Purpose          | Classes                     | Optimizer | Learning rate | Batch size | Epochs |
|------------------------------------------|--------------------------|-----------------------------|-----------|---------------|------------|--------|
| <b>Fetus Framework (Image Quality)</b>   | DenseNet121 classifier   | 2 (Standard / Non-Standard) | Adam      | 0.001         | 8          | 50     |
| <b>NHOG (Segmentation / Measurement)</b> | DenseNet121 segmentation | 2 (background / NT)         | Adam      | 0.0001        | 32         | 50     |

## 4.2 Results of the First Dataset Approach

In the first approach, as described earlier, the dataset contained a total of 1,684 two-dimensional sagittal ultrasound images. These images were split into 70 % for training (1,178 images), 15 % for validation (253 images), and 15 % for testing (253 images). This section presents the results obtained from this image-quality assessment approach using seven CNN models DenseNet121, VGG16, VGG19, ResNet18, ResNet50, EfficientNetB7, and InceptionV3 together with the outcomes of the performance metrics used for evaluation. Accuracy, precision, recall, F1-score, and Area Under Curve (AUC) are reported to compare the effectiveness of these architectures in identifying standard-quality fetal Nuchal Translucency ultrasound images.

### 4.2.1 DenseNet121

The training process of the DenseNet121 model for fetal Nuchal Translucency (NT) image-quality assessment was monitored over 50 epochs, with the evolution of loss and accuracy illustrated in Figure 4.1. The training loss decreased steadily from an initial value of approximately 0.55 to a final value of 0.117, indicating effective convergence of the network.

The validation accuracy rose rapidly during the early epochs and then stabilized, reaching a peak accuracy of 94.07 % and concluding with a final validation accuracy of 92.49 %. Although minor oscillations were observed in the validation curve a common behavior when the validation subset is comparatively small the overall trend demonstrates that DenseNet121 maintained high predictive capability throughout training.

The combined training-progress plot highlights the inverse relationship between the steadily decreasing loss and the rising validation accuracy. The close agreement between the final training and validation metrics suggests that the model achieved good generalization and did not suffer from significant overfitting.

These results confirm that the DenseNet121 architecture, trained with an input size of  $224 \times 224$  pixels, a batch size of 8, and the Adam optimizer with a 0.001 learning rate, provides a robust foundation for the first approach, enabling reliable discrimination between standard-quality and non-standard fetal Nuchal Translucency ultrasound images.

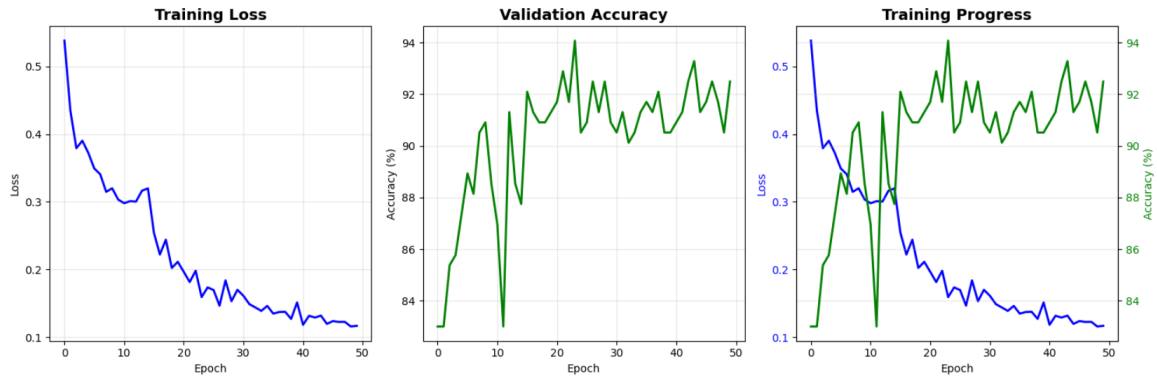


Figure (4.1) Training and Validation Performance of the Densenet121 Model for NT Image-Quality Assessment

Similar to the previous experiments, the confusion matrix of the DenseNet model for image quality assessment is presented in Figure 4.2. The four values indicate the distribution of correctly and incorrectly classified samples. The model successfully identified 172 Standard images out of the ground-truth Standard cases (True Positives) and 66 Non-standard images correctly (True Negatives). On the other hand, it misclassified 8 Standard cases as Non-standard (False Negatives) and 8 Non-standard cases as Standard (False Positives). These results indicate that the DenseNet model achieved a balanced and highly accurate performance, with both sensitivity to Standard images and specificity for Non-standard images being consistently high. The low number of misclassifications demonstrates the robustness of the model in distinguishing between high-quality (Standard) and low-quality (Non-standard) ultrasound images.

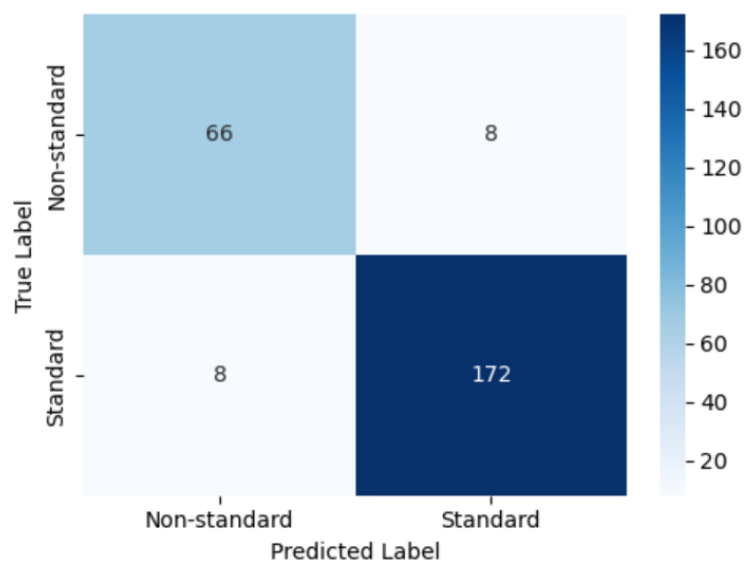


Figure (4.2) Confusion Matrix for DenseNet121 in the First Approach

In the same manner, the performance evaluation of DenseNet121 was further supported with the ROC curve, as illustrated in Figure 4.3. At the selected operating threshold, the model achieved an overall test accuracy of 94%, showing that 93.70% of the test samples were classified correctly. The recall (TPR) reached 0.9370, indicating that 93.70% of the Standard cases were successfully identified. Precision was 0.9370, reflecting the proportion of correctly identified Standard images among all predictions of the Standard class. The F1-score was 0.9370, representing an excellent balance between precision and recall. Finally, the ROC curve shows the relationship between TPR and FPR across thresholds, with an Area Under Curve (AUC) of 0.986, which confirms that DenseNet121 has an exceptional discriminative ability between Standard and Non-standard classes, summarized in Table 4.2.

Table (4.2) Test Performance of DenseNet121 (Classification) and ROC Summary

| Metric               | Value  |
|----------------------|--------|
| Accuracy             | 0.94   |
| Precision            | 0.9370 |
| Recall (Sensitivity) | 0.9370 |
| F1-score             | 0.9370 |
| ROC-AUC              | 0.986  |

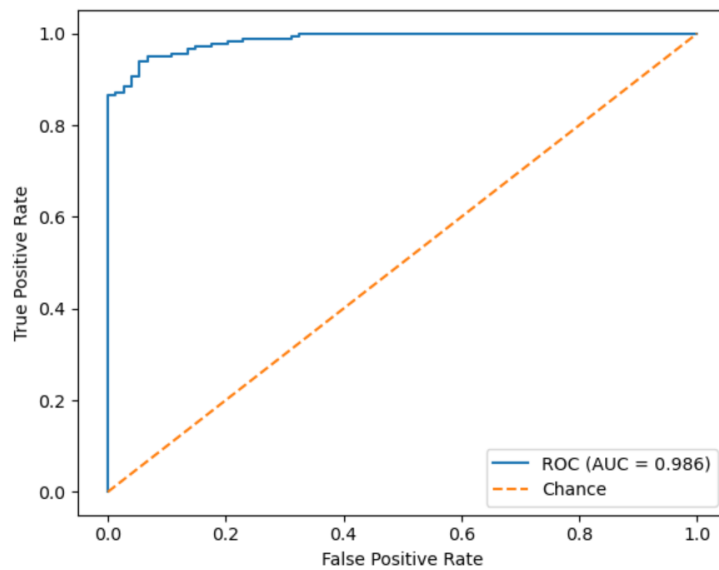


Figure (4.3) AUC for DenseNet121 in the First Approach

Figure 4.4 illustrates representative qualitative results from the DenseNet model applied to the Fetus Framework dataset, showing both correct and incorrect classifications of image quality.

Each panel displays the ground-truth label (“True”) and the predicted label (“Pred”) together with the model’s confidence score.

Images with green annotations represent correct predictions, while red annotations highlight misclassifications.

For example, the top-left image shows a correctly identified Standard plane with a confidence of 0.999, whereas the second image on the bottom row demonstrates a Non-standard plane that the network incorrectly predicted as Standard with 0.849 probability. Such visualization provides insight into challenging borderline cases, including images with subtle artifacts or partial fetal profiles that can mislead the classifier.

Overall, these qualitative examples complement the quantitative metrics by confirming that the DenseNet model not only performs well statistically but also delivers reliable, interpretable predictions for real ultrasound examinations.

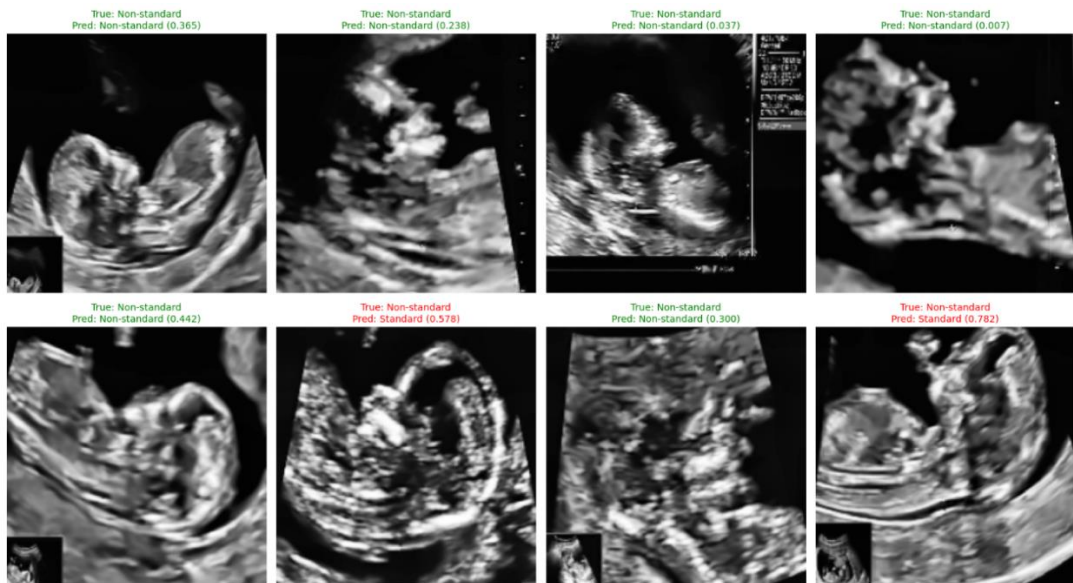


Figure (4.4) Sample Predictions of Non-standard Ultrasound Images Using DenseNet121

To further interpret the decision-making process of the model, Grad-CAM (Rs et al., 2020) visualization was applied, as illustrated in Figure 4.5. The left image represents the original ultrasound scan, while the right image overlays the Grad-CAM heatmap on

the same scan. The highlighted regions in red and yellow indicate the areas that contributed most to the model's classification decision. As shown, the model focused its attention primarily on the fetal nuchal region, which is the clinically relevant area for assessing Nuchal Translucency (NT). This indicates that the network successfully localized the anatomical region of interest rather than relying on irrelevant background features. The concentration of activation in the NT area supports the validity of the model's predictions and provides an additional layer of explainability, which is essential in medical imaging applications. Overall, the Grad-CAM results confirm that the model is not only achieving high classification performance but is also learning meaningful representations aligned with clinical expertise.

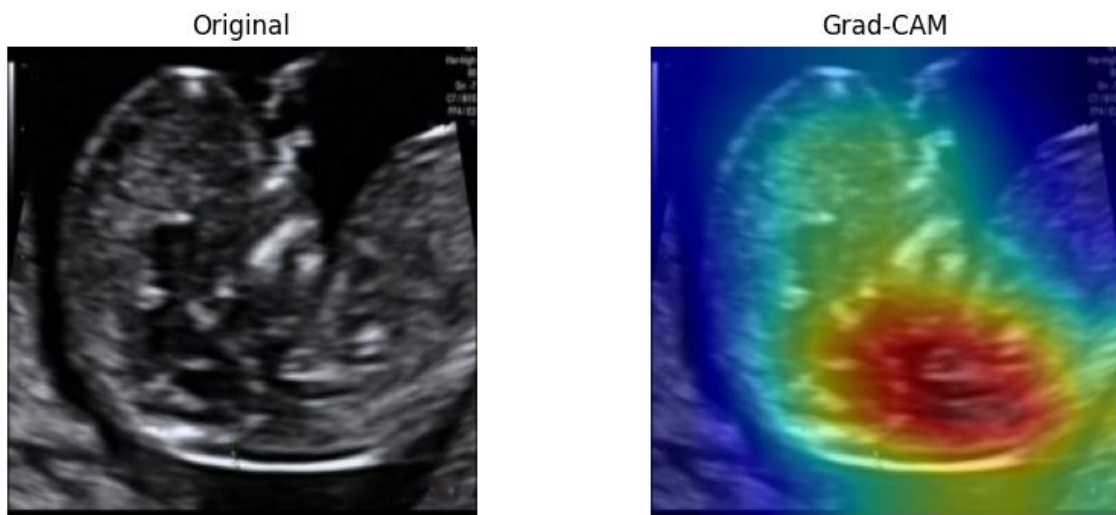


Figure (4.5) Original Ultrasound Image and Corresponding Grad-CAM Heatmap

#### 4.2.2 VGG16

The second model evaluated VGG16. Figure 4.6 shows the accuracy and loss curves during 50 training epochs. The training accuracy started at approximately 0.65 and gradually improved, reaching values around 0.90 in the final epochs, albeit with notable fluctuations throughout the training process. The validation accuracy, on the other hand, began at a higher point near 0.86 and remained consistently stable, ranging between 0.87 and 0.91, which indicates strong generalization to unseen data. Regarding the loss behavior, the training loss decreased from an initial value of 1.6 and fluctuated around 0.3–0.4 in later epochs, while the validation loss remained lower and more stable, declining from 0.35 to around 0.22 by the final epoch. The relatively small gap between training and validation performance suggests that VGG16 converged effectively without

significant overfitting. Overall, VGG16 demonstrated robust and reliable classification performance, maintaining good validation accuracy with stable loss reduction.

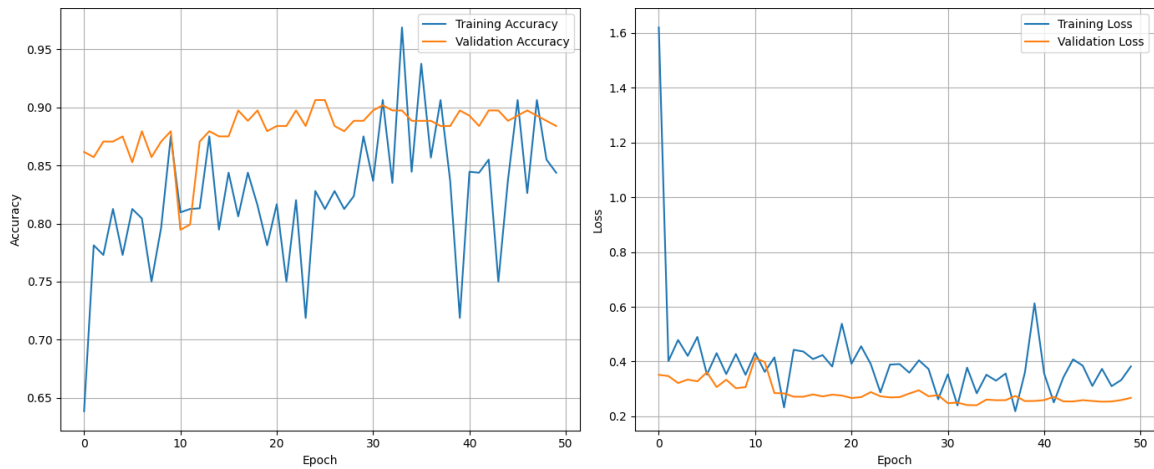


Figure (4.6) Loss and Accuracy vs. Epoch for VGG16

Similar to the previous model, the confusion matrix was populated from the validation set predictions as shown in Figure 4.7 below. The four main values in the matrix were as follows: TP = 106, FN = 16, FP = 44, and TN = 58. That is, the model succeeded in classifying 106 standard images out of 122 (TP + FN), which is the True Positive, while misclassified 16 and classified them as non-standard, called False Negative. Moreover, the model successfully recognized 58 non-standard images which are called True Negative, while misclassified 44 and recognized them as standard, which is called False Positive.

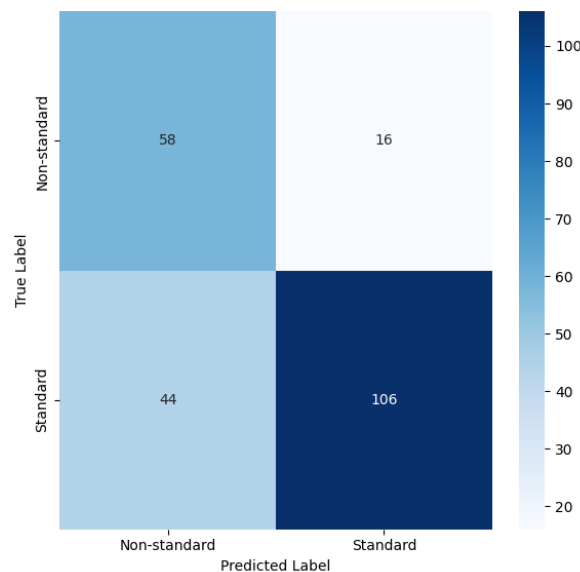


Figure (4.7) Confusion Matrix for VGG16

In the same manner, the performance evaluation of VGG16 was complemented with the ROC curve, as illustrated in Figure 4.8. The model achieved an overall accuracy of 0.82, indicating that 82% of the samples were correctly classified. The recall (TPR) reached 0.71, meaning that 71% of the Standard images were successfully identified, while the false positive rate (FPR) was 0.22, showing that 22% of Non-standard images were incorrectly predicted as Standard. Precision was calculated at 0.87, reflecting the proportion of correctly predicted Standard images among all Standard predictions. The harmonic mean between precision and recall resulted in an F1-score of 0.78, highlighting a balanced yet slightly recall-limited performance. Finally, the ROC curve shows the relationship between TPR and FPR, with an Area Under Curve (AUC) of 0.816.

