Arab American University

Faculty of Graduate Studies

Irregularities Detection in Non-Verbal Cues Using Machine Learning.

By

Noura Jamal Said Abulail

Supervisor

Dr. Amani Yousef Owda

Co-Supervisor

Dr. Majdi Owda

This thesis was submitted in partial fulfillment of the requirements for the Master’s degree in Data Science and Business Analytics

January/ 2024

Thesis Approval

Irregularities Detection in Non-Verbal Cues Using Machine Learning.

By

Noura Jamal Said Abulail

This thesis was defended successfully on 29/1/2024 and approved by:

Committee members	Signature
1. Dr. Amani Owda: Supervisor	Amani Owda
2. Dr. Majdi Owda: Co- Supervisor	Majdi Owda
3. Dr. Amjad Rattroot: Internal Examiner	……………………Dr Amjad RATTROUT.
4. Dr. Radi Jarrar: External Examiner	

Declaration

I declare that the thesis titled " Irregularities Detection in Non-Verbal Cues Using Machine Learning " is my work, it does not contain work from other researchers and has not been submitted for any other degree or qualification. This thesis is conducted for the Master's Degree in Data Science and Business Analytics at the Arab American University Palestine.

The Name of The Student: Noura Jamal Said Abulail

ID: 202012340

Signature: Noura Jamal Abulail

Date: 21/01/2025

Acknowledgments

I would like to acknowledge my deep gratitude and my warmest thanks to my supervisors Dr. Majdi Owda and Dr. Amani Owda who made this work possible. Their guidance, useful suggestions, and huge contributions have carried me through all the stages of developing and writing my project.

Abstract

Effective human communication across varied fields such as healthcare, politics, law, business, education, and social interactions It depends on understanding and expressing situations). Detecting facial gestures makes it easy to recognize the communicator’s emotional motivational (happiness, anger, fear, sadness disgust, surprise, and contempt), and health state (neurological weakness, and disorders, and strokes).

Since the diagnosis clinically of facial tics disorders involves various complex processes, patient behavior observations and evaluation usually require time and effective cooperation between the doctors and the patients.

This study proposed an effective and novel framework for detecting the irregularities in (head position, Eyelid movement detection, iris position, yawning drowsiness, mouth deviation(mouth droopy corners)) and applied as real-time assisting system for on real-time front face laptop camera, and uploaded videos and uploaded images.

first, Mediapipe face landmark model is initiated. Preprocessed frames or images using Open CV library, retrieving and extracting and landmarks from the Mediapipe models to identify the specific points or landmarks on a face, find distances between the chosen specific points, detect irregularities depending on the distance between the chosen points, classify the movement among the distance or angles and the predefined thresholds and the number of frames the facial movement lasts.

The system was evaluated on a diverse dataset of labeled images. Following preprocessing and comparison with defined thresholds, evaluation metrics (accuracy, precision, recall, and F1) were calculated. Results indicated high accuracy: 100% for head position, 96% for iris position, 86% for eyelid status, 96% for yawning detection, 88% for mouth deviation, 97% for drowsy eye detection, and 100% for mouth movement.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Expansion
ADE	Average Displacement Error
AFLW	Annotated Facial Landmarks in The Wild
AFW	Annotated Faces in The Wild
AI	Artificial Intelligence
mAP	Mean Average Precision
API	Application Programming Interface
AR	Aspect Ratio
AU	Action Units
BGR	Blue Green Red
BP	Belief Propagation
CAVE	Dataset Concerns Towards COVID Vaccines
CHT	Circular Hough Transformation
CNN	Convolutional Neural Network
COFW	Caltech Occluded Faces in The Wild
CPN	Cascaded Pyramid Network
CPU	Central Processing Unit
CQT	Comparison Question Test
DDD	Driver Drowsiness Dataset
DSDF	Digital Europe Service Data Format
EAR	Eye Aspect Ratio
ERT	Ensemble of Regression Trees
EVAL	Evaluating Large Language Dataset
FACS	Facial Action Coding System
FMD	Functional Movement Disorders
FN	False Negative
FP	False Positive
GB	Gigabyte
GHz	Gigahertz
GKT	Guilty Knowledge Test
GOTURN	Generic Object Tracking Using Regression