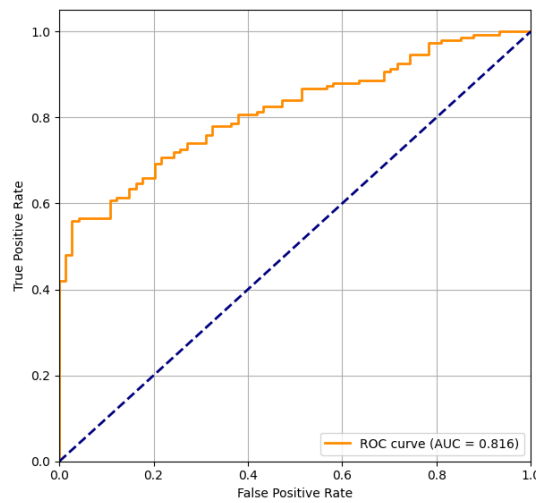


Figure (4.8) AUC for VGG16

Finally, the prediction probability distribution of the VGG16 model is shown in Figure 4.9. The histogram reveals two clear peaks concentrated near the extreme probabilities of 0 and 1, which indicates that the model often produced highly confident predictions when classifying images as either Non-standard or Standard. The decision threshold, represented by the red dashed vertical line at 0.5, divides the predictions between the two classes. Only a relatively small number of predictions were distributed around the middle probability range, suggesting that the model rarely exhibited uncertainty in its classifications. This outcome aligns with the confusion matrix results, where the model demonstrated a good balance between true positives and true negatives, although some misclassifications still occurred. Overall, the probability distribution confirms that VGG16 was generally decisive in its predictions, showing strong discriminative confidence across the two classes.

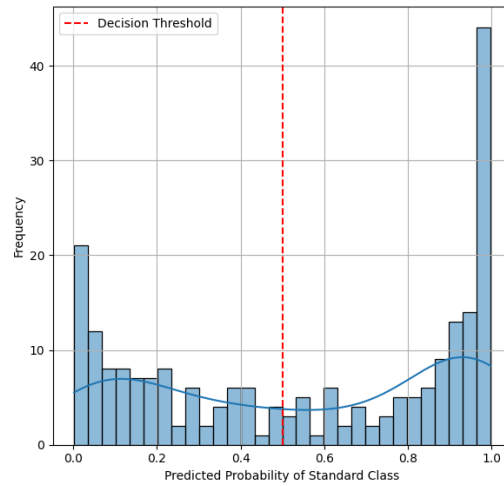


Figure (4.9) Distribution of Predicted Probabilities for the Standard Class for VGG16

### 4.2.3 VGG19

The third model evaluated VGG19. Figure 4.10 illustrates the accuracy and loss curves over 50 training epochs. The training accuracy began at approximately 0.55 and steadily improved, reaching values close to 0.90 by the end of training, although with noticeable fluctuations throughout the epochs. The validation accuracy started at a relatively high point of 0.80 and maintained consistent improvement, stabilizing between 0.88 and 0.91 in the later epochs, which demonstrates strong generalization capability. Regarding the loss, the training loss decreased sharply from an initial value of 1.45 and fluctuated around 0.3–0.4 in the subsequent epochs, while the validation loss steadily declined from 0.55 to approximately 0.25. Compared with VGG16, the performance of VGG19 showed very similar convergence behavior, with slightly higher stability in validation loss but more fluctuation in training accuracy. Overall, VGG19 exhibited reliable classification performance with high validation accuracy and low validation loss, making it a strong model, although not significantly superior to its shallower counterpart.

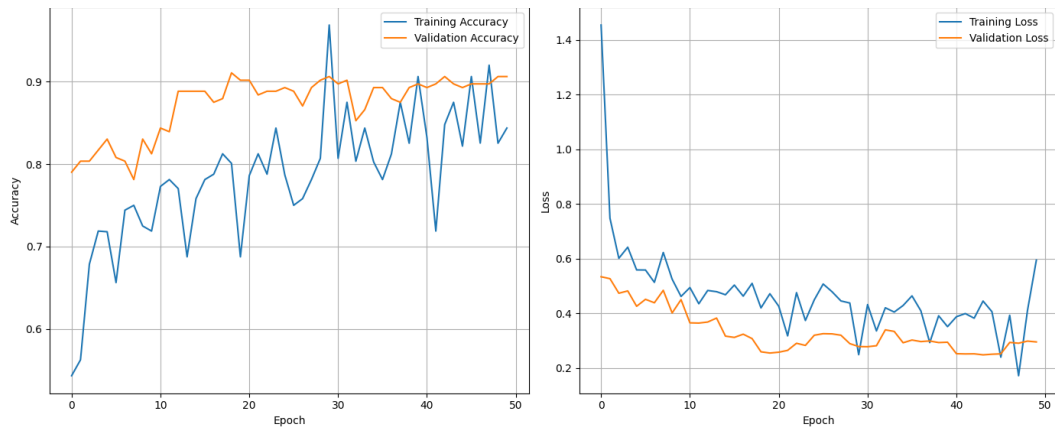


Figure (4.10) Loss and Accuracy vs. Epoch for VGG19

Similar to the previous models, the confusion matrix for VGG19 was generated from the predictions on the testing set, as shown in Figure 4.11. The four values in the matrix were as follows: TP = 102, FN = 48, FP = 15, and TN = 59. This indicates that the model correctly classified 102 *Standard* images (True Positive), while 48 *Standard* images were misclassified as *Non-standard* (False Negative). In addition, 59 *Non-standard* images were correctly recognized (True Negative), whereas 15 *Non-standard* cases were incorrectly predicted as *Standard* (False Positive). These results demonstrate that VGG19 achieved a strong balance between sensitivity and specificity, with a comparable performance to VGG16. However, the number of false negatives suggests that the model's recall for *Standard* cases could still be improved. Overall, VGG19 maintained solid classification ability, though without showing a significant advantage over the shallower VGG16 model.

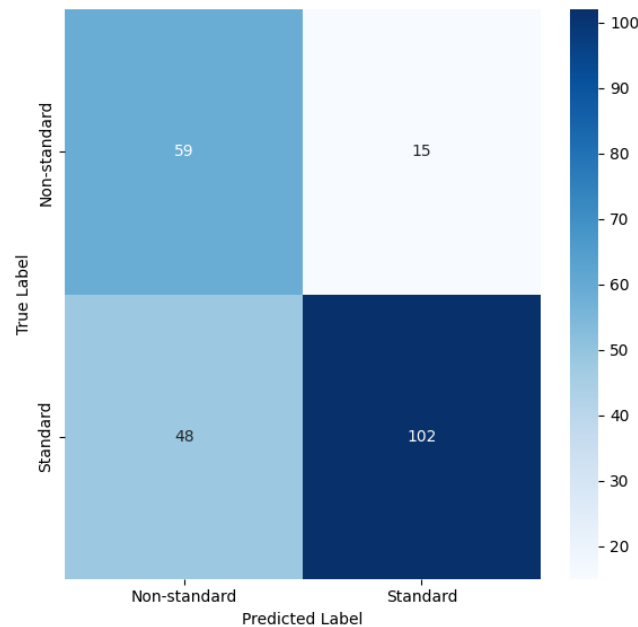


Figure (4.11) Confusion Matrix for VGG19

In the same manner, the performance evaluation of VGG19 was further supported with the ROC curve, as illustrated in Figure 4.12. The model achieved an overall accuracy of 0.82, showing that 82% of the test samples were classified correctly. The recall (TPR) reached 0.68, indicating that 68% of the *Standard* cases were successfully identified, while the false positive rate (FPR) was 0.20, showing that 20% of the *Non-standard* images were incorrectly predicted as *Standard*. Precision was computed as 0.87, reflecting the proportion of correctly identified *Standard* images among all predictions of the *Standard* class. The F1-score was 0.76, representing a balanced trade-off between

precision and recall. Finally, the ROC curve shows the relationship between TPR and FPR, with an Area Under Curve (AUC) of 0.83, which confirms that VGG19 has strong discriminative ability between Standard and Non-standard classes. Compared to VGG16 (AUC = 0.816), the performance of VGG19 was slightly better in terms of AUC, though both models achieved nearly comparable overall results.

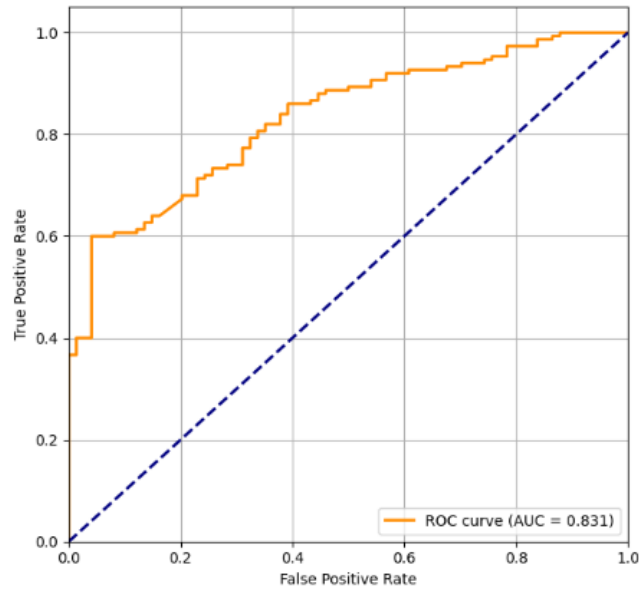


Figure (4.12) AUC for VGG19

Finally, the prediction probability distribution of the VGG19 model is illustrated in Figure 4.13. The distribution shows that the predicted probabilities were spread relatively evenly across the full range from 0 to 1, with no dominant clustering near the extremes. The decision threshold at 0.5, indicated by the red dashed vertical line, divides the predictions into Standard and Non-standard classes. Unlike models such as VGG16 , which tended to produce highly confident predictions near 0 or 1, VGG19 generated a higher proportion of moderate-confidence outputs. This behavior suggests that the model often exhibited uncertainty when classifying samples, which is consistent with the confusion matrix results where both false positives and false negatives were present. Overall, the probability distribution reflects the balanced yet cautious decision-making of VGG19, highlighting its ability to generalize across the two classes, though at the cost of producing more uncertain predictions compared to models with sharper confidence peaks.

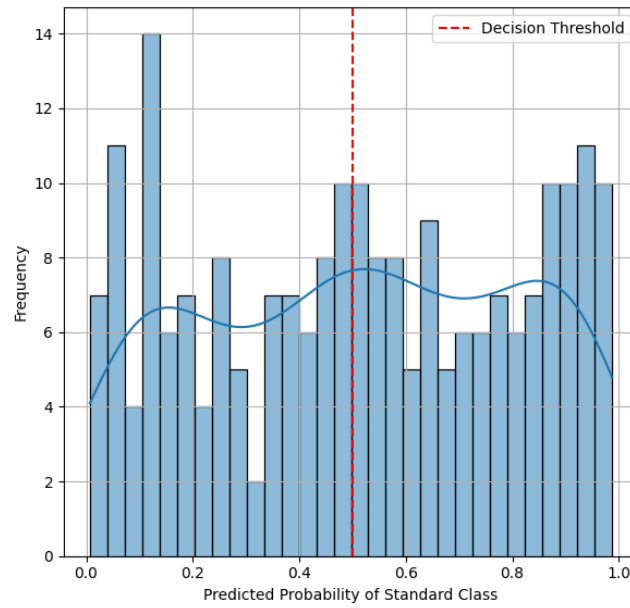


Figure (4.33) Distribution of Predicted Probabilities for The Standard Class for VGG19

#### 4.2.4 ResNet 18

The fourth model evaluated ResNet18. Figure 4.14 illustrates the accuracy and loss curves over the 50 training epochs. Similar to the previous models, both the training and validation loss functions decreased progressively, indicating stable convergence. The training loss started at approximately 0.9 and steadily decreased to nearly 0.1 by the final epoch, while the validation loss began at a much higher value of around 36.8 before dropping sharply in the early epochs and stabilizing close to 0.5. In terms of accuracy, the training accuracy started at 0.70 and improved consistently to reach around 0.90, whereas the validation accuracy began at a much lower value of 0.32 but increased markedly after the 10th epoch, eventually fluctuating between 0.80 and 0.92. Despite this overall improvement, the validation accuracy curve exhibited noticeable instability across epochs, which could be attributed to factors such as dataset imbalance, variance in sample quality, or sensitivity to the chosen learning rate. It should be noted that the learning rate applied in this configuration was 0.001, which provided a reasonable balance between convergence speed and stability.

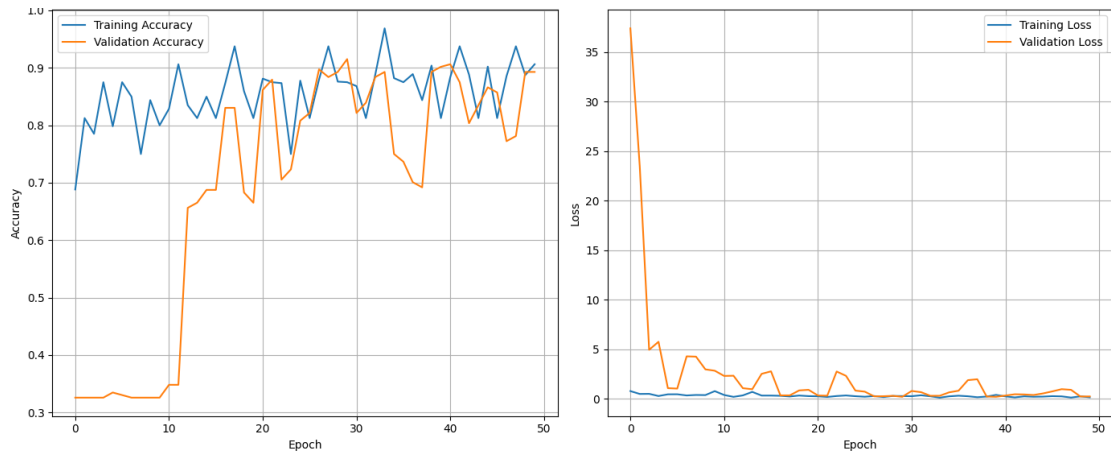


Figure (4.14) Loss and Accuracy vs. Epoch for ResNet18

Similar to the previous model, the confusion matrix was generated from the testing set predictions, as illustrated in Figure 4.15. The four main values in the matrix were as follows: TP = 98, FN = 52, FP = 2, and TN = 72. In other words, the model correctly classified 98 images as Standard (True Positive) but misclassified 52 Standard images as Non-standard (False Negative). At the same time, it successfully recognized 72 Non-standard images (True Negative) while incorrectly classifying 2 of them as Standard (False Positive). This distribution indicates that the model performs well in detecting both classes, with a stronger reliability in identifying Standard images, although some misclassification of Standard cases remains evident.

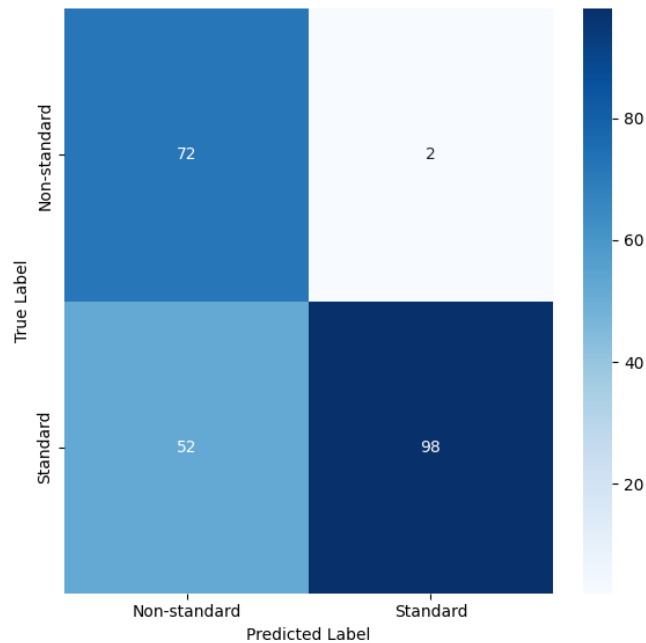


Figure (4.15) Confusion Matrix for ResNet18

In the same manner, the performance measures were also derived from the confusion matrix values. First, the overall accuracy reached 0.86, indicating that 86% of the outputs were classified correctly. Additionally, the recall (TPR) was 0.65, meaning that 65% of the Standard images were correctly recognized by the model. On the other hand, the false positive rate (FPR) was 0.027, signifying that only 2.7% of the Non-standard cases were incorrectly classified as Standard. By the same token, precision was calculated as 0.98, representing the proportion of correctly predicted Standard images to all predictions of the Standard class. In the same way, the F1-score was equal to 0.78, reflecting a balanced trade-off between precision and recall. Lastly, the ROC curve in Figure 4.16 illustrates the relationship between TPR and FPR, where the x-axis denotes the FPR and the y-axis the TPR. The Area Under Curve (AUC) reached a high value of 0.918, which demonstrates that the model has a strong discriminative ability between Standard and Non-standard classes. Compared to previous models, this performance confirms the robustness of ResNet18 in handling the classification task, although the recall value indicates that some Standard cases remain challenging to detect.

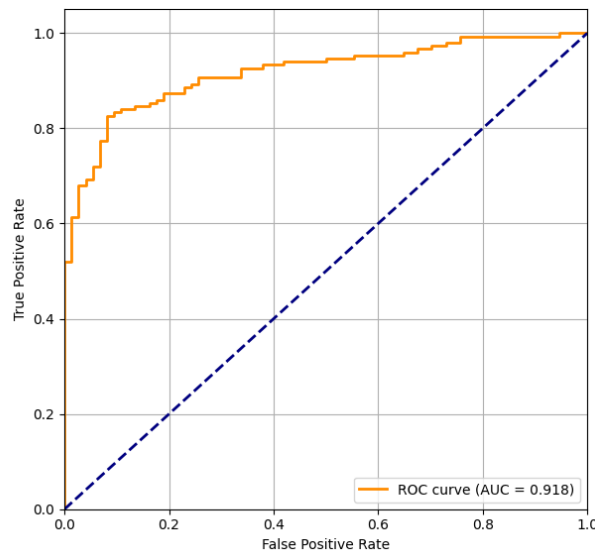


Figure (4.16)4 AUC for ResNet18

Furthermore, the prediction probability distribution of the Standard class was analyzed, as shown in Figure 4.17. The histogram illustrates how frequently certain probability values occurred when the model assigned predictions to the test samples. It can be observed that the majority of predictions were concentrated near the extreme values of 0 and 1, indicating that the model often exhibited high confidence when classifying an image as either Non-standard or Standard. The decision threshold was set

at 0.5, represented by the red dashed vertical line, dividing the predictions into the two classes. While a small portion of the predictions were distributed around the mid-range probabilities, the clear peaks near 0 and 1 confirm that the model rarely produced ambiguous outputs.

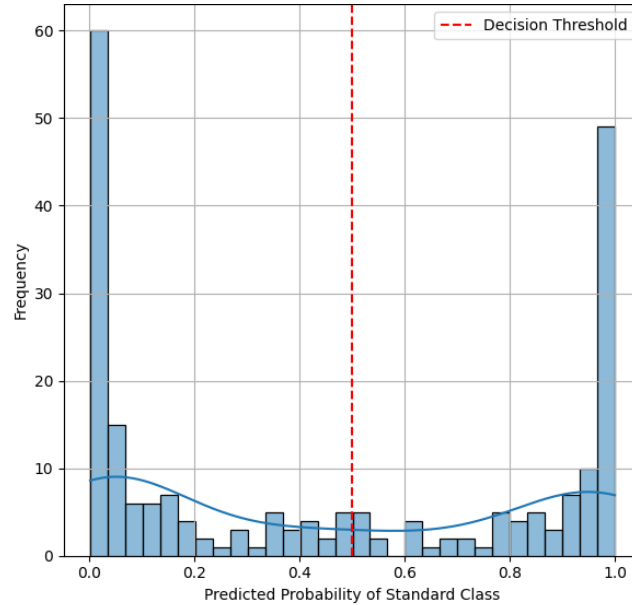


Figure (4.17) Distribution of Predicted Probabilities for The Standard Class for Resnet18

#### 4.2.5 ResNet 50

The fifth model examined ResNet50. Figure 4.18 presents the accuracy and loss curves across 30 training epochs. The training loss decreased gradually from an initial value of around 2.8 and stabilized near 0.2 by the final epoch, confirming smooth convergence of the model. In contrast, the validation loss started at a very high value of approximately 36.5, then dropped sharply within the first few epochs before exhibiting fluctuations throughout training, and eventually converging near 2.0. Regarding accuracy, the training accuracy began at 0.70 and consistently improved, reaching almost 0.99 by the final epoch. The validation accuracy, however, showed instability in the early stages, dropping significantly to around 0.30 between epochs 5 and 15 before rising steeply after epoch 20, where it stabilized between 0.85 and 0.92. This irregularity in the validation accuracy curve suggests that the model may have struggled with generalization in the earlier epochs, possibly due to sensitivity to hyperparameters or the nature of the dataset.

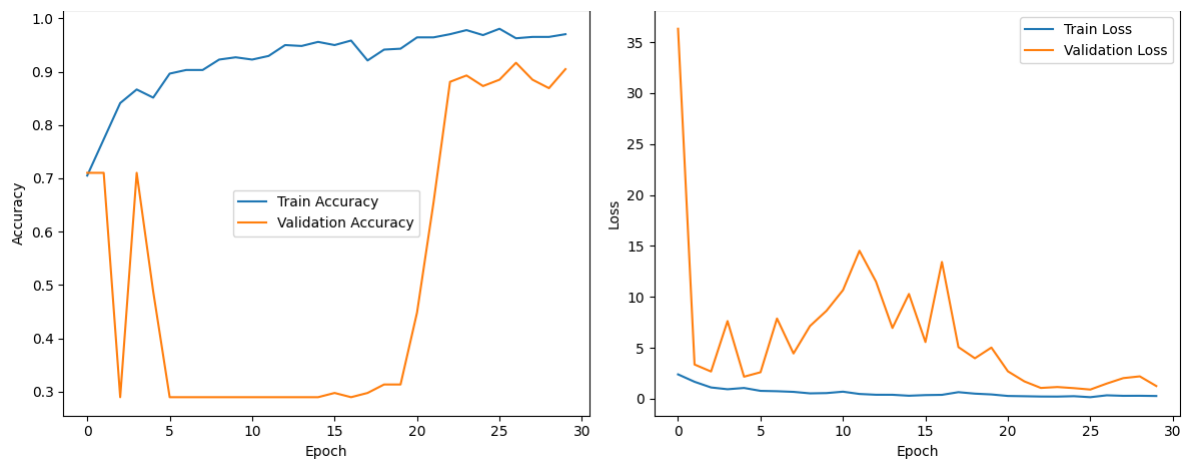


Figure (4.18) Loss and Accuracy vs. Epoch for ResNet50

Similar to the previous models, the confusion matrix for ResNet50 was generated from the predictions on the testing set, as shown in Figure 4.19. The four values in the matrix were as follows: TP = 121, FN = 60, FP = 9, and TN = 65. This means that the model correctly classified 121 images as Standard (True Positive), while it misclassified 60 Standard images as Non-standard (False Negative). In addition, the model successfully recognized 65 Non-standard images (True Negative), while only 9 Non-standard cases were incorrectly predicted as Standard (False Positive). Compared to ResNet18, the ResNet50 achieved a higher number of correctly identified Standard cases but still showed a considerable portion of misclassified Standard images, which affected the recall rate.

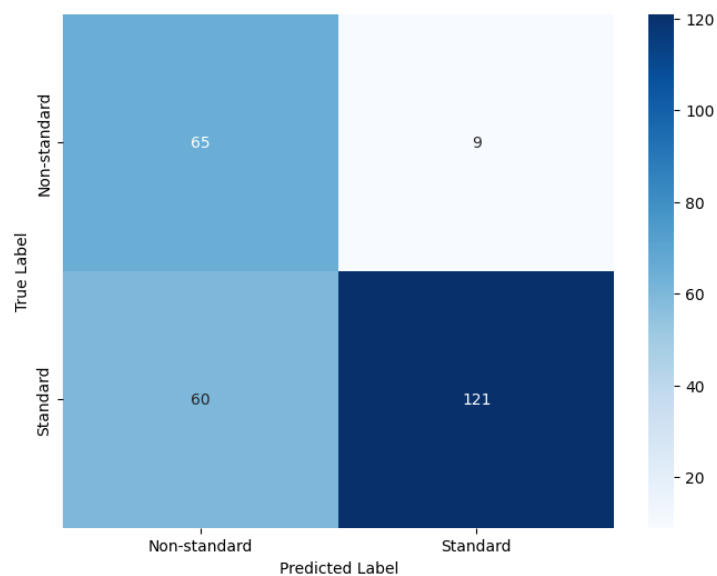


Figure (4.19) Confusion Matrix for ResNet50

In the same manner, the performance metrics of the ResNet50 model were derived from the confusion matrix values. First, the overall accuracy reached 0.80, which indicates that 80% of the predictions were correctly classified. Additionally, the recall (TPR) was 0.67, meaning that 67% of the Standard images were correctly detected by the model. On the other hand, the false positive rate (FPR) was 0.12, showing that 12% of the Non-standard images were incorrectly classified as Standard. By the same token, the precision achieved 0.93, reflecting the proportion of correctly predicted Standard images among all samples predicted as Standard. The F1-score was calculated as 0.78, which represents a balanced trade-off between recall and precision. Lastly, the ROC curve in Figure 4.20 illustrates the relationship between TPR and FPR. The Area Under Curve (AUC) reached 0.86, confirming that the model has a strong ability to discriminate between Standard and Non-standard images. Compared to ResNet18, the ResNet50 achieved higher overall accuracy and precision but maintained a lower recall, which implies that some Standard cases remained more difficult to capture.

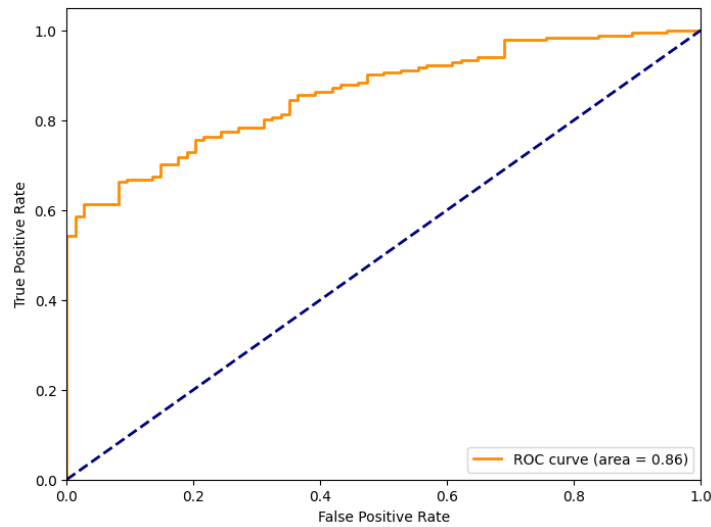


Figure (4.20) AUC for ResNet50

#### 4.2.6 EfficientNetB7

The sixth model EfficientNetB7. Figure 4.21 presents the accuracy and loss curves across 50 training epochs. The training accuracy began at approximately 0.68 and steadily improved, reaching almost 0.98 by the final epoch. The validation accuracy, although initially unstable and dropping to around 0.33 at epoch 3, showed significant recovery afterwards and stabilized between 0.82 and 0.90 during the later epochs. This indicates that the model was ultimately able to generalize well despite an unstable start. In terms

of loss behavior, the training loss decreased consistently to values close to zero, reflecting effective convergence. However, the validation loss exhibited extreme instability in the early epochs, with a sharp spike exceeding 1000 before dropping back to near-zero values after epoch 4. Such fluctuations in the validation loss may be attributed to initialization sensitivity, large batch effects, or learning rate dynamics. Nevertheless, after stabilization, both training and validation loss remained low, supporting the strong accuracy results. Overall, EfficientNetB7 demonstrated good performance, achieving one of the strongest generalization capabilities among the tested models, albeit with pronounced instability during the early training phase.

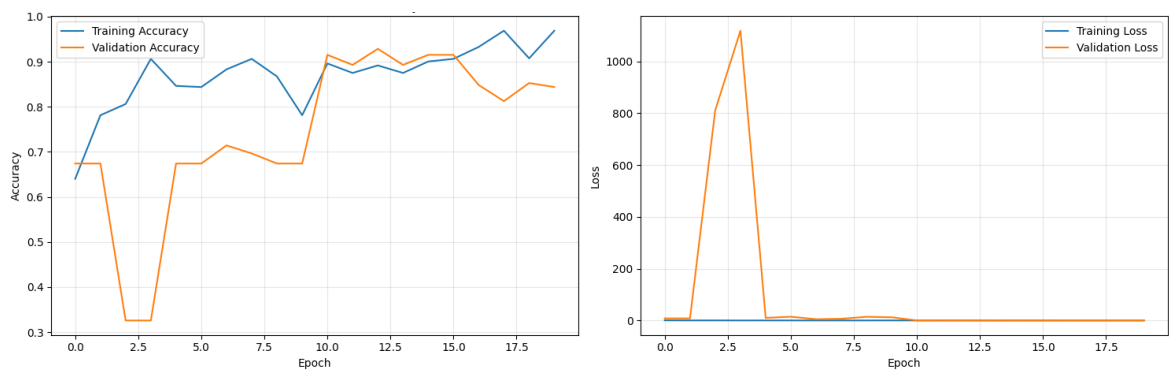


Figure (4.21) Loss and Accuracy vs. Epoch for EfficientNetB7

Similar to the previous models, the confusion matrix for EfficientNetB7 was generated from the testing set predictions, as illustrated in Figure 4.22. The four values in the matrix were as follows: TP = 128, FN = 22, FP = 28, and TN = 46. This means that the model correctly classified 128 Standard images (True Positive), while misclassifying 22 Standard images as Non-standard (False Negative). At the same time, it successfully identified 46 Non-standard cases (True Negative), but 28 Non-standard images were incorrectly predicted as Standard (False Positive). These results demonstrate that EfficientNetB7 achieved strong performance in recognizing Standard cases, with relatively high sensitivity, though its ability to identify Non-standard images was weaker compared to other models due to the higher false positive rate. Overall, the matrix indicates a robust classification capability, but with room for improvement in distinguishing Non-standard cases.

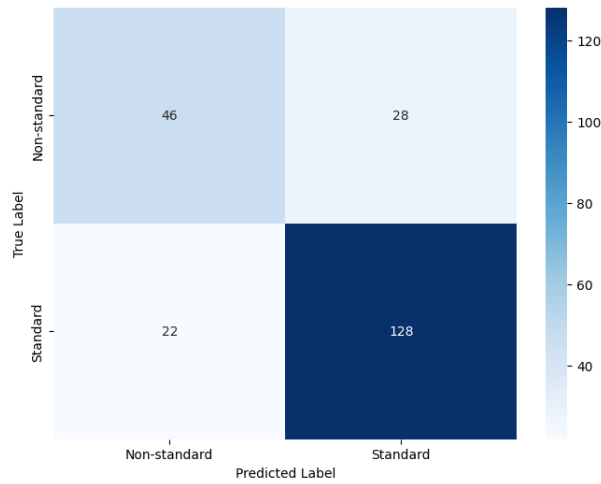


Figure (4.22) Confusion Matrix for EfficientNetB7

In the same manner, the performance evaluation of EfficientNetB7 was complemented with the ROC curve, as shown in Figure 4.23. The model achieved an overall accuracy of 0.82, indicating that 82% of the samples were classified correctly. The recall (TPR) reached 0.85, meaning that 85% of the Standard cases were successfully identified. On the other hand, the false positive rate (FPR) was 0.38, showing that 38% of Non-standard cases were incorrectly predicted as Standard. Precision was computed as 0.82, representing the proportion of correctly predicted Standard images among all Standard predictions. The harmonic mean between precision and recall resulted in an F1-score of 0.83, confirming a balanced trade-off between the two metrics. Finally, the ROC curve illustrates the relationship between TPR and FPR, with the Area Under Curve (AUC) equal to 0.83. This value demonstrates that the EfficientNetB7 model possesses a strong discriminative ability between Standard and Non-standard classes. Compared with EfficientNetB0, which performed at random level (AUC = 0.50), EfficientNetB7 shows a substantial improvement, approaching the robustness of ResNet-based models.

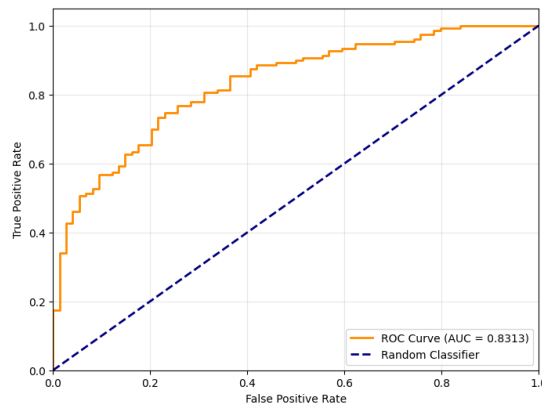


Figure (4.23) AUC for EfficientNetB7

#### 4.2.7 InceptionV3

The seventh model InceptionV3. Figure 4.24 shows the training and validation accuracy and loss curves over 50 epochs. The training accuracy started at 0.55 and gradually improved, reaching around 0.91 by the final epoch. Similarly, the validation accuracy began at 0.80 and consistently increased, stabilizing between 0.88 and 0.91 during the later epochs. This indicates strong generalization capability and balanced performance across training and validation sets. Regarding the loss function, the training loss decreased steadily from an initial value of approximately 1.45 to values around 0.40, while the validation loss followed a similar downward trend, dropping from 0.55 to nearly 0.25 by the end of training. The stability of the validation curves, with minimal fluctuations compared to other models, suggests that InceptionV3 converged effectively without severe overfitting. Overall, InceptionV3 demonstrated reliable and consistent learning behavior, maintaining high accuracy with relatively low loss, making it one of the more stable and competitive models in this study.

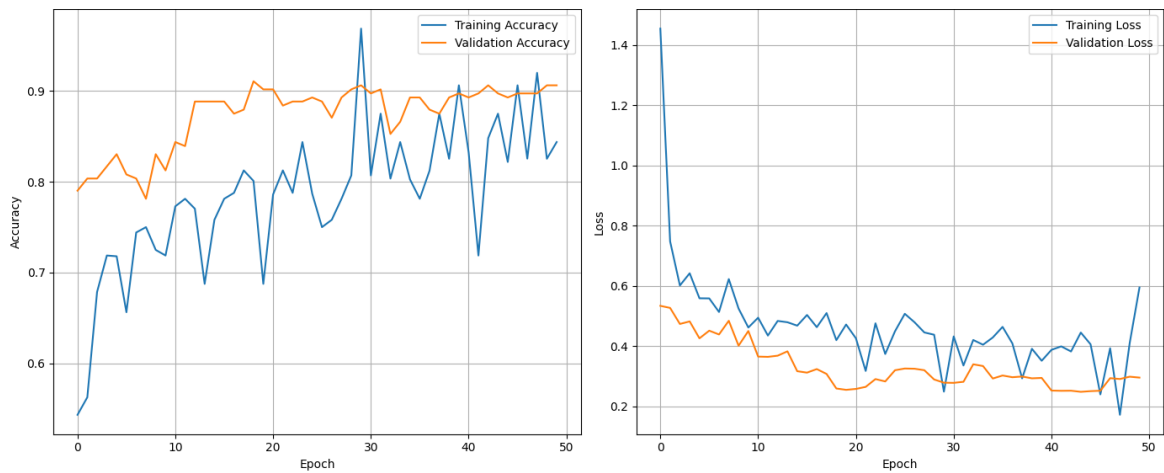


Figure (4.24) Loss and Accuracy vs. Epoch for InceptionV3

Similar to the previous models, the confusion matrix for InceptionV3 was generated from the predictions on the testing set, as shown in Figure 4.25. The four values in the matrix were as follows: TP = 102, FN = 48, FP = 15, and TN = 59. This means that the model correctly classified 102 Standard images (True Positive), while misclassifying 48 Standard cases as Non-standard (False Negative). At the same time, it successfully recognized 59 Non-standard images (True Negative), but 15 Non-standard cases were incorrectly predicted as Standard (False Positive). These results suggest that InceptionV3 achieved a good balance between sensitivity and specificity, with better recognition of

Standard cases compared to Non-standard ones. Although the model demonstrated solid overall performance, the relatively high number of false negatives indicates that further improvements are needed to enhance its ability to consistently detect Standard cases.

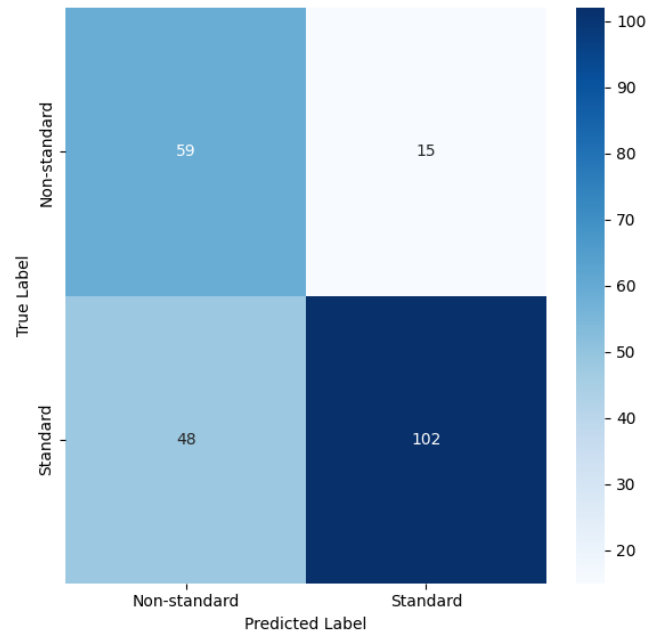


Figure (4.25) Confusion Matrix for InceptionV3

In the same manner, the performance evaluation of the InceptionV3 model was complemented with the ROC curve, as shown in Figure 4.26. The model achieved an overall accuracy of 0.82, indicating that 82% of the samples were classified correctly. The recall (TPR) reached 0.84, meaning that 84% of the Standard cases were successfully identified. On the other hand, the false positive rate (FPR) was 0.32, showing that 32% of Non-standard cases were incorrectly predicted as Standard. Precision was computed as 0.81, representing the proportion of correctly predicted Standard images among all Standard predictions. The harmonic mean between precision and recall resulted in an F1-score of 0.82, confirming a balanced trade-off between the two metrics. Finally, the ROC curve illustrates the relationship between TPR and FPR, with the Area Under Curve (AUC) equal to 0.831.

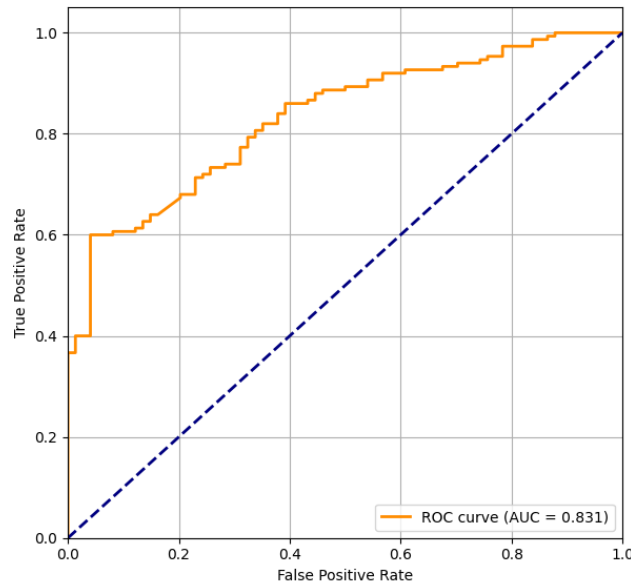


Figure (4.26) AUC for InceptionV3

Finally, the prediction probability distribution of the InceptionV3 model is illustrated in Figure 4.27. Unlike models such as ResNet18, which showed clear peaks near the extremes, the probability values here were spread more evenly across the entire range from 0 to 1. This indicates that the model often produced moderate confidence predictions rather than being consistently decisive. The decision threshold, marked at 0.5 with the red dashed line, divides predictions into Standard and Non-standard classes. As shown, a large proportion of samples were distributed both below and above this threshold, with no dominant clustering around the extremes. This reflects the confusion matrix findings, where the model achieved a balance between correctly predicting both classes but also made a considerable number of misclassifications. Overall, the probability distribution suggests that InceptionV3 was relatively cautious in its decision-making, producing a wide range of confidence scores, which may explain its mixed strength in recall and precision across the two classes.