GUI	Graphical User Interface
HOG	Histogram of Oriented Gradients
HPE	Human Pose Estimation
HPEG	Head Pose and Eye Gaze Dataset
MPI-INF-3DHP	Monocular 3D Human Pose Estimation
IR	Infrared
KLT	Kanade-Lucas-Tomasti Tracker
KNN	K-Nearest Neighbours
LA	Los Angeles Classification
LSP	Leeds Sports Pose (Dataset)
LSTM	Long Short-Term Memory
MDN	Multi-Domain Convolutional Neural Networks
ML	Machine Learning
MOTA	Multi-Object Tracking Accuracy
MPJPE	Mean Per Joint Position Error
MRF	Markov Random Field
NIR	Near-Infrared Spectroscopy

Chapter 1: Introduction

1.1. Overview

Facial expressions, arms, and legs movements, these expressions and movements form part of interpersonal communication, and often they are unconscious processes for humans, detecting these gestures makes it easy to recognize the communicator’s emotional motivational, and mental state.

Humans often perceive their emotions through Facial expressing to facial expressions, which account for up to 30% of nonverbal expressions, and the most easily can be proceeded by visual recognition regardless of language, culture, or personal background, become versed in reading faces, learning to recognize when an emotional response is beginning, there are seven basic facial expressions: happiness, anger, fear, sadness disgust, surprise, and contempt. Facial muscle movements and head movements are responsible for bringing many of those emotional expressions, research classified these facial movements into two classes: macro expressions, and micro expressions [1].

Macro expressions are normal expressions that cover a large face area. and are easy to notice, match the tone of the verbal communication, and last between 0.5 a second to 4 seconds, but these types of expressions in many times do not express the real emotions people feel because people tend to display fake expressions, while micro expressions display unconsciously concealed emotion, it determines feelings and the state of mind, it occurs with less than a second of movement and with vibration lasting between 0.04 and 0.5 second, it conveys hidden emotions, even that person is not aware how he feel [2].

Detecting and identifying facial expressions is necessary in several fields, such as health, politics, airport security, criminal investigations, and education.

Problem Statement and Motivation

It is essential in human communication to interpret and express people’s attitudes, non-verbal behaviors are so integral to human interaction. This project aims to explore non-verbal cues and facial expressions, providing users with valuable insights into a person's emotions, health status and to indicate his or her desired action.

However, people may hide their true feelings and produce fake expressions. Identifying fake expressions is crucial for discerning deception and understanding a person's genuine intentions. False expressions manifest briefly, marked by subtle changes [3] [4].

In facial tics, the clinical diagnosis of facial tics disorders involves various complex processes; facial disorders are complex and multifaceted, requiring thorough assessments, medical histories, physical examinations, and specialized tests. They can manifest in subtle twitches or repetitive movements, and their diagnosis requires careful observation and analysis. Facial tics often coexist with other conditions, so a comprehensive and individualized approach is crucial for accurate diagnosis and effective treatment strategies. patient behavior observations and evaluation usually require time and effective cooperation between the doctors and the patients.

Recognizing irregularities and facial tics is challenging due to involuntary, short detections lasting less than half a second. They are often missed or not fully realized so needed information may be lost and so assisting systems and advanced intelligent systems advanced intelligent systems aid in recognizing these detections for images and recording videos are used to make them easier to perceive.

1.1 Objectives

This project aims to develop and present a computational framework for the automatic detection of multiple nonverbal cues and tics simultaneously.

Face expressions, tics, and head bending are analyzed to be detected and to be used in several applications.

Facial Tic Detection Algorithm Development: implement and develop a model for real-time detection of facial tics. The following objectives in this study were stated:

- MediaPipe Framework: explore and configure the MediaPipe framework into the project involving real-time perception, tracking and detection facial landmark.