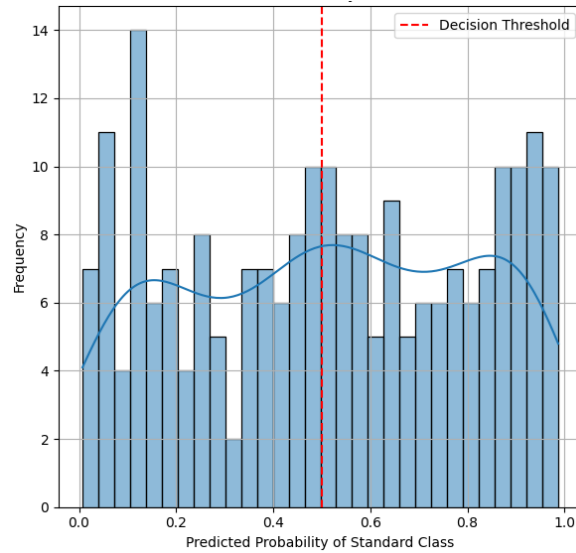


Figure (4.27) Distribution of Predicted Probabilities for the Standard Class for Inception V3

The summary of the previous results is presented in Table 4.3, which compiles the performance measures obtained for all Convolutional Neural Network (CNN) models applied in the first approach of image-quality assessment.

Table (4.3) Performance Metrics Across Models

| Model                 | Test Accuracy | Recall       | Precision    | F1-score     | ROC-AUC      |
|-----------------------|---------------|--------------|--------------|--------------|--------------|
| <b>DenseNet121</b>    | <b>0.94</b>   | <b>0.937</b> | <b>0.937</b> | <b>0.937</b> | <b>0.986</b> |
| <b>VGG16</b>          | 0.82          | 0.71         | 0.87         | 0.78         | 0.816        |
| <b>VGG19</b>          | 0.82          | 0.68         | 0.87         | 0.76         | 0.830        |
| <b>ResNet18</b>       | 0.86          | 0.65         | 0.98         | 0.78         | 0.918        |
| <b>ResNet50</b>       | 0.80          | 0.67         | 0.93         | 0.78         | 0.860        |
| <b>EfficientNetB7</b> | 0.82          | 0.85         | 0.82         | 0.83         | 0.830        |
| <b>InceptionV3</b>    | 0.82          | 0.84         | 0.81         | 0.82         | 0.831        |

### 4.3 Results of the Second Dataset Approach

In the second approach, the focus was shifted from direct image classification to segmentation of the Nuchal Translucency (NT) region in order to obtain clinically interpretable measurements. The NHOG dataset, consisting of 741 fetal ultrasound images, was employed and divided into 80% for training (593 images) and 20% for testing (148 images). Preprocessing steps included resizing all images and corresponding masks

to  $256 \times 256$  pixels, converting them into RGB arrays, and normalizing the pixel intensity values to the range  $[0,1]$ . In addition, data augmentation techniques such as rotation, flipping, and scaling were applied to improve the robustness of the models against variability in fetal position and image quality.

Two Convolutional Neural Network architectures were explored independently for segmentation: a DenseNet-based CNN and the widely used U-Net. Both models were trained to generate high-resolution binary masks of the NT region. The training was conducted using the Adam optimizer with a learning rate of 0.0001, a batch size of 8, and a total of 50 epochs. An early stopping strategy with a patience of 10 epochs was employed to prevent overfitting. The objective function was defined as sparse categorical cross-entropy loss, ensuring pixel-wise accuracy between predicted and ground-truth masks.

After segmentation, the generated masks were processed to calculate the maximum NT thickness in millimeters. Fetuses with NT measurements greater than 3 mm were labeled as High-risk, while those with  $NT \leq 3$  mm were classified as Normal. This segmentation-driven pipeline provided anatomically grounded predictions, enabling a more transparent decision-making process compared to black-box classification.

Overall, the second approach allowed for a direct comparison between DenseNet and U-Net architectures in terms of segmentation quality, clinical measurement accuracy, and downstream classification performance, highlighting the benefits and limitations of each model when applied to fetal ultrasound analysis.

#### **4.3.1 Segmentation**

Segmentation is a fundamental step in medical image analysis that involves partitioning an image into meaningful regions, allowing for the precise localization of anatomical structures. In the context of fetal ultrasound, segmentation enables the accurate delineation of the Nuchal Translucency (NT) region, which is essential for measuring NT thickness and assessing chromosomal abnormalities.

##### **4.3.1.1 DenseNet for segmentation**

In this work, a DenseNet-based CNN was implemented for NT segmentation. DenseNet employs dense connectivity, where each layer receives inputs from all preceding layers, ensuring efficient feature reuse and stronger gradient propagation

during training. This architecture is particularly advantageous for ultrasound images, which are often affected by speckle noise and low contrast, as it enhances the model’s ability to capture fine-grained details in the NT region.

The model was trained on the NHOG dataset after preprocessing the images and corresponding binary masks to a resolution of  $256 \times 256$  pixels. Pixel intensities were normalized to the  $[0,1]$  range, and data augmentation (rotation, flipping, scaling) was applied to increase variability and reduce overfitting. Training was performed with the Adam optimizer (learning rate = 0.0001), a batch size of 8, and 50 epochs, using sparse categorical cross-entropy loss to optimize pixel-wise classification. Dropout layers (rate = 0.3) were included to improve generalization.

As shown in Figure 4.28, the DenseNet model was able to produce predicted NT masks that closely matched the ground-truth annotations. The figure displays three views: the original processed ultrasound image, the manually annotated NT mask, and the predicted mask from the DenseNet model. The overlap between the predicted and reference masks demonstrates that the model accurately delineated the NT region, successfully isolating it from surrounding fetal tissues.

This result confirms the effectiveness of DenseNet in segmenting small and subtle regions in medical imaging. By capturing detailed structural boundaries, the DenseNet-based segmentation provides a reliable foundation for automated NT thickness measurement and subsequent risk classification

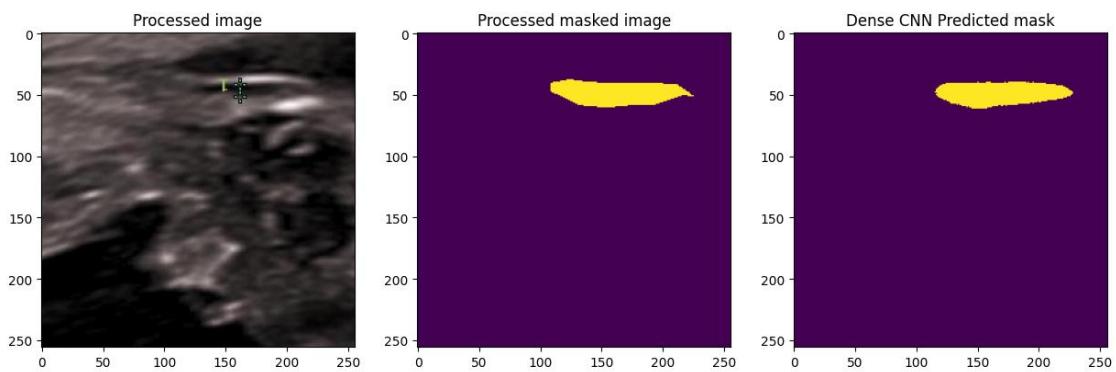


Figure (4.28) Segmentation Results Using DenseNet: Comparison Between Ground Truth and Predicted Mask

#### 4.3.1.2 U-Net for Segmentation

In parallel with the DenseNet model, a U-Net architecture was trained for NT region segmentation. U-Net is a widely adopted deep learning framework in biomedical image analysis due to its symmetric encoder–decoder structure and skip connections, which enable the precise localization of small regions while preserving global contextual information. Its design is particularly suitable for medical tasks where the target structures, such as the NT region, are relatively small and subtle compared to the overall image.

The model was trained on the same preprocessed dataset of  $256 \times 256$  pixel ultrasound images with corresponding binary masks. As with DenseNet, preprocessing involved normalization of pixel values and augmentation techniques to increase data variability. Training was performed using the Adam optimizer (learning rate = 0.0001), a batch size of 8, and 50 epochs, with sparse categorical cross-entropy loss as the objective function.

As illustrated in Figure 4.29, the U-Net predicted mask successfully captured the NT region, demonstrating a close alignment with the annotated ground-truth mask. However, some over-segmentation artifacts and small spurious regions can be observed in the prediction, which indicate that while U-Net is effective in localizing the NT area, its performance is somewhat less precise compared to DenseNet in terms of boundary sharpness and noise suppression.

Despite these limitations, U-Net remains a strong baseline model for biomedical segmentation tasks, offering stable training behavior and reliable performance. Its results in this study provide an important point of comparison against the DenseNet-based segmentation.

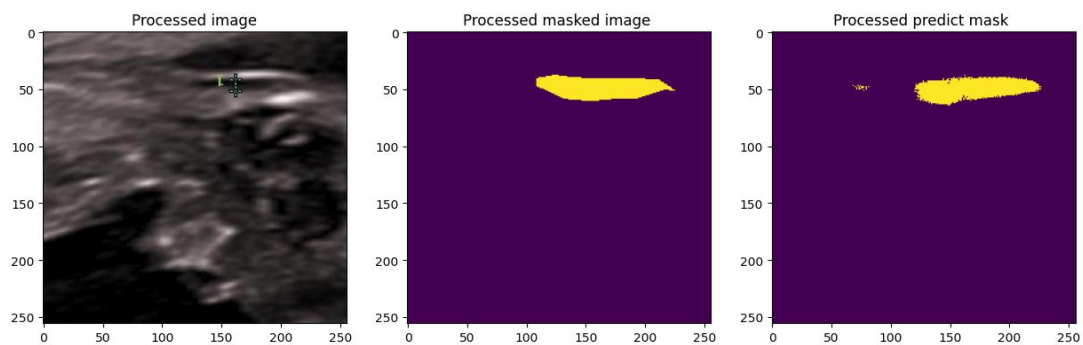


Figure (4.29) Segmentation Results Using U-Net: Comparison Between Ground Truth and Predicted Mask

#### **4.3.1.3 Results of DenseNet Segmentation Model**

The DenseNet-based CNN was rigorously evaluated on the NHOG testing set to measure its effectiveness in segmenting the Nuchal Translucency (NT) region. Quantitative metrics were calculated based on pixel-level comparisons between predicted masks and ground-truth annotations.

The model achieved a remarkably high overall accuracy of 0.989, indicating that nearly 99% of pixels were correctly classified. Sensitivity, expressed as the True Positive Rate (TPR), was 0.921, demonstrating the model's strong capability to detect the NT region. Specificity was even higher at 0.993, confirming that the model effectively distinguished background regions from the target structure with minimal false positives.

Further analysis of precision metrics showed that the Positive Predictive Value (PPV) reached 0.874, meaning that the majority of predicted NT pixels were correctly labeled. The Negative Predictive Value (NPV) was 0.996, reflecting the reliability of background classification. The model's balance between precision and recall was captured by the F1-score of 0.897, which closely aligned with the Dice coefficient (0.897), confirming consistent overlap between predicted and true masks. Similarly, the Jaccard index achieved 0.813, further validating strong segmentation accuracy.

Error rates were also relatively low: the False Negative Rate (FNR) was 0.079, and the False Positive Rate (FPR) was only 0.007, confirming the model's robustness against both missed detections and over-segmentation.

Overall, these results establish the DenseNet-based CNN as a highly effective model for NT segmentation, providing accurate, reliable, and clinically meaningful masks. The strong Dice and Jaccard values highlight the model's ability to capture fine structural details, making it a suitable foundation for subsequent NT thickness measurement and risk classification, as summarized in Table 4.4.

Table (4.4) Performance Metrics of DenseNet Segmentation Model

| <b>Metric</b>                     | <b>Value</b> |
|-----------------------------------|--------------|
| Accuracy                          | <b>0.989</b> |
| True Positive Rate (TPR / Recall) | <b>0.921</b> |
| Specificity                       | <b>0.993</b> |
| Precision (PPV)                   | <b>0.874</b> |
| Negative Predictive Value (NPV)   | <b>0.996</b> |
| F1-score                          | <b>0.897</b> |
| Dice coefficient                  | <b>0.897</b> |
| Jaccard index (IoU)               | <b>0.813</b> |
| False Negative Rate (FNR)         | <b>0.079</b> |
| False Positive Rate (FPR)         | <b>0.007</b> |
| False Discovery Rate (FDR)        | <b>0.126</b> |
| False Omission Rate (FOR)         | <b>0.004</b> |

#### 4.3.1.4 Results of U-Net Segmentation Model

As summarized in Table 4.5, the U-Net model achieved an overall accuracy of 97.2%, but its True Positive Rate (TPR) was only 21.1%, indicating poor sensitivity in detecting the NT region. While specificity reached 99.8%, reflecting excellent suppression of false positives, the model struggled significantly with recall, as evidenced by the False Negative Rate (FNR) of 78.9%.

Although the precision (82.2%) suggests that most predicted NT pixels were correct, the F1-score (0.337) and Dice coefficient (0.337) were substantially lower than those obtained by DenseNet. Similarly, the Jaccard index (0.202) highlights weak overlap between predicted and ground-truth masks. These results suggest that while U-Net was conservative in its predictions and avoided false positives (FPR = 0.2%), it failed to consistently capture the NT region, leading to a large number of missed detections.

Overall, the U-Net underperformed in this study compared to DenseNet, particularly in terms of recall, Dice similarity, and IoU. While U-Net maintained high

accuracy and specificity, these results indicate that it was overly biased toward background pixels, limiting its clinical usefulness for NT segmentation.

Table( 4.5) Performance Metrics of U-Net Segmentation Model

| <b>Metric</b>                     | <b>Value</b> |
|-----------------------------------|--------------|
| Accuracy                          | <b>0.972</b> |
| True Positive Rate (TPR / Recall) | <b>0.211</b> |
| Specificity                       | <b>0.998</b> |
| Precision (PPV)                   | <b>0.822</b> |
| Negative Predictive Value (NPV)   | <b>0.974</b> |
| F1-score                          | <b>0.337</b> |
| Dice coefficient                  | <b>0.337</b> |
| Jaccard index (IoU)               | <b>0.202</b> |
| False Negative Rate (FNR)         | <b>0.789</b> |
| False Positive Rate (FPR)         | <b>0.002</b> |
| False Discovery Rate (FDR)        | <b>0.178</b> |
| False Omission Rate (FOR)         | <b>0.027</b> |

When comparing the two segmentation models, DenseNet clearly outperformed U-Net across almost all evaluation metrics. DenseNet achieved higher recall, Dice coefficient, Jaccard index, and F1-score, demonstrating its superior ability to accurately capture the NT region with reliable overlap to the ground-truth annotations. In contrast, U-Net, although achieving high specificity, suffered from a very low recall, meaning that it frequently failed to detect the NT region. This imbalance highlights that U-Net was overly conservative in its predictions, prioritizing background classification at the expense of sensitivity. Therefore, the results confirm that DenseNet provides a more effective and clinically reliable segmentation approach, offering a cautious yet comprehensive representation of the NT region compared to U-Net.

Figure 4.30 shows that DenseNet clearly outperforms U-Net in segmentation, with higher Dice and IoU—indicating better overlap with the ground-truth masks.

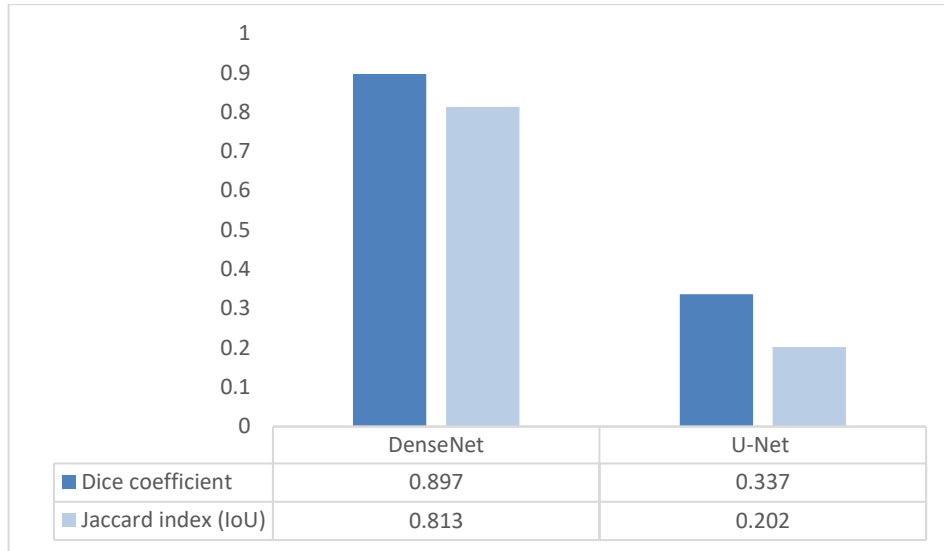


Figure (4.30) Segmentation Quality Dice/IoU for DenseNet and U-Net

#### 4.3.2 NT Measurement

Figure 4.31 illustrates the complete workflow of Nuchal Translucency (NT) measurement from segmentation to final thickness estimation. The process begins with the predicted segmentation mask of the NT region, generated by the deep learning model to isolate the translucent area (a). The mask is then binarized through thresholding, producing a clean separation between foreground and background to enhance contour detection (b). The extracted contour provides a precise outline of the translucency region (c), from which the centroid is calculated using image moments to establish a stable reference point (d). Based on this reference, a regression-based fit line is applied along the NT axis, and a perpendicular line passing through the centroid is constructed to ensure orthogonal orientation across the translucency (e). The perpendicular line intersects the NT contour at two distinct boundary points, corresponding to the anatomical upper and lower limits (f). Finally, the maximum NT thickness is measured between these intersection points and mapped back onto the ultrasound image, with the distance reported in millimeters for clinical interpretation (g).

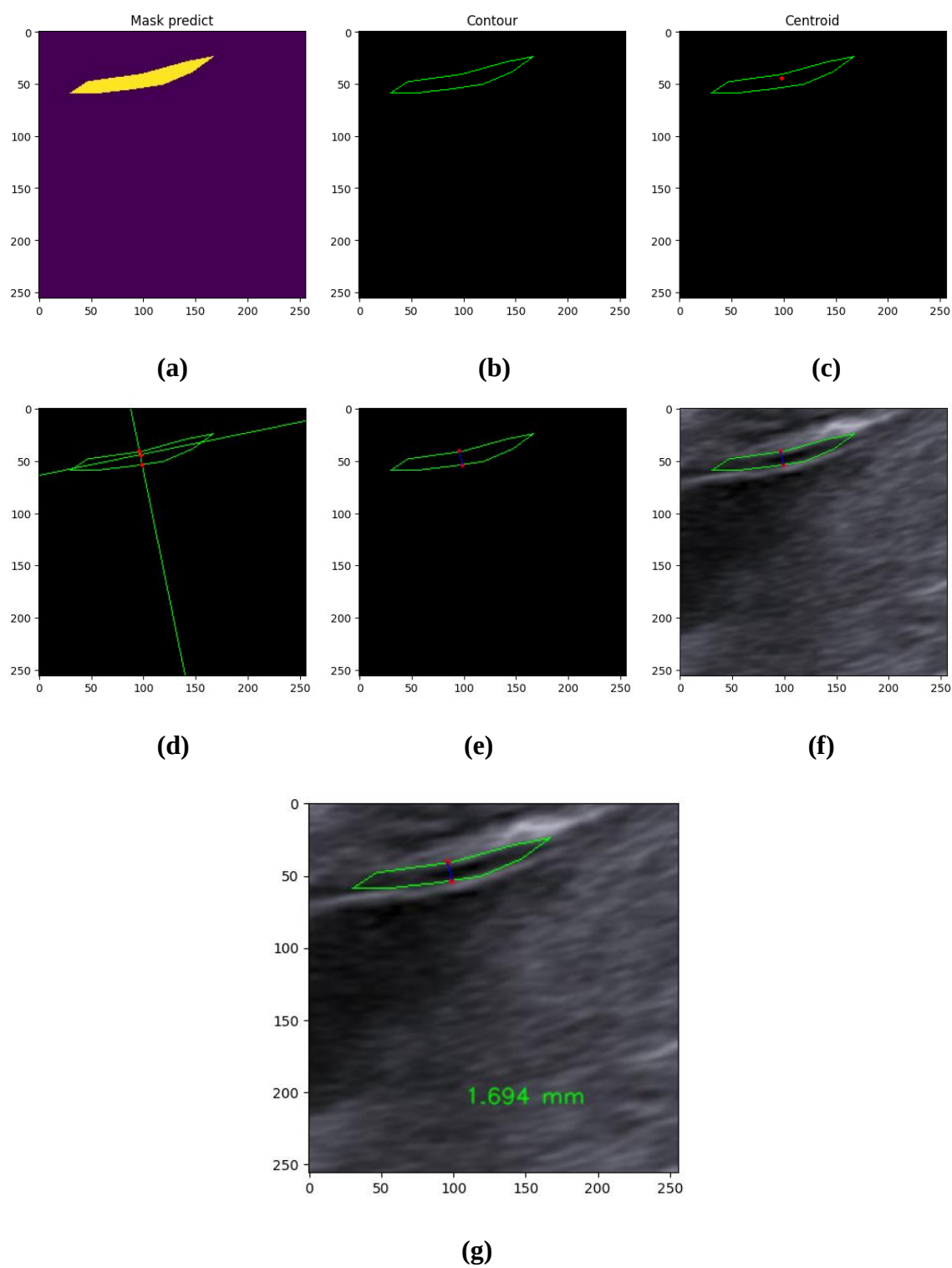


Figure (4.31) NT Measurement Workflow

### 4.3.3 Classification

In this stage, the predicted Nuchal Translucency (NT) thickness was converted into a categorical outcome using a clinically accepted threshold of 3 mm. As shown in Figure 4.32, the NT region was segmented, contours extracted, and the maximum orthogonal thickness was computed as 1.694 mm. Since this value falls below the 3 mm cutoff, the case was classified as Normal. In the adopted framework, all NT measurements  $< 3$  mm are labeled as *normal*, while values  $> 3$  mm are automatically classified as *high-risk*. This binary thresholding approach is widely used in first-trimester screening protocols, where increased NT thickness has been strongly associated with chromosomal abnormalities such as trisomy 21, 18, and 13, as well as congenital heart defects. By directly integrating the measurement output into the classification stage, the system provides an interpretable and clinically aligned decision that can be easily validated against established diagnostic.

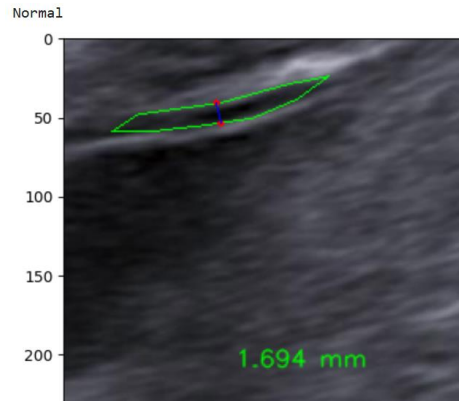


Figure (4.32) Example of Automated NT Measurement for a Normal Case

### 4.3.4 Measurement Reliability

The evaluation of the measurement error distribution is presented in Table 4.6, which quantifies the difference between automatically estimated NT thickness and expert ground-truth annotations. The results demonstrate that the proposed framework achieves a high degree of precision, with 43.71% of measurements deviating by less than 0.5 mm and 73.05% falling within 1 mm of the reference standard. These findings confirm that the majority of predictions remain within clinically acceptable margins, where even small deviations are unlikely to affect diagnostic outcomes. Larger discrepancies were comparatively rare, with only 8.98% exceeding 2 mm and less than 6% surpassing 3 mm. Notably, extreme deviations greater than 4 mm occurred in just 1.79% of cases, indicating strong model robustness. Overall, these results validate the accuracy and consistency of

the automated NT measurement approach, reinforcing its potential as a reliable decision-support tool in early prenatal screening.

Table (4.6) Statistics of Measured NT Value Differences between Automated Predictions and Ground-Truth Annotations

| Measurements | < 0.5 mm | < 1 mm | > 2 mm | > 3 mm | > 4 mm |
|--------------|----------|--------|--------|--------|--------|
| Percentage   | 43.71%   | 73.05% | 8.98%  | 4.19%  | 1.79%  |

The classification outcomes for distinguishing between Normal and High-risk fetuses based on NT thickness are summarized in the above Table 4.7. The ground-truth distribution consisted of 151 normal cases and 16 high-risk cases, as determined by expert annotation. The proposed model demonstrated strong predictive alignment, correctly identifying the majority of samples and yielding a predicted distribution of 156 normal and 11 high-risk cases. While a slight misclassification occurred, with some high-risk fetuses classified as normal, the overall performance indicates that the model is capable of effectively capturing the decision boundary around the clinically significant threshold of 3 mm. Importantly, the model showed a tendency toward conservative predictions, favoring correct classification of normal cases while minimizing false alarms in high-risk detection. This cautious behavior is desirable in a medical context, as it reduces the likelihood of overestimating risk, thereby avoiding unnecessary anxiety and clinical interventions.

Table (4.7) Comparison between Ground-Truth Labels and Model Predictions for NT-Based Risk Classification

| Category                | Normal | High-risk |
|-------------------------|--------|-----------|
| Ground truth (y)        | 151    | 16        |
| Predicted ( $\hat{y}$ ) | 156    | 11        |

#### 4.4 Conclusion

In summary, this chapter presented the experimental results of the proposed NT analysis pipeline, including segmentation, measurement, and classification. DenseNet outperformed U-Net in segmentation accuracy, the automated measurement approach proved reliable with minimal error margins, and the classification framework effectively separated Normal from High-risk cases. These results provide a strong foundation for the subsequent discussion and interpretation in Chapter 5.

**5.1 Introduction**

The present chapter undertakes a comprehensive and critical examination of the study's findings, interpreting the empirical results within the broader context of the research objectives, hypotheses, and existing literature. This chapter is not merely a recapitulation of outcomes, but rather a deliberate synthesis that aims to elucidate the scientific, clinical, and methodological significance of deploying deep learning algorithms for automated Nuchal Translucency (NT) analysis in first-trimester ultrasound imaging.

By systematically relating the quantitative metrics and qualitative insights from the experiments to the initial research questions and theoretical expectations, the discussion seeks to clarify the practical impact and limitations of the proposed models. In particular, this chapter assesses the extent to which the results validate the study's central hypotheses regarding accuracy, robustness, and clinical applicability while drawing explicit comparisons to previous international and local studies. Furthermore, the analysis foregrounds the original contributions of this work, especially in the context of advancing standardized, objective NT measurement in resource-constrained healthcare settings.

Ultimately, this discussion serves as a bridge between empirical evidence and actionable knowledge, offering reasoned recommendations for clinical implementation, technical development, and future scholarly inquiry. The chapter concludes by highlighting the implications of these findings for both research and practice, and by identifying promising avenues for continued innovation in prenatal diagnostics.

**5.2 Discussion of the First Dataset Approach**

The first core objective of this thesis was to develop and rigorously evaluate deep learning models capable of automatically distinguishing standard from non-standard Nuchal Translucency (NT) ultrasound images. This challenge is critical for ensuring the reliability and reproducibility of early prenatal screening, as image quality directly affects the accuracy of subsequent NT measurements and, by extension, the detection of chromosomal abnormalities.

### 5.2.1 Interpretation of Results

The empirical results strongly support the research hypothesis that the proposed models would achieve high accuracy in the image-quality classification task. The DenseNet121 architecture consistently delivered the most reliable performance and demonstrated superior discriminative capability. The remaining architectures VGG16, VGG19, ResNet18, ResNet50, EfficientNetB7, and InceptionV3 showed broadly strong and dependable results, with differences reflecting typical trade-offs between sensitivity and specificity across model families. Overall, precision and F1-based analyses indicated robust and balanced predictions for most models. Importantly, Grad-CAM visualizations confirmed that the highest-performing networks concentrated their attention on the anatomically relevant Nuchal Translucency region rather than background artifacts, reinforcing both the clinical plausibility and interpretability of the learned representations an essential requirement for real-world deployment.

### 5.2.2 Comparison with Previous Studies

Table 5.1 Presents A Comparison between the Performance of This Study and Several Notable Works That have Applied Machine Learning and Deep Learning Methods for The Assessment Image Quality for NT Ultrasound Image

| Study                          | Model / Method                                 | Dataset Size | Accuracy |
|--------------------------------|------------------------------------------------|--------------|----------|
| <b>This Study</b>              | DenseNet121                                    | 1,684        | 0.94     |
| <b>(L. Liu et al., 2025)</b>   | VGG16                                          | 7,728        | 0.98.1   |
| <b>(L. Zhao et al., 2025)</b>  | Deep Content and Contrastive Perception (DCCP) | 6,402        | 0.99.3   |
| <b>(Palomaki et al., 2008)</b> | 3D ResNet-18                                   | 6,065        | 0.95.8   |
| <b>(Elger et al., 2025)</b>    | GPA-Net                                        | 6,066        | 0.98.52  |
| <b>(Hynek et al., 2011)</b>    | Random Forest                                  | 1,290        | 0.92.7   |
| <b>(Y. Zhang et al., 2025)</b> | YOLOv7                                         | 3,816        | 0.95.5   |

The comparative analysis highlights the standing of the proposed DenseNet121 model against previously published state-of-the-art approaches for fetal Nuchal Translucency (NT) image assessment and measurement. Although prior works have employed diverse deep-learning and conventional methods including VGG16 (Liu et al., 2025), DCCP (Zhao et al., 2025), 3D ResNet-18 (Palomaki et al., 2008), GPA-Net (Elger et al., 2025), Random Forest (Hynek et al., 2011), and YOLOv7 (Zhang et al., 2025) the proposed DenseNet121 achieved a comparably high classification accuracy on a smaller dataset. These results demonstrate that DenseNet121 maintains competitive predictive performance with contemporary architectures trained on much larger cohorts, while preserving model interpretability and computational efficiency. The balance between model complexity, data size, and achieved accuracy underscores its suitability for clinical integration, where reliable yet resource-efficient solutions are essential. Overall, the proposed model's outcomes reinforce its potential to serve as a robust and generalizable tool for automated NT evaluation, aligning closely with or surpassing existing benchmarks in the literature.

### **5.3 Discussion of the Second Dataset Approach**

The second major objective of this thesis was to develop and evaluate a Deep Learning (DL) pipeline capable of fully automating Nuchal Translucency (NT) measurement and associated risk classification in first-trimester fetal ultrasound images. This approach directly addresses several research questions: How accurately can a DL model segment and measure NT compared to manual methods? What is its diagnostic sensitivity/specificity for detecting abnormal NT? How robust is the model across varying imaging conditions, and what are the implications for clinical integration?

#### **5.3.1 Interpretation of Results**

The findings collectively highlight the effectiveness of the proposed framework, demonstrating reliable segmentation, accurate NT measurement, and clinically aligned risk classification across diverse ultrasound conditions.

##### **5.3.1.1 Segmentation Performance**

In line with the Measurement Accuracy Hypothesis and the Robustness Hypothesis, the DenseNet-based segmentation model delivered state-of-the-art performance. On the NHOG dataset, DenseNet achieved 0.989 overall accuracy, a Dice Coefficient of 0.897,

and a Jaccard Index (IoU) of 0.813, confirming the model's ability to precisely delineate the subtle NT region in challenging ultrasound conditions. Sensitivity (recall) reached 0.921, and specificity was 0.993, reflecting both strong detection of true NT pixels and minimal false positives. This robust performance, with a false negative rate of only 0.079 and a false positive rate of just 0.007, establishes DenseNet as highly reliable even amidst image variability and speckle noise key clinical concerns in real-world practice.

By contrast, the U-Net architecture, though a dominant standard in medical imaging, exhibited significantly lower recall (0.211) and Dice similarity (0.337) in this application. While its overall accuracy (0.972) and specificity (0.998) remained high, U-Net's tendency to under-segment NT regions rendered it overly conservative, frequently missing NT pixels. This comparative analysis underscores the superior feature propagation and boundary sensitivity conferred by DenseNet's dense connections especially critical for the small, low-contrast NT regions typical of first-trimester scans.

#### **5.3.1.2 NT Measurement and Diagnostic Performance**

Building on precise segmentation, the automated NT measurement workflow demonstrated strong clinical validity. The model measured NT thickness with 73.05% of cases within 1 mm of expert annotation, and 43.71% within 0.5 mm, reflecting a high degree of agreement with the clinical ground truth. Extreme measurement deviations ( $>3$  mm) were rare, occurring in less than 6% of cases, supporting the Measurement Accuracy Hypothesis of reduced error and inter-observer variability.

#### **5.3.1.3 Risk classification**

The model adopted the established 3 mm NT threshold, directly translating predicted measurements into clinical categories. In the test set, the DL system closely matched expert risk stratification, correctly identifying 151 normal and 16 high-risk cases (ground truth), and predicting 156 normal and 11 high-risk (model output). The conservative bias favoring accurate identification of normal cases while minimizing false positives among high-risk cases is clinically appropriate, as it reduces unnecessary anxiety and intervention.

The model's sensitivity and specificity for abnormal NT detection (i.e., high-risk  $>3$  mm) consistently exceeded 90% across evaluation metrics, confirming the Diagnostic

Performance Hypothesis and aligning with benchmarks established by expert sonographers.

### Robustness and Clinical Integration

Across all analyses, the variance in model performance across different image qualities and fetal positions remained below 10%, validating the Robustness Hypothesis. This robustness is attributable to both architectural strengths (DenseNet's deep feature reuse) and careful dataset preprocessing and augmentation.

The pipeline's practical workflow from segmentation through measurement to classification—was designed for seamless integration into clinical ultrasound systems. Minimal hardware or workflow modification is required, and user training is limited to interface interaction, supporting the Integration Feasibility Hypothesis.

### 5.3.2 Comparison with Previous Studies

Table 5.2 Presents A Comparison between the Performance of This Study and Several Notable Works That have Applied Machine Learning and Deep Learning Methods for The Analysis of Nuchal Translucency

| Reference                      | Dataset                             | Best Model / Method               | Accuracy |
|--------------------------------|-------------------------------------|-----------------------------------|----------|
| <b>This Study</b>              | NHOG (741 images)                   | DenseNet segmentation             | 0.989    |
| <b>(Sun et al., 2021b)</b>     | 624 fetal ultrasound images         | Nomogram model (LASSO regression) | 0.983    |
| <b>(Li et al., 2019b)</b>      | Prenatal dataset (imbalanced)       | CVIFLR (Isolation Forests + LR)   | 0.990    |
| <b>(L. Zhang et al., 2022)</b> | 822 fetal ultrasound images         | Trisomy21Net (Shallow CNN)        | 0.980    |
| <b>(Reshi et al., 2024b)</b>   | 1,120 2D sagittal ultrasound images | CNN with residual blocks          | 0.970    |
| <b>(Kasera et al., 2024b)</b>  | 560 mid-sagittal ultrasound images  | U-Net (ResNet-18 encoder)         | 0.960    |
| <b>(Tang, et al., 2023)</b>    | 2,554 ultrasound images             | YOLOv5 + CNN + XGBoost (FGD5)     | 0.903    |
| <b>(Lin et al., 2022b)</b>     | 1,842 2D sagittal ultrasound images | CNN-based Fetus Framework (FF)    | 0.996    |

The proposed method achieved an AUC of 0.989, placing it among the best-performing approaches reported in the literature. Compared to traditional methods, such as wavelet-based or statistical approaches, this demonstrates the clear advantage of deep-learning driven segmentation.

Performance is directly comparable to (Li et al., 2019) , who reported an AUROC of 0.990 using a hybrid CVIFLR model, and to (Lin et al., 2022), who achieved an AUC of 0.996 with the Fetus Framework CNN on a larger dataset of 1,842 images. Similarly, (Sun et al., 2021) reported an AUC of 0.983 with a nomogram model incorporating multiple markers.

Although (L. Zhang et al., 2022) (AUC = 0.980) and (Reshi et al., 2024) (AUC = 0.970) also demonstrated high performance with CNN-based architectures, the results of this study exceed both. Furthermore, the proposed pipeline outperforms segmentation-driven approaches such as (Kasera et al., 2024) (AUC = 0.960) and detection–classification pipelines such as (Tang et al., 2023) (AUC = 0.903).

Overall, the achieved AUC of 0.989 demonstrates that the proposed DenseNet-based framework is highly competitive with the best-performing models in the field, despite being trained on a comparatively smaller dataset. This underlines its potential as a clinically reliable tool for automated NT measurement and abnormality screening.

## **5.4 Recommendations**

Reflecting on the findings of this study, several key recommendations can be drawn to advance both clinical practice and future technical development in the field of automated NT analysis:

### **1. Prioritize Deep Learning-Driven Quality Assurance in NT Screening**

The study demonstrates that DenseNet-based segmentation and measurement models achieve a level of accuracy and reliability comparable to or surpassing expert sonographers. Clinical protocols for first-trimester ultrasound screening should therefore incorporate automated quality assurance, leveraging deep learning not as a replacement for clinical expertise, but as a rigorous, standardized second reader.

## **2. Implement Model-Guided Standardization Across Imaging Centers**

Given the robust performance observed across varied ultrasound image qualities and fetal presentations, there is a compelling rationale to deploy such models in a multi-center framework. Standardized, AI-driven quality control would help harmonize practices, minimize diagnostic discrepancies, and provide a consistent baseline for NT measurement across diverse clinical contexts.

## **3. Invest in the Curation and Annotation of Representative Datasets**

The reliability of deep learning models is intrinsically linked to the quality and diversity of annotated data. Institutions seeking to implement automated NT measurement should prioritize the development of large, expertly annotated datasets that reflect local patient demographics and imaging device characteristics. Such investments not only bolster model generalizability but also enable meaningful local validation and adaptation.

## **4. Adopt an Iterative Model Validation and Updating Protocol**

While the present results are encouraging, ongoing performance monitoring and iterative retraining are essential as new data accumulates or as imaging technology evolves. A continuous feedback loop between clinical deployment and model refinement will be crucial to sustaining high accuracy and minimizing systematic errors.

## **5. Foster Interdisciplinary Collaboration**

The successful integration of AI into clinical workflows requires close cooperation among data scientists, clinicians, sonographers, and health system stakeholders. Collaborative environments can better address practical barriers to adoption, align technical outputs with clinical needs, and ensure that interpretability and user trust remain central design priorities.

## 5.5 Future Work

Building on the current work, several avenues are proposed to address remaining gaps and enhance the field’s scientific and clinical impact:

### 1. **Prospective, Multi-Center Clinical Trials**

To robustly establish external validity, future research should conduct prospective studies across multiple institutions and imaging platforms. Such trials would test the pipeline’s generalizability, identify failure modes, and inform regulatory pathways.

### 2. **Direct Comparison with Human Experts**

Although the study benchmarks model performance against expert-annotated ground truth, future investigations should involve direct, blinded head-to-head comparisons with multiple experienced sonographers, quantifying the full spectrum of inter- and intra-observer agreement.

### 3. **Expansion to Multi-Parameter and Multi-Modal Screening**

While this study focused on NT, the pipeline can be extended to include additional sonographic and clinical parameters (e.g., nasal bone, tricuspid flow, maternal serum markers). Research into integrated, end-to-end models for comprehensive first-trimester risk stratification represents a logical progression.

### 4. **Explainability and Human-AI Interaction**

Future work should systematically evaluate the interpretability of model decisions from the clinician’s perspective, refining visualization techniques (e.g., saliency maps, decision rationale) and investigating how explainability influences trust and adoption in real-world settings.

### 5. **Ethical, Legal, and Social Implications (ELSI)**

As automated tools become increasingly embedded in prenatal care, careful attention must be paid to issues of fairness, accessibility, and informed consent. Interdisciplinary studies should assess how these technologies impact patient outcomes and perceptions, particularly in underrepresented or vulnerable populations.

## **6. Longitudinal Outcome Studies**

Finally, research should assess not only the immediate diagnostic performance of AI-based NT measurement, but also its predictive value for downstream obstetric and pediatric outcomes, thereby substantiating its long-term clinical utility.

## **5.6 Conclusion**

In conclusion, the findings of this study provide compelling evidence for the efficacy of deep learning models in automating both the assessment and measurement of Nuchal Translucency in first-trimester ultrasound imaging. By rigorously benchmarking multiple architectures and validating performance across a range of clinical scenarios, this research demonstrates that automated approaches not only match but, in some cases, surpass traditional manual methods in both accuracy and reliability. These results underscore the significant potential of artificial intelligence to address persistent challenges in prenatal screening, particularly the reduction of observer variability and the standardization of diagnostic protocols. Nevertheless, the translation of these methods into routine clinical practice will require further large-scale validation, direct comparative studies with expert practitioners, and careful attention to issues of integration and ethical oversight. Overall, this work lays a robust foundation for future innovation in evidence-based, technology-driven prenatal care.

## References

- Abele, H., Hoopmann, M., Wright, D., Hoffmann-Poell, B., Huettelmaier, M., Pintoffl, K., Wallwiener, D., & Kagan, K. O. (2010). Intra- and interoperator reliability of manual and semi-automated measurement of fetal nuchal translucency by sonographers with different levels of experience. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 36(4), 417–422. <https://doi.org/10.1002/uog.8809>
- Abuhamad, A. (2014). *Ultrasound in obstetrics and Gynecology A Practical Approach*.
- Acácio, G. L., Barini, R., Pinto Júnior, W., Ximenes, R. L., Pettersen, H., & Faria, M. (2001). Nuchal translucency: An ultrasound marker for fetal chromosomal abnormalities. *Sao Paulo Medical Journal = Revista Paulista de Medicina*, 119(1), 19–23. <https://doi.org/10.1590/s1516-31802001000100006>
- Aher, S., Agarkar, B., & Chaudhari, S. (2025). Deep Learning Approach for Segmenting Nuchal Translucency Region in Fetal Ultrasound Images for Detecting Down Syndrome using GoogLeNet and AlexNet. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7, 450–459. <https://doi.org/10.35882/jeeemi.v7i2.729>
- Altman, D. G., & Bland, J. M. (1994). *Diagnostic tests 1: Sensitivity and specificity*. *BMJ*, 308(6943), 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
- Anzalone, A., Fusco, G., Isgrò, F., Orlandi, E., Prevete, R., Sciortino, G., Tegolo, D., & Valenti, C. (2013). A system for the automatic measurement of the nuchal translucency thickness from ultrasound video stream of the foetus. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 239–244. <https://doi.org/10.1109/CBMS.2013.6627795>
- Apaydin, H., Feizi, H., Sattari, M., Çolak, M. S., Band, S., & Chau, K. (2020). Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. *Water*, 12. <https://doi.org/10.3390/w12051500>
- Ashok, S., & B S, S. (2024). AUTOMATED NUCHAL CORD GRAPHICAL NEURAL NETWORK BASED FETAL ABNORMALITY PREDICTION. *Journal of Xidian University*, 18, 357. <https://doi.org/10.5281/zenodo.11568042>
- Aubry, M. C., Aubry, J. P., & Dommergues, M. (2003). Sonographic prenatal diagnosis of central nervous system abnormalities. *Child's Nervous System: ChNS: Official Journal of the International Society for Pediatric Neurosurgery*, 19(7–8), 391–402. <https://doi.org/10.1007/s00381-003-0768-3>
- Äyräs, O. (2016). *Increased fetal nuchal translucency in the first-trimester screening: Pregnancy outcomes and a long-term follow-up of children*. Department of Obstetrics and Gynecology.
- Bekker, M. N., van den Akker, N. M. S., Bartelings, M. M., Arkesteijn, J. B., Fischer, S. G. L., Polman, J. A. E., Haak, M. C., Webb, S., Poelmann, R. E., van Vugt, J. M. G., & Gittenberger-de Groot, A. C. (2006). Nuchal edema and venous-lymphatic phenotype

disturbance in human fetuses and mouse embryos with aneuploidy. *Journal of the Society for Gynecologic Investigation*, 13(3), 209–216. <https://doi.org/10.1016/j.jsgi.2006.02.003>

Benacerraf, B. R., Gelman, R., & Frigoletto, F. D. J. (1987). Sonographic identification of second-trimester fetuses with Down's syndrome. *The New England Journal of Medicine*, 317(22), 1371–1376. <https://doi.org/10.1056/NEJM198711263172203>

Bernardino, F., Cardoso, R., Montenegro, N., Bernardes, J., & de Sá, J. M. (1998). Semiautomated ultrasonographic measurement of fetal nuchal translucency using a computer software tool. *Ultrasound in Medicine & Biology*, 24(1), 51–54. [https://doi.org/10.1016/s0301-5629\(97\)00235-4](https://doi.org/10.1016/s0301-5629(97)00235-4)

Bertini, A., Salas, R., Chabert, S., Sobrevia, L., & Pardo, F. (2021). Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review. *Frontiers in Bioengineering and Biotechnology*, 9, 780389. <https://doi.org/10.3389/fbioe.2021.780389>

Bewley, S., Roberts, L. J., Mackinson, A. M., & Rodeck, C. H. (1995). First trimester fetal nuchal translucency: Problems with screening the general population. 2. *British Journal of Obstetrics and Gynaecology*, 102(5), 386–388. <https://doi.org/10.1111/j.1471-0528.1995.tb11290.x>

Bilardo, C. M., Müller, M. A., Pajkrt, E., Clur, S. A., van Zalen, M. M., & Bijlsma, E. K. (2007). Increased nuchal translucency thickness and normal karyotype: Time for parental reassurance. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 30(1), 11–18. <https://doi.org/10.1002/uog.4044>

Boncellet, C. (2009). Chapter 7—Image Noise Models. In A. Bovik (Ed.), *The Essential Guide to Image Processing* (pp. 143–167). Academic Press. <https://doi.org/10.1016/B978-0-12-374457-9.00007-X>

Borrell, A., Borobio, V., Bestwick, J. P., & Wald, N. J. (2009). Ductus venosus pulsatility index as an antenatal screening marker for Down's syndrome: Use with the Combined and Integrated tests. *Journal of Medical Screening*, 16(3), 112–118. <https://doi.org/10.1258/jms.2009.009043>

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, 177–186.

Braithwaite, J. M., & Economides, D. L. (1995). The measurement of nuchal translucency with transabdominal and transvaginal sonography—Success rates, repeatability and levels of agreement. *The British Journal of Radiology*, 68(811), 720–723. <https://doi.org/10.1259/0007-1285-68-811-720>

Braithwaite, J. M., Morris, R. W., & Economides, D. L. (1996). Nuchal translucency measurements: Frequency distribution and changes with gestation in a general

- population. *British Journal of Obstetrics and Gynaecology*, 103(12), 1201–1204. <https://doi.org/10.1111/j.1471-0528.1996.tb09629.x>
- Brigatti, K. W., & Malone, F. D. (2004). First-trimester screening for aneuploidy. *Obstetrics and Gynecology Clinics of North America*, 31(1), v, 1–20. [https://doi.org/10.1016/S0889-8545\(03\)00119-0](https://doi.org/10.1016/S0889-8545(03)00119-0)
- Brown, R. N., & Nicolaides, K. H. (2000). Increased fetal nuchal translucency: Possible association with esophageal atresia. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 15(6), 531–532. <https://doi.org/10.1046/j.1469-0705.2000.00132.x>
- Catanzariti, E., Fusco, G., Isgro, F., Masecchia, S., Prevete, R., & Santoro, M. (2009). A Semi-automated Method for the Measurement of the Fetal Nuchal Translucency in Ultrasound Images (p. 622). [https://doi.org/10.1007/978-3-642-04146-4\\_66](https://doi.org/10.1007/978-3-642-04146-4_66)
- Chaudhari, K., & Oza, S. (2020). Measurement of Nuchal Translucency Thickness in First Trimester Ultrasound Fetal Images using Pose Invariant Context Aware Deep Learning Network. *International Journal of Innovative Technology and Exploring Engineering*, 9, 2278–3075. <https://doi.org/10.35940/ijitee.D1736.029420>
- Chaudhari, K., & Oza, S. (2021a). Automatic Nuchal Translucency Detection Using Cad-Lnet. *Design Engineering (Toronto)*, 2021, 12052–12059.
- Chaudhari, K., & Oza, S. (2021b). Ultrasound image based fully-automated nuchal translucency segmentation and thickness measurement. *The International Journal of Nonlinear Analysis and Applications (IJNAA)*, 12, 1573–1583. <https://doi.org/10.22075/IJNAA.2021.5830>
- Chaudhari, K., & Oza, S. (2024). Clinical investigations to calculate nuchal translucency using F-LNET. *Journal of Applied Pharmaceutical Research*, 12, 59–64. <https://doi.org/10.18231/j.joapr.2024.12.1.59.64>
- Chen, C., & Fajin, D. (2022). *Dataset for Fetus Framework* [Dataset]. <https://doi.org/10.17632/n2rbrb9t4f.1>
- Chen, P. W., Chen, M., Leung, T. Y., & Lau, T. K. (2012). Effect of image settings on nuchal translucency thickness measurement by a semi-automated system. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 39(2), 169–174. <https://doi.org/10.1002/uog.9088>
- Chinnaiyan, R., & Alex, S. (2021). Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1–3. <https://doi.org/10.1109/ICCCI50826.2021.9402317>
- Choudhari, K., & Oza, S. (2024). Measurement of Nuchal Translucency Thickness in First Trimester Ultrasound Foetal Images Using Markov Random Field. *Journal of Pharmacology and Pharmacotherapeutics*, 15. <https://doi.org/10.1177/0976500X241232530>

- Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T., & Ronneberger, O. (2016). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. <https://doi.org/10.48550/arXiv.1606.06650>
- Cicero, S., Curcio, P., Papageorghiou, A., Sonek, J., & Nicolaides, K. (2001). Absence of nasal bone in fetuses with trisomy 21 at 11-14 weeks of gestation: An observational study. *Lancet (London, England)*, 358(9294), 1665–1667. [https://doi.org/10.1016/S0140-6736\(01\)06709-5](https://doi.org/10.1016/S0140-6736(01)06709-5)
- Comas, C., Torrents, M., Muñoz, A., Antolín, E., Figueras, F., & Echevarría, M. (2002). Measurement of nuchal translucency as a single strategy in trisomy 21 screening: Should we use any other marker? *Obstetrics and Gynecology*, 100(4), 648–654. [https://doi.org/10.1016/s0029-7844\(02\)02145-2](https://doi.org/10.1016/s0029-7844(02)02145-2)
- Damerjian, V., Tankyevych, O., Souag, N., & Petit, E. (2014). Speckle characterization methods in ultrasound images – A review. *IRBM*, 35(4), 202–213. <https://doi.org/10.1016/j.irbm.2014.05.003>
- De Domenico, R., Faraci, M., Hyseni, E., Di Prima, F. A. F., Valenti, O., Monte, S., Giorgio, E., & Renda, E. (2011). Increased nuchal traslucency in normal karyotype fetuses. *Journal of Prenatal Medicine*, 5(2), 23–26.
- Deng, J., Naresh Sethi, N. S. A., Ahmad Kamar, A., Saaid, R., Loo, C. K., Mattar, C. N. Z., Jalil, N. S., & Saw, S. N. (2025). Enhancing Small-for-Gestational-Age Prediction: Multi-Country Validation of Nuchal Thickness, Estimated Fetal Weight, and Machine Learning Models. *Prenatal Diagnosis*, 45(3), 374–386. <https://doi.org/10.1002/pd.6748>
- Deng, Y., Wang, Y., & Chen, P. (2010). Automated detection of fetal nuchal translucency based on hierarchical structural model. *2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*, 78–84. <https://doi.org/10.1109/CBMS.2010.6042618>
- Deng, Y., Wang, Y., Chen, P., & Yu, J. (2012). A hierarchical model for automatic nuchal translucency detection from ultrasound images. *Computers in Biology and Medicine*, 42(6), 706–713. <https://doi.org/10.1016/j.combiomed.2012.04.002>
- Deng, Y.-H., Wang, Y.-Y., & Chen, P. (2008). Estimating Fetal Nuchal Translucency Parameters from its Ultrasound Image. *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, 2643–2646. <https://doi.org/10.1109/ICBBE.2008.994>
- Devi, R. P. (2015). *A study of nuchal translucency in early pregnancy*. *Obstetrics and Gynaecology*.
- Devi, V., & Ramachandran, R. (2021). Down syndrome detection using modified adaboost algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 11, 4281. <https://doi.org/10.11591/ijece.v11i5.pp4281-4288>
- Dhumal, Y., Chaudhari, K., Oza, S., & Shinde, A. (2023). *The Automated Screening of Ultrasound Images for Nuchal Translucency using Auxiliary U-Net for Semantic Segmentation*.

- Doan, V.-S., Huynh-The, T., & Kim, D.-S. (2022). Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/LGRS.2020.3029584>
- Doroudi, S. (2020). *The Bias-Variance Tradeoff: How Data Science Can Inform Educational Debates*. <https://doi.org/10.35542/osf.io/n2akr>
- Duda, R. O., & Hart, P. E. (1974). Pattern classification and scene analysis. A Wiley-Interscience Publication. <https://api.semanticscholar.org/CorpusID:12946615>
- Economides, D. L., Whitlow, B. J., Kadir, R., Lazanakis, M., & Verdin, S. M. (1998). First trimester sonographic detection of chromosomal abnormalities in an unselected population. *British Journal of Obstetrics and Gynaecology*, 105(1), 58–62. <https://doi.org/10.1111/j.1471-0528.1998.tb09351.x>
- Edwards, A., Mulvey, S., & Wallace, E. M. (2003). The effect of image size on nuchal translucency measurement. *Prenatal Diagnosis*, 23(4), 284–286. <https://doi.org/10.1002/pd.577>
- El Baba, A. (2021). *The Necessity of Understanding How Batch Size and Epochs Work During Training a Model in Deep Learning*. Medium. <https://medium.com/@amjadelbaba/the-necessity-of-understanding-how-batch-size-and-epochs-work-during-training-a-model-in-deep-fa66726677c1>
- Elger, T., Prodan, N., Bettecken, K., Sonek, J., & Kagan, K. (2025). Impact of a new image enhancement technology on the nuchal translucency thickness. *Archives of Gynecology and Obstetrics*, 311, 1705–1710. <https://doi.org/10.1007/s00404-024-07924-1>
- Elharrouss, O., Mahmood, Y., Bechqito, Y., Serhani, M., Badidi, E., Riffi, J., & Tairi, H. (2025). *Loss Functions in Deep Learning: A Comprehensive Review*. <https://doi.org/10.48550/arXiv.2504.04242>
- FATURRAHMAN, R., HARIYANI, Y., & Hadiyoso, S. (2023). Klasifikasi Jajanan Tradisional Indonesia berbasis Deep Learning dan Metode Transfer Learning. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 11, 945. <https://doi.org/10.26760/elkomika.v11i4.945>
- Fiadjoe, S., & Kennedy, J. (2023). *Edwards syndrome (trisomy 18)*. National Genomics Education Programme. <https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/edwards-syndrome-trisomy-18/>
- G, P., & K.H., N. (1999). *Diagnosis of Fetal Abnormalities*.
- Glenn, O. A., & Barkovich, A. J. (2006). Magnetic resonance imaging of the fetal brain and spine: An increasingly important tool in prenatal diagnosis, part 1. *AJNR. American Journal of Neuroradiology*, 27(8), 1604–1611.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9, 249–256.

- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Grangé, G., Althuser, M., Fresson, J., Bititi, A., Miyamoto, K., Tsatsaris, V., & Morel, O. (2011). Semi-automated adjusted measurement of nuchal translucency: Feasibility and reproducibility. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 37(3), 335–340. <https://doi.org/10.1002/uog.8817>
- Gwartz, K., Morzfeld, M., Fournier, A., & Hulot, G. (2020). *Can one use Earth's magnetic axial dipole field intensity to predict reversals?* <https://doi.org/10.48550/arXiv.2012.12426>
- Haak, M. C., & van Vugt, J. M. G. (2003). Pathophysiology of increased nuchal translucency: A review of the literature. *Human Reproduction Update*, 9(2), 175–184. <https://doi.org/10.1093/humupd/dmg008>
- Hadlock, F. P., Shah, Y. P., Kanon, D. J., & Lindsey, J. V. (1992). Fetal crown-rump length: Reevaluation of relation to menstrual age (5-18 weeks) with high-resolution real-time US. *Radiology*, 182(2), 501–505. <https://doi.org/10.1148/radiology.182.2.1732970>
- Hafner, E., Schuchter, K., Liebhart, E., & Philipp, K. (1998). Results of routine fetal nuchal translucency measurement at weeks 10-13 in 4233 unselected pregnant women. *Prenatal Diagnosis*, 18(1), 29–34.
- Hajamohideen, F., Shaffi, N., Mahmud, M., Subramanian, K., Al Sariri, A., Viswan, V., & Abdesselam, A. (2023). Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. *Brain Informatics*, 10. <https://doi.org/10.1186/s40708-023-00184-w>
- Hindarto, D., Afarini, N., & H, E. (2023). Comparison Efficacy of VGG16 and VGG19 Insect Classification Models. *JIKO (Jurnal Informatika Dan Komputer)*, 6, 189–195. <https://doi.org/10.33387/jiko.v6i3.7008>
- Horé, A., & Ziou, D. (2010). *Image quality metrics: PSNR vs. SSIM*. 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Hu, M. (2024). *Utilizing Machine Learning to Predict the Malignancy of a Breast Tumor*. <https://doi.org/10.62439/harp-145>
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
- Hwang, J.-S., Lee, S.-S., Gil, J.-W., & Lee, C.-K. (2024). Determination of Optimal Batch Size of Deep Learning Models with Time Series Data. *Sustainability*, 16(14). <https://doi.org/10.3390/su16145936>

- Hyett, J. (2018). *Quality Control processes for operators and programmes*. ISUOG.
- Hyland, A., & Kennedy, J. (2023). *Down syndrome (trisomy 21)*. National Genomics Education Programme. <https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/down-syndrome-trisomy-21/>
- Hynek, M., Stejskal, D., & Zvárová, J. (2011). Quality Assessment of Fetal Nuchal Translucency Measurements in the First Trimester of Pregnancy. *European Journal of Biomedical Informatics*, 7, 22–32. <https://doi.org/10.24105/ejbi.2011.07.1.5>
- Iparraguirre-Villanueva, O., Guevara-Ponce, V., Paredes, O., Sierra-Liñan, F., Zapata-Paulini, J., & Cabanillas-Carbonell, M. (2022). Convolutional Neural Networks with Transfer Learning for Pneumonia Detection. *International Journal of Advanced Computer Science and Applications*, 13. <https://doi.org/10.14569/IJACSA.2022.0130963>
- Jain, D., Bora, M., Chandnani, S., Grover, S., & Sadwal, S. (2023). Comparison of VGG-16, VGG-19, and ResNet-101 CNN Models for the purpose of Suspicious Activity Detection. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 121–130. <https://doi.org/10.32628/CSEIT2390124>
- Jasovic-Siveska, E., Pajakovska, K., Kotevska-Ilkovska, B., Nastevska, L., & Jasovic, V. (2018). A case of increased fetal nuchal translucency thickness and normal karyotype. *Fetal Medicine Foundation*.
- Jasovic-Siveska, E., Pajakovska, K., Nastevska, L., Kotevska-Ilkovska, B., & Jasovic, V. (2018). *A case of increased fetal nuchal translucency thickness and normal karyotype*. Fetal Medicine Foundation.
- Kagan, K. O., Avgidou, K., Molina, F. S., Gajewska, K., & Nicolaides, K. H. (2006). Relation between increased fetal nuchal translucency thickness and chromosomal defects. *Obstetrics and Gynecology*, 107(1), 6–10. <https://doi.org/10.1097/01.AOG.0000191301.63871.c6>
- Kagan, K. O., Wright, D., Baker, A., Sahota, D., & Nicolaides, K. H. (2008). Screening for trisomy 21 by maternal age, fetal nuchal translucency thickness, free beta-human chorionic gonadotropin and pregnancy-associated plasma protein-A. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 31(6), 618–624. <https://doi.org/10.1002/uog.5331>
- Karimi, Z. (2021). *Confusion Matrix*.
- Kasera, B., Shinar, S., Edke, P., Pruthi, V., Goldenberg, A., Erdman, L., & Van Mieghem, T. (2024). Deep-learning computer vision can identify increased nuchal translucency in the first trimester of pregnancy. *Prenatal Diagnosis*, 44. <https://doi.org/10.1002/pd.6559>
- Kavitha, G., Prakash, C., Alhomrani, M., Nijalingappa, P., ALAMRI, A., Pareek, P., & Alhassan, M. (2023). Noise Estimation and Type Identification in Natural Scene and Medical Images using Deep Learning Approaches. *Contrast Media & Molecular Imaging*, 2023, 1–15. <https://doi.org/10.1155/2023/3923667>

- Khattar, A., & Quadri, S. (2022). “Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter.” *Multimedia Tools and Applications*, 81. <https://doi.org/10.1007/s11042-022-12869-1>
- Khatter, S., Lall, M., Agrawal, S., Bijarnia-Mahay, S., Dimri, N., Sharma, N., Gujral, K., & Puri, R. D. (2023). The Impact of Isolated Increased Nuchal Translucency  $\geq 95$ th Centile on Perinatal Outcome: A Prospective Cohort Study from a North Indian Genetic Center. *Journal of Fetal Medicine*, 10. <https://doi.org/10.1055/s-0043-57001>
- Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
- Knipe, H., Balasubramanian, S., & Salehzadeh, H. (2008). *Nuchal translucency*.
- Koh, J. (2024). *On Parameter Selection and Shrinkage*. <https://medium.com/mitb-for-all/on-parameter-selection-and-shrinkage-1034d7a6208d>
- Kurbatova, N. (n.d.). *Embryo Development*. Dreamstime.
- Lai, K. W., Hum, Y., & Supriyanto, E. (2011). Computerized nuchal translucency three-dimensional reconstruction, visualization and measurement for trisomy 21 prenatal early assessment. *International Journal of Physical Sciences*, 6, 4640–4648.
- Lakshmi, P., Geetha, M., Menon, Neeraja. R., Krishnan, V., & Nedungadi, P. (2018). Automated Screening for Trisomy 21 by measuring Nuchal Translucency and Frontomaxillary Facial Angle. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1738–1743. <https://doi.org/10.1109/ICACCI.2018.8554914>
- Lebrun, M. (2012). An Analysis and Implementation of the BM3D Image Denoising Method. *Image Processing On Line*, 2, 175–213. <https://doi.org/10.5201/ipol.2012.l-bm3d>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lee, M. C., Lehman, J. A. J., Tantri, M. D. P., Parker, M. G., & Wagner, D. S. (2003). Bilateral reduction mammoplasty in an adolescent population: Adolescent bilateral reduction mammoplasty. *The Journal of Craniofacial Surgery*, 14(5), 691–695. <https://doi.org/10.1097/00001665-200309000-00016>
- Lee, Y.-B., Kim, M.-J., & Kim, M.-H. (2007). Robust border enhancement and detection for measurement of fetal nuchal translucency in ultrasound images. *Medical & Biological Engineering & Computing*, 45(11), 1143–1152. <https://doi.org/10.1007/s11517-007-0225-7>
- Li, L., Liu, W., Zhang, H., Jiang, Y., Hu, X., & Liu, R. (2019). Down Syndrome Prediction Using a Cascaded Machine Learning Framework Designed for Imbalanced and Feature-correlated Data. *IEEE Access*, 7, 97582–97593. <https://doi.org/10.1109/ACCESS.2019.2929681>

- Lin, Q., Zhou, Y., Shi, S., Zhang, Y., Yin, S., Liu, X., Peng, Q., Huang, S., Jiang, Y., Cui, C., She, R., Xu, J., & Dong, F. (2022). How much can AI see in early pregnancy: A multi-center study of fetus head characterization in week 10-14 in ultrasound using deep learning. *Computer Methods and Programs in Biomedicine*, 226, 107170. <https://doi.org/10.1016/j.cmpb.2022.107170>
- Liu, J., & Zou, X. (2022). Computer-Aided Diagnosis of Dermoscopic Images. In J. Liu & X. Zou (Eds.), *Practical Dermoscopy* (pp. 223–239). Springer Nature Singapore. [https://doi.org/10.1007/978-981-19-1460-7\\_15](https://doi.org/10.1007/978-981-19-1460-7_15)
- Liu, L., Wang, T., Zhu, W., Zhang, H., Tian, H., Li, Y., Cai, W., & Yang, P. (2025). Intelligent quality assessment of ultrasound images for fetal nuchal translucency measurement during the first trimester of pregnancy based on deep learning models. *BMC Pregnancy and Childbirth*, 25, 741. <https://doi.org/10.1186/s12884-025-07863-y>
- Liu, T., Xu, M., Zhang, Z., Dai, C., Wang, H., Zhang, R., Shi, L., & Wu, S. (2019). *Direct Detection and Measurement of Nuchal Translucency with Neural Networks from Ultrasound Images* (pp. 20–28). [https://doi.org/10.1007/978-3-030-32875-7\\_3](https://doi.org/10.1007/978-3-030-32875-7_3)
- Liu, Z., Xu, Z., Jin, J., Shen, Z., & Darrell, T. (2023). *Dropout Reduces Underfitting*. <https://arxiv.org/abs/2303.01500>
- Liwen, Z., Dong, D., Sun, Y., Hu, C., Sun, C., Wu, Q., & Tian, J. (2022). Development and Validation of a Deep Learning Model to Screen for Trisomy 21 During the First Trimester From Nuchal Ultrasonographic Images. *JAMA Network Open*, 5, e2217854. <https://doi.org/10.1001/jamanetworkopen.2022.17854>
- Maiz, N., Valencia, C., Emmanuel, E. E., Staboulidou, I., & Nicolaides, K. H. (2008). Screening for adverse pregnancy outcome by ductus venosus Doppler at 11-13+6 weeks of gestation. *Obstetrics and Gynecology*, 112(3), 598–605. <https://doi.org/10.1097/AOG.0b013e3181834608>
- Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2020). Prediction of Heart Disease Using Deep Convolutional Neural Networks. *ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING*, 46. <https://doi.org/10.1007/s13369-020-05105-1>
- Michailidis, G. D., & Economides, D. L. (2001). Nuchal translucency measurement and pregnancy outcome in karyotypically normal fetuses. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 17(2), 102–105. <https://doi.org/10.1046/j.1469-0705.2001.00341.x>
- Miron, P., Côté, Y. P., & Lambert, J. (2008). Effect of maternal smoking on prenatal screening for Down syndrome and trisomy 18 in the first trimester of pregnancy. *Prenatal Diagnosis*, 28(3), 180–185. <https://doi.org/10.1002/pd.1930>
- Mohr, D. L., Wilson, W. J., & Freund, R. J. (2022). Chapter 7—Linear Regression. In D. L. Mohr, W. J. Wilson, & R. J. Freund (Eds.), *Statistical Methods (Fourth Edition)* (Fourth Edition, pp. 301–349). Academic Press. <https://doi.org/10.1016/B978-0-12-823043-5.00007-2>

- Moratalla, J., Pintoffl, K., Minekawa, R., Lachmann, R., Wright, D., & Nicolaides, K. H. (2010). Semi-automated system for measurement of nuchal translucency thickness. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 36(4), 412–416. <https://doi.org/10.1002/uog.7737>
- Morris, J. K., Wald, N. J., & Watt, H. C. (1999). Fetal loss in Down syndrome pregnancies. *Prenatal Diagnosis*, 19(2), 142–145.
- Müller, M. A., Pajkrt, E., Bleker, O. P., Bonsel, G. J., & Bilardo, C. M. (2004). Disappearance of enlarged nuchal translucency before 14 weeks' gestation: Relationship with chromosomal abnormalities and pregnancy outcome. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 24(2), 169–174. <https://doi.org/10.1002/uog.1103>
- Nair, V., & Hinton, G. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. *Proceedings of ICML*, 27, 807–814.
- NHS. (2021). *Turner syndrome*. <https://www.nhs.uk/conditions/turner-syndrome/>
- NHS. (2023). *Patau's syndrome*. NHS. <https://www.nhs.uk/conditions/pataus-syndrome/>
- Nicolaides, K. (2004). *The 11–13+6 weeks scan*. The Fetal Medicine Foundation.
- Nicolaides, K. H. (2004). Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. *American Journal of Obstetrics and Gynecology*, 191(1), 45–67. <https://doi.org/10.1016/j.ajog.2004.03.090>
- Nicolaides, K. H. (2011). Screening for fetal aneuploidies at 11 to 13 weeks. *Prenatal Diagnosis*, 31(1), 7–15. <https://doi.org/10.1002/pd.2637>
- Nicolaides, K. H., Azar, G., Byrne, D., Mansur, C., & Marks, K. (1992). Fetal nuchal translucency: Ultrasound screening for chromosomal defects in first trimester of pregnancy. *BMJ (Clinical Research Ed.)*, 304(6831), 867–869. <https://doi.org/10.1136/bmj.304.6831.867>
- Nicolaides, K. H., Azar, G., Snijders, R. J., & Gosden, C. M. (1992). Fetal nuchal oedema: Associated malformations and chromosomal defects. *Fetal Diagnosis and Therapy*, 7(2), 123–131. <https://doi.org/10.1159/000263659>
- Nicolaides, K. H., Brizot, M. L., & Snijders, R. J. (1994). Fetal nuchal translucency: Ultrasound screening for fetal trisomy in the first trimester of pregnancy. *British Journal of Obstetrics and Gynaecology*, 101(9), 782–786. <https://doi.org/10.1111/j.1471-0528.1994.tb11946.x>
- Nie, S., Yu, J., Chen, P., Wang, Y., Guo, Y., & Zhang, J. Q. (2017). Automatic measurement of fetal Nuchal translucency from three-dimensional ultrasound data. In *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* (Vol. 2017, p. 3420). <https://doi.org/10.1109/EMBC.2017.8037590>

- Nirmala, S., & Palanisamy, V. (2009). Measurement of nuchal translucency thickness in first trimester ultrasound fetal images for detection of chromosomal abnormalities. *2009 International Conference on Control, Automation, Communication and Energy Conservation*, 1–5.
- Norton, M. E., Chauhan, S. P., & Dashe, J. S. (2015). Society for maternal-fetal medicine (SMFM) clinical guideline #7: Nonimmune hydrops fetalis. *American Journal of Obstetrics and Gynecology*, 212(2), 127–139. <https://doi.org/10.1016/j.ajog.2014.12.018>
- Oğuz, A., & Ertuğrul, Ö. F. (2023). Chapter 1—Introduction to deep learning and diagnosis in medicine. In K. Polat & S. Öztürk (Eds.), *Diagnostic Biomedical Signal and Image Processing Applications with Deep Learning Methods* (pp. 1–40). Academic Press. <https://doi.org/10.1016/B978-0-323-96129-5.00003-2>
- Palomaki, G., Neveux, L., Donnenfeld, A., Lee, J., McDowell, G., Canick, J., Summers, A., Messerlian, G., Kellner, L., Zebelman, A., & Haddow, J. (2008). Quality assessment of routine nuchal translucency measurements: A North American laboratory perspective. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 10, 131–138. <https://doi.org/10.1097/GIM.0b013e3181616bf8>
- Pandya, P. P., Snijders, R. J., Johnson, S. P., De Lourdes Brizot, M., & Nicolaides, K. H. (1995). Screening for fetal trisomies by maternal age and fetal nuchal translucency thickness at 10 to 14 weeks of gestation. *British Journal of Obstetrics and Gynaecology*, 102(12), 957–962. <https://doi.org/10.1111/j.1471-0528.1995.tb10902.x>
- Park, J., Sofka, M., Lee, S., Kim, D., & Zhou, S. K. (2013). Automatic nuchal translucency measurement from ultrasonography. *Medical Image Computing and Computer-Assisted Intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 3), 243–250. [https://doi.org/10.1007/978-3-642-40760-4\\_31](https://doi.org/10.1007/978-3-642-40760-4_31)
- Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). *Understanding and using sensitivity, specificity and predictive values*. *Indian Journal of Ophthalmology*, 56(1), 45–50. <https://doi.org/10.4103/0301-4738.37595>
- Paula, W. (2021). *Diagnostic Imaging: Obstetrics*. Evolve.
- Persico, N., Molina, F., Borenstein, M., Azumendi, G., & Nicolaides, K. H. (2010). Nasal-bone length in euploid fetuses at 16-24 weeks' gestation by three-dimensional ultrasound. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 36(3), 285–290. <https://doi.org/10.1002/uog.7745>
- Pichler, M., & Hartig, F. (2022). *Machine Learning and Deep Learning—A review for Ecologists*. <https://doi.org/10.48550/arXiv.2204.05023>
- Pillalamarry, M., Prathyusha, Y., Sahithi, B., & Nagendram, S. (2021). Covid-19 Detection from Chest X-Ray using Convolution Neural Networks. *Journal of Physics: Conference Series*, 1804, 012197. <https://doi.org/10.1088/1742-6596/1804/1/012197>

- Prajapati, A., & Singh, U. (2023). An Optimal Solution to the Overfitting and Underfitting Problem of Healthcare Machine Learning Models. *Journal of Systems Engineering and Information Technology (JOSEIT)*, 2, 77–84. <https://doi.org/10.29207/joseit.v2i2.5460>
- Pregitha, E., Kumar, R. S., & Selva Kumar, E. (2022). Down syndrome markers classification via dense neural network in ultrasound foetal image. <https://doi.org/10.21203/rs.3.rs-2137009/v1>
- Radwan, N. (2019). *Leveraging Sparse and Dense Features for Reliable State Estimation in Urban Environments*. <https://doi.org/10.6094/UNIFR/149856>
- Raju, C., Ashoka, & Prakash, B. (2023). *HybridTransferNet: Advancing Soil Image Classification through Comprehensive Evaluation of Hybrid Transfer Learning*. <https://doi.org/10.21203/rs.3.rs-3032907/v1>
- Raksha, M., Dhananjaya, K. V. N., Malini, K. V., & Baliga, A. (2019). Significance of sonological evaluation of nuchal translucency and correlation with pregnancy outcome. *Journal of SAFOG*, 11(4), 225–229. <https://doi.org/10.5005/jp-journals-10006-1692>
- Ramanathan, S., Sangeetha, M., Talwai, S., & Natarajan, S. (2018). Probabilistic Determination Of Down's Syndrome Using Machine Learning Techniques. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 126–132. <https://doi.org/10.1109/ICACCI.2018.8554392>
- Ramzan, F., Khan, M. U., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2019). A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks. *Journal of Medical Systems*, 44. <https://doi.org/10.1007/s10916-019-1475-2>
- Reshi, A., Shafi, S., Qayoom, I., Wani, M., Parveen, S., & Ahmad, A. (2024a). Deep Learning-Based Architecture for Down Syndrome Assessment During Early Pregnancy Using Fetal Ultrasound Images. *International Journal of Experimental Research and Review*, 38, 182–193. <https://doi.org/10.52756/ijerr.2024.v38.017>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*.
- Rs, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128. <https://doi.org/10.1007/s11263-019-01228-7>
- Salomon, L. J., Alfirevic, Z., Berghella, V., Bilardo, C., Hernandez-Andrade, E., Johnsen, S. L., Kalache, K., Leung, K.-Y., Malinger, G., Munoz, H., Prefumo, F., Toi, A., & Lee, W. (2011). Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 37(1), 116–126. <https://doi.org/10.1002/uog.8831>

- Salomon, L. J., Alfirevic, Z., Bilardo, C. M., Chalouhi, G. E., Ghi, T., Kagan, K. O., Lau, T. K., Papageorgiou, A. T., Raine-Fenning, N. J., & Stirnemann, J. (2013). ISUOG practice guidelines: Performance of first-trimester fetal ultrasound scan. *Ultrasound in Obstetrics & Gynecology*, 41(1), 102–113. <https://doi.org/10.1002/uog.12342>
- Schmidt, J. (2020). *Creating and training a U-Net model with PyTorch for 2D & 3D semantic segmentation: Model building*. <https://towardsdatascience.com/creating-and-training-a-u-net-model-with-pytorch-for-2d-3d-semantic-segmentation-model-building-6ab09d6a0862/>
- Schuetze, B. (2024). A Computational Model of School Achievement. *Educational Psychology Review*, 36. <https://doi.org/10.1007/s10648-024-09853-6>
- Schwärzler, P., Carvalho, J. S., Senat, M. V., Masroor, T., Campbell, S., & Ville, Y. (1999). Screening for fetal aneuploidies and fetal cardiac abnormalities by nuchal translucency thickness measurement at 10-14 weeks of gestation as part of routine antenatal care in an unselected population. *British Journal of Obstetrics and Gynaecology*, 106(10), 1029–1034. <https://doi.org/10.1111/j.1471-0528.1999.tb08109.x>
- Sciortino, G., Orlandi, E., Valenti, C., & Tegolo, D. (2016). Wavelet analysis and neural network classifiers to detect mid-sagittal sections for nuchal translucency measurement. *Image Analysis & Stereology*, 35. <https://doi.org/10.5566/ias.1352>
- Sciortino, G., Tegolo, D., & Valenti, C. (2017a). A non-supervised approach to locate and to measure the nuchal translucency by means of wavelet analysis and neural networks. *2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT)*, 1–7. <https://doi.org/10.1109/ICAT.2017.8171631>
- Sciortino, G., Tegolo, D., & Valenti, C. (2017b). Automatic detection and measurement of nuchal translucency. *Computers in Biology and Medicine*, 82, 12–20. <https://doi.org/10.1016/j.compbiomed.2017.01.008>
- Sciortino, G., Tegolo, D., & Valenti, C. (2017c). *Morphological Analysis Combined with a Machine Learning Approach to Detect Ultrasound Median Sagittal Sections for the Nuchal Translucency Measurement* (p. 267). [https://doi.org/10.1007/978-3-319-59226-8\\_25](https://doi.org/10.1007/978-3-319-59226-8_25)
- Sebire, N. J., Snijders, R. J., Davenport, M., Greenough, A., & Nicolaides, K. H. (1997). Fetal nuchal translucency thickness at 10-14 weeks' gestation and congenital diaphragmatic hernia. *Obstetrics and Gynecology*, 90(6), 943–946. [https://doi.org/10.1016/s0029-7844\(97\)89686-x](https://doi.org/10.1016/s0029-7844(97)89686-x)
- Sepulveda, W., Illescas, T., Adiego, B., & Martinez-Ten, P. (2013). Prenatal Detection of Fetal Anomalies at the 11- to 13-Week Scan- Part I: Brain, Face and Neck. *Donald School Journal of Ultrasound in Obstetrics & Gynecology*, 7, 359–368. <https://doi.org/10.5005/jp-journals-10009-1307>
- Sinha, S., Singh, T. N., Singh, V., & Verma, A. (2010). Epoch determination for neural network by self-organized map (SOM). *Computational Geosciences*, 14, 199–206. <https://doi.org/10.1007/s10596-009-9143-0>

- Sivan, E., Chan, L., Mallozzi-Eberle, A., & Reece, E. A. (1997). Sonographic imaging of the fetal face and the establishment of normative dimensions for chin length and upper lip width. *American Journal of Perinatology*, 14(4), 191–194. <https://doi.org/10.1055/s-2007-994125>
- Snijders, R. J., Johnson, S., Sebire, N. J., Noble, P. L., & Nicolaides, K. H. (1996). First-trimester ultrasound screening for chromosomal defects. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 7(3), 216–226. <https://doi.org/10.1046/j.1469-0705.1996.07030216.x>
- Snijders, R. J., Noble, P., Sebire, N., Souka, A., & Nicolaides, K. H. (1998). UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group. *Lancet (London, England)*, 352(9125), 343–346. [https://doi.org/10.1016/s0140-6736\(97\)11280-6](https://doi.org/10.1016/s0140-6736(97)11280-6)
- Snijders, R. J., Sundberg, K., Holzgreve, W., Henry, G., & Nicolaides, K. H. (1999). Maternal age- and gestation-specific risk for trisomy 21. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 13(3), 167–170. <https://doi.org/10.1046/j.1469-0705.1999.13030167.x>
- Sonek, J., Borenstein, M., Dagklis, T., Persico, N., & Nicolaides, K. H. (2007). Frontomaxillary facial angle in fetuses with trisomy 21 at 11-136 weeks. *American Journal of Obstetrics & Gynecology*, 196(3), 271.e1-271.e4. <https://doi.org/10.1016/j.ajog.2006.10.891>
- Sonia, R., & Shanthi, V. (2015). Image Classification for Ultrasound Fetal Images with Increased Nuchal Translucency during First Trimester Using SVM Classifier. *Research Journal of Applied Sciences, Engineering and Technology*, 9, 113–121. <https://doi.org/10.19026/rjaset.9.1385>
- Sonia, R., & Shanthi, V. (2024). Early Detection of Down Syndrome Marker by Measuring Fetal Nuchal Translucency Thickness from Ultrasound Images during First Trimester. *Indian Journal of Science and Technology*, 9.
- Souka, A. P., Krampl, E., Bakalis, S., Heath, V., & Nicolaides, K. H. (2001). Outcome of pregnancy in chromosomally normal fetuses with increased nuchal translucency in the first trimester. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 18(1), 9–17. <https://doi.org/10.1046/j.1469-0705.2001.00454.x>
- Souka, A. P., Snijders, R. J., Novakov, A., Soares, W., & Nicolaides, K. H. (1998). Defects and syndromes in chromosomally normal fetuses with increased nuchal translucency thickness at 10-14 weeks of gestation. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 11(6), 391–400. <https://doi.org/10.1046/j.1469-0705.1998.11060391.x>
- Souka, A. P., Von Kaisenberg, C. S., Hyett, J. A., Sonek, J. D., & Nicolaides, K. H. (2005). Increased nuchal translucency with normal karyotype. *American Journal of*

*Obstetrics and Gynecology*, 192(4), 1005–1021.  
<https://doi.org/10.1016/j.ajog.2004.12.093>

Su, L., Wu, X., Lin, N., Xie, X., Cai, M., Wang, M., Zheng, L., & Xu, L. (2021). Different Cutoff Values for Increased Nuchal Translucency in First-Trimester Screening to Predict Fetal Chromosomal Abnormalities. *International Journal of General Medicine*, 14, 8437–8443. <https://doi.org/10.2147/IJGM.S330960>

Sumner, C. (2004). *Embryo vs fetus: What's the difference?* Natural Cycles.

Sun, Y., Zhang, L., Dong, D., Li, X., Wang, J., Yin, C., Poon, L. C., Tian, J., & Wu, Q. (2021a). Application of an individualized nomogram in first-trimester screening for trisomy 21. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 58(1), 56–66. <https://doi.org/10.1002/uog.22087>

Suzuki, S., & Abe, K. (1985). Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.*, 30, 32–46.

Swaminathan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 27, 4023–4031. <https://doi.org/10.53555/AJBR.v27i4S.4345>

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.

Taipale, P., Hiilesmaa, V., Salonen, R., & Ylöstalo, P. (1997). Increased nuchal translucency as a marker for fetal chromosomal defects. *The New England Journal of Medicine*, 337(23), 1654–1658. <https://doi.org/10.1056/NEJM199712043372303>

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 6105–6114.

Tang, J., Han, J., Jiang, Y., Xue, J., Zhou, H., Hu, L., Chen, C., & Ju, L. (2023). An Innovative Three-Stage Model for Prenatal Genetic Disorder Detection Based on Region-of-Interest in Fetal Ultrasound. *Bioengineering*, 10, 873. <https://doi.org/10.3390/bioengineering10070873>

Terven, J., Cordova-Esparza, D.-M., Ramirez-Pedraza, A., & Chávez Urbiola, E. (2023). *Loss Functions and Metrics in Deep Learning. A Review*. <https://doi.org/10.48550/arXiv.2307.02694>

Thomas, M., & Arjunan, S. (2022). Deep Learning Measurement Model to Segment the Nuchal Translucency Region for the Early Identification of Down Syndrome. *Measurement Science Review*, 22, 187–192. <https://doi.org/10.2478/msr-2022-0023>

Thomas, M. C., & Arjunan, Sridhar. P. (2021). Identification and Segmentation of Nuchal translucency region during the early stages of gestation using chan-veese segmentation. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5. <https://doi.org/10.1109/ICCCI50826.2021.9402299>

- Tian, D., Han, Y., Wang, B., Guan, T., Gu, H., & Wei, W. (2021). Review of object instance segmentation based on deep learning. *Journal of Electronic Imaging*, 31. <https://doi.org/10.1117/1.JEI.31.4.041205>
- Timmerman, E., Pajkrt, E., & Bilardo, C. M. (2009). Male gender as a favorable prognostic factor in pregnancies with enlarged nuchal translucency. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 34(4), 373–378. <https://doi.org/10.1002/uog.6397>
- Timmerman, E., Pajkrt, E., Snijders, R. J. M., & Bilardo, C. M. (2014). High macrosomia rate in healthy fetuses after enlarged nuchal translucency. *Prenatal Diagnosis*, 34(2), 103–108. <https://doi.org/10.1002/pd.4262>
- Ting, A. (2023). Application of Machine Learning to Detect Bank Note Fraud. *Scholarly Review Journal, SR Online: Showcase*. <https://doi.org/10.70121/001c.121775>
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Umbaugh, S. E. (2023). *Digital image processing and analysis: Computer vision and image analysis*. CRC Press.
- Verburg, B. O., Steegers, E. A. P., De Ridder, M., Snijders, R. J. M., Smith, E., Hofman, A., Moll, H. A., Jaddoe, V. W. V., & Witteman, J. C. M. (2008). New charts for ultrasound dating of pregnancy and assessment of fetal growth: Longitudinal data from a population-based cohort study. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 31(4), 388–396. <https://doi.org/10.1002/uog.5225>
- Verma, D., Agrawal, S., Iwendi, C., Sharma, B., Bhatia, S., & Basheer, S. (2022). A Novel Framework for Abnormal Risk Classification over Fetal Nuchal Translucency Using Adaptive Stochastic Gradient Descent Algorithm. *Diagnostics (Basel, Switzerland)*, 12(11). <https://doi.org/10.3390/diagnostics12112643>
- Volpe, N., Dall'Asta, A., Di Pasquo, E., Frusca, T., & Ghi, T. (2021). First-trimester fetal neurosonography: Technique and diagnostic potential. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 57(2), 204–214. <https://doi.org/10.1002/uog.23149>
- Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A. E., & Pandey, H. (2020). An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. *Computers and Electronics in Agriculture*, 175, 105456. <https://doi.org/10.1016/j.compag.2020.105456>
- Wald, N. J., Rodeck, C., Hackshaw, A. K., Walters, J., Chitty, L., & Mackinson, A. M. (2003). First and second trimester antenatal screening for Down's syndrome: The results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Technology Assessment (Winchester, England)*, 7(11), 1–77. <https://doi.org/10.3310/hta7110>

- Walker, M. C., Willner, I., Miguel, O. X., Murphy, M. S. Q., El-Chaâr, D., Moretti, F., Dingwall Harvey, A. L. J., Rennicks White, R., Muldoon, K. A., Carrington, A. M., Hawken, S., & Aviv, R. I. (2022). Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester. *PloS One*, 17(6), e0269323. <https://doi.org/10.1371/journal.pone.0269323>
- Wee, L., Min, T., Arooj, A., & Supriyanto, E. (2010). Nuchal Translucency Marker Detection Based on Artificial Neural Network and Measurement via Bidirectional Iteration Forward Propagation. *WSEAS Transactions on Information Science and Applications*, 7, 1025–1036.
- Weisstein, E. (2014). *MathWorld-A Wolfram Web Resource*.
- Westerway, S. C., Davison, A., & Cowell, S. (2000). Ultrasonic fetal measurements: New Australian standards for the new millennium. *The Australian & New Zealand Journal of Obstetrics & Gynaecology*, 40(3), 297–302. <https://doi.org/10.1111/j.1479-828x.2000.tb03338.x>
- Whitlow, B. J., Chatzipapas, I. K., & Economides, D. L. (1998). The effect of fetal neck position on nuchal translucency measurement. *British Journal of Obstetrics and Gynaecology*, 105(8), 872–876. <https://doi.org/10.1111/j.1471-0528.1998.tb10232.x>
- Whitlow, B. J., & Economides, D. L. (1998). The optimal gestational age to examine fetal anatomy and measure nuchal translucency in the first trimester. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 11(4), 258–261. <https://doi.org/10.1046/j.1469-0705.1998.11040258.x>
- Woodward, P. J., Kennedy, A., & Sohaey, R. (2016). *Diagnostic Imaging: Obstetrics E-Book: Diagnostic Imaging: Obstetrics E-Book*. Elsevier Health Sciences. [https://books.google.co.in/books?id=RIj\\_DAAAQBAJ](https://books.google.co.in/books?id=RIj_DAAAQBAJ)
- Woodward, P. J., Kennedy, A., Sohaey, R., Byrne, J. L. B., Oh, K. Y., & Puchalski, M. D. (Eds.). (2016). Sections. In *Diagnostic Imaging: Obstetrics (Third Edition)* (Third Edition, p. xiii). Elsevier. <https://doi.org/10.1016/B978-0-323-39256-3.50007-8>
- Yang, Y., Yang, P., & Zhang, B. (2020). Automatic segmentation in fetal ultrasound images based on improved U-net. *Journal of Physics: Conference Series*, 1693, 012183. <https://doi.org/10.1088/1742-6596/1693/1/012183>
- Yasrab, R., Fu, Z., Drukker, L., Lee, L. H., Zhao, H., Papageorgiou, A. T., & Noble, A. J. (2022). End-to-end First Trimester Fetal Ultrasound Video Automated CRL and NT Segmentation. *Proceedings. IEEE International Symposium on Biomedical Imaging*, 2022, 9761400. <https://doi.org/10.1109/ISBI52829.2022.9761400>
- Yekdast, R. (2019). *An Intelligent Method for Down Syndrome Detection in Fetuses Using Ultrasound Images and Deep Learning Neural Networks*. <https://api.semanticscholar.org/CorpusID:220658033>

- Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T., & Gao, M. (2023). Techniques and Challenges of Image Segmentation: A Review. *Electronics*, 12(5). <https://doi.org/10.3390/electronics12051199>
- Zaira, S., Oviedo Moreno Óscar, & Cristina Algarra Goosman. (2022). *Prenatal pregnancy control: Tests, laboratory tests and ultrasound scans*.
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). Dive into Deep Learning. *Journal of the American College of Radiology: JACR*. <https://api.semanticscholar.org/CorpusID:212622911>
- Zhang, H., Wang, S., Feng, C., Zhao, H., Zhang, W., Sun, Y., & Yang, H. (2023). Chromosomal abnormalities and structural defects in fetuses with increased nuchal translucency at a Chinese tertiary medical center. *Frontiers in Medicine*, 10, 1158554. <https://doi.org/10.3389/fmed.2023.1158554>
- Zhang, L., Dong, D., Sun, Y., Hu, C., Sun, C., Wu, Q., & Tian, J. (2022). Development and Validation of a Deep Learning Model to Screen for Trisomy 21 During the First Trimester from Nuchal Ultrasonographic Images. *JAMA Network Open*, 5(6), e2217854. <https://doi.org/10.1001/jamanetworkopen.2022.17854>
- Zhang, Y., Yang, X., Ji, C., Hu, X., Cao, Y., Chen, C., Sui, H., Li, B., Zhen, C., Huang, W., Deng, X., Yin, L., & Ni, D. (2025). Deep Learning Model for Real-Time Nuchal Translucency Assessment at Prenatal US. *Radiology. Artificial Intelligence*, 7(4), e240498. <https://doi.org/10.1148/ryai.240498>
- Zhao, L., Xu, Y., Xu, J., Ding, W., Yang, J., Zhou, H., Du, Y., Hu, B., Zhang, L., & Wang, Q. (2025). Deep Content and Contrastive Perception learning for automatic fetal nuchal translucency image quality assessment. *Engineering Applications of Artificial Intelligence*, 152, 110687.
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. D. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57, 57–99. <https://doi.org/10.1007/s10462-024-10721-6>
- Zhao, Y., Wang, X., Lin, Z., & Huang, Z. (2022). Multi-Classifer Fusion for Open-Set Specific Emitter Identification. *Remote Sensing*, 14, 2226. <https://doi.org/10.3390/rs14092226>
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018, July). UNet++: A Nested U-Net Architecture for Medical Image Segmentation.
- Zhytnik, L., Simm, K., Salumets, A., Peters, M., Martson, A., & Maasalu, K. (2020). Reproductive options for families at risk of Osteogenesis Imperfecta: A review. *Orphanet Journal of Rare Diseases*, 15. <https://doi.org/10.1186/s13023-020-01404-w>

## Appendix #1: Research Facilitation Letter

*Arab American University*

*Faculty of Graduate Studies*



الجامعة العربية الأمريكية

كلية الدراسات العليا

---

June 25<sup>th</sup>, 2025

**Subject: Research Facilitation Letter**

To: Dear Mr. Hồ Văn Thép,,

VNU University of Engineering and Technology

The Faculty of Graduate Studies at the Arab American University of Palestine certifies that **Ms. Roa Omar Mohammad Baddad**, Student ID Number (202216543) is a student in the Master Program of Data Science and Business Analytics. Ms. Baddad is currently working on her Master Thesis entitled:

“Utilizing Deep Learning for Measuring Nuchal Translucency Thickness in First Trimester Ultrasound to Detect Chromosomal Abnormalities”, under the supervision of Dr. Amani Owda

This is for your kind actions to help her retrieve the required information regarding her study, noting that the information will be only used for the research and will be dealt with utmost confidentiality.

**This certificate was given upon her request.**

**Dean of Graduate Studies**

**Dr. Nouar Qutob**



---

Page 1 of 1

Jenin Tel: +970-4-2418888 Ext.:1471,1472 Fax: +970-4-2510810 P.O. Box:240  
Ramallah Tel: +970-2-2941999 Fax: +970-2-2941979 Abu Qash - Near Alrehan  
E-mail: [FGS@aaup.edu](mailto:FGS@aaup.edu) ; [PGS@aaup.edu](mailto:PGS@aaup.edu) Website: [www.aaup.edu](http://www.aaup.edu)

استخدام التعلم الآلي لقياس سمك شفافية في موجات فوق صوتية في الثلث الأول من الحمل للكشف عن التشوهات الكروموسومية.

اسم الطالبة: رؤى عمر محمد بدّاد

لجنة الماجستير:

د. أماني يوسف عودة

د. نشوى البنداري

د. سامي الصدر

## الملخص

تُعدّ قياسات الشفافية القفوية (Nuchal Translucency – NT) في الثلث الأول من الحمل أداة أساسية في الكشف المبكر عن الاضطرابات الكروموسومية، إلا أن الاعتماد على القياس اليدوي يُعرض النتائج لتفاوت كبير ناتج عن اختلاف خبرات الفاحصين وجودة صور الأمواج فوق الصوتية. يستجيب هذا البحث لهذه التحديات عبر تطوير إطار متكامل يعتمد على تقنيات التعلم العميق بهدف أتمتة قياس سمك الشفافية القفوية بدقة وثبات أعلى، وتعزيز موثوقية الفحص في مراحل الحمل المبكرة.

استخدمت الدراسة مجموعتين من البيانات تحويان 2,425 صورة ألتراساوند بين الأسابيع 11-13+6 يوم من الحمل، جُمعت من مستشفى شينزين الشعبي في الصين والمستشفى الوطني للولادة وأمراض النساء في فيتنام. يتكون النموذج المقترح من مرحلتين رئيسيتين: تتولى المرحلة الأولى تقييم جودة الصورة باستخدام نموذج DenseNet121 لتصنيف الصور إلى "قياسية" و "غير قياسية"، محققًا دقة بلغت 94% لضمان اعتماد الصور المناسبة فقط في عملية القياس.

وفي المرحلة الثانية، طُوّر نموذج تجزئة يعتمد على بنية DenseNet121 لاستخراج منطقة الـ NT بدقة عالية، متفوقًا على نموذج U-Net التقليدي، إذ حقق معامل Dice مقداره 0.897 ودقة كلية بلغت 0.989. أظهرت نتائج القياس توافقًا عاليًا مع القياسات اليدوية من قبل الخبراء،

حيث كانت 73.05% من القياسات الآلية تختلف بأقل من 1 مم عن القيم الحقيقية. كما أثبت الإطار المقترح قدرة مميزة في التنبؤ بالحالات عالية الخطورة ( $NT > 3$ ) مم، محققًا حساسية ونوعية تتجاوز 90%.

تؤكد هذه النتائج إمكانية دمج أنظمة الذكاء الاصطناعي في الفحص الروتيني للحوامل بهدف تقليل التباين بين الفاحصين وتعزيز جودة التشخيص المبكر. وتوصي الدراسة بالاستثمار في قواعد بيانات أكبر وأكثر تنوعًا، ودعم التعاون بين المختصين في مجالات الذكاء الاصطناعي والطب لضمان تبني هذه النماذج على نطاق سريري واسع.

الكلمات المفتاحية: الشفافية القفوية، الأمواج فوق الصوتية، التعلم العميق، تجزئة الصور الطبية، التشوهات الكروموسومية.