- Real-Time Processing: the system should detect and analyze facial tics in real environment scenarios without significant delays.
- Non-Verbal Communication Analysis: this approach is applied to detect and classify the irregularities (yawning or blinking or looking directions.).
- User Interface Development: A user-friendly interface will be implemented for the application, allowing users to interact with the system, visualize the detected facial tics and obtain the results.
- Accuracy and Reliability Evaluation: the project will conduct thorough evaluations of the developed system to assess its reliability and accuracy, that involve testing the implemented system with a diverse dataset.

1.2 Contribution

Our project uses advanced computational techniques, integrating the MediaPipe framework, thresholds and mathematical equations to achieve detection accuracy and reliability by meticulously defining detections.

Studies in this field still confront difficult problems even after many years of research. The proposed system will detect multiple non-verbal cues in real time.

The implemented system will be used to estimate patients' status or to interpret and express their attitudes through their non-verbal cues; The motivation behind Thesis Structure

The thesis is structured as follows:

Chapter 2 literature review has presented an overview of verbal and nonverbal cues and discussed used techniques and the types of abnormal facial movements (facial tics and nonverbal cues symptoms. The relevant works related to machine learning approaches used in facial disorder movement.

In Chapter 3 the system implementation requirements were discussed. The MediaPipe Face Mesh, which creates a mesh of the approximate face geometry, and a face landmark subgraph from the face landmark module, then detects irregularities based on distance, angles, predefined thresholds, and the number of frames between selected points.

In Chapter 4 This chapter explains methodology and techniques for detecting head position, eyelid movement, iris position, yawning, drowsiness, and mouth deviation, and introduces a real-time assisting system using a front-facing laptop camera.

In Chapter 5 evaluated the implemented model and the MediaPipe model to identify potential issues, to assess its performance, and to make informed decisions about its improvement and deployment.

In Chapter 6 provide conclusions and motivations for future work.

Chapter 2: Background and Literature Review

2.1 Overview

Different cues or signals were being used to interpret and express people’s attitudes and to share information and ideas. These signals might be verbal or nonverbal [5]. Verbal communication is an all-encompassing communication involving words, involves the use of words to convey a message to the receiver .It could be oral communication, written communication, and sign language, which refers to transmitting information using a single channel (words), Verbal communication is assumed to a conscious, distinct (identifiable start-stop) communication, it has patterns and grammar rules while, primarily mode used to express our thoughts and describe how the person feels [6]. While nonverbal is a set of signs and non-linguistic sign systems, that use body language, and facial expressions, it is generally assumed as unconscious communication, and involves multiple channels (Touch (haptics)). All non-verbal cues’ expressions form part of interpersonal communication and often they are unconscious processes for humans, detecting these gestures makes it easy to recognize the communicator’s emotional motivational, and mental state [7].

2.2 Types of Nonverbal Cues

1. Aesthetic communication where people express their feelings in creative ways. This would include all the art forms: dance, music, theater, crafts, art, and sculpture. Ballet is an example of this, as there is music and dance, where there are words, there are still posture, facial expressions, gestures, and costumes [8].

2. Signs are concerned with using tools and mechanical behaviors of non-verbal communication including signal lights or flags, horns, a 21-gun salute, a display of airplanes in formation, and sirens [9].

3. Symbols of communication a person uses to build self-esteem. This includes jewelry, clothing, cars, and other things to communicate social status, financial means, or religion [10].

4. Physical communication: is being used widely for many applications that work with attitudes and attributes to get valuable information about a person’s feelings and to indicate his or her desired action [11], it can be divided into four main categories:

- Appearance and Adornment Cues: the first nonverbal cues can be noticed, it is related to the general appearance; face and body symmetry, skin color, dress, a person’s attractiveness, artifacts to make inferences about that individual’s age, gender, tastes, financial well-being, cultural background, class, values, and class [12]. Attire (clothing) can be used to convey social status, economic status, belief system (political, religious, philosophical), athletic ability and/or interests moral standards, education, and level of sophistication” [13] [14].
- Vocalists: is the study of paralanguage that contributes an extensive context for the verbal content of speaking in this type of non-verbal cue many variables can be considered, for example, regulating conversational flow and pitch helps convey meaning, and communicating the intensity of a message. Many of them can easily be recognized a higher-pitched ending as a question. People often place emphasis on greetings and lowering emphasis on farewells. Voice volume helps communicate the intensity, that louder voice is usually reflected as more intense, on the other hand, if someone flirting, sending a secret message, or with a romantic partner, it's best to lower the volume or whisper, but when used in a professional presentation lower tone wouldn’t enhance a person’s credibility. Speaking rate indicates how fast or slow a person speaks and can give others’ ideas about our emotional state, credibility, and intelligence. variations in speaking rate, as well as volume, can affect the Ability to Receive and understand other verbal messages [15].
- Kinesics cues: cues are those visible body movements and posture micromovements, that send messages about Attitude towards the other person, Emotional states, and Desire to control the environment [16]. Gestures are arm and hand movements and include adaptors to make very precise meanings known within a cultural, sub-cultural and ethnic; a hand-to-ear "please speak louder, a wave hello or goodbye, ", a "come-

here" beckoning hand movement, a hitch hiker’s thumb Head movements, and posture send different messages about our attitudes, feelings, and intentions [17]. Head movements can indicate agreement, disagreement, and interest. The posture may be self-assertion, defense, interest, motivation, or intimidation. Eye contact has been studied about eye movements, specifically making eye contact with other people's faces, heads, and eyes, and looking back during interactions. It is related to the pattern. Facial expressions can be conveyed by using the forehead, eyebrows, and facial muscles to convey the meaning [18]. Nose and mouth, Facial expressions can show happiness, sadness, fear, anger, and other emotions.

- Contact Codes: Proxemics and Haptics: Proxemics studies the effect of space and distance in communication. space, communication, and relationships are closely related. Often when a group of people is attracted to each other, we say they are "close" together. When people lose connection with someone, we can say that he or she is "distant". In general, space affects how people communicate and behave [19]. Haptics can be defined as the study of touch as a type of nonverbal communication. In our daily lives, people touch each other for a variety of purposes; they shake hands to greet and give affection if with someone, hug and kiss their friends or their amorous partner [20].

2.3 Non-Verbal Cues Detection Applications and Used Techniques

Detecting and identifying facial expressions is necessary in several fields, such as law, politics, healthcare, business, and education, and it will be discussed in the following sections:

2.3.1 Nonverbal Communication in Social and Cultural Issues

It is essential in human communication to interpret and express people’s attitudes and attributes, non-verbal behaviors are integral to human interaction, which in turn will help users get valuable information about a person’s feelings to indicate his desired action referred to as kinesics, this area of non-verbal communication encompasses posture, facial expressions, eye contact, gestures, and body orientation [21]. Individuals use these channels to convey a variety of emotions as well as to display important clues regarding their personality.

There are various sender emotional state and moods, that contribute to non-verbal encoding and is affected by the observance of an interaction, interaction length, and acquaintanceship between perceiver and sender, the same factors also affect the perceiver’s decoding process. Perceiver qualities such as personality traits and demographic attributes also have impacts on non-verbal communication and subsequent impressions [22].

Many studies worked on this area, such as [23], which summarized the nonverbal expressions of behavior and their contribution to the accuracy of personality judgments. They presented the different domains of non-verbal cues: facial expressions, paralanguage, body language, and appearance, and then presented a catalog about non-verbal cues' validity as indicators of personality judgment accuracy, as well as how they can be utilized to make personality impressions. And, authors in [24] focused on personality recognition using machine learning techniques and nonverbal cues detection based on the big five dimensions (openness, extraversion, agreeableness, conscientiousness, and neuroticism).

2.3.2 Nonverbal Communication and Health Care

Nonverbal forms of communication play a big role among people in healthcare sectors, accurate interpretation of nonverbal signals allows the understanding of patients’ communications of confusion, or disagreement, both at the interpersonal and cultural level, and lack of nonverbal sensitivity affects patients’ satisfaction and liking of the physician [25].

The facial expression posture, tone of voice, and other forms of nonverbal increases the culturally diverse world and overcome language differences and any verbal communication disagreements between patient and clinician obstacle that provides optimal medical care [26].

Regardless of cultural differences; Many patients are reluctant to disagree with their clinicians, and accurate detection of delicate nonverbal cues may be the main factor for discussions leading to shared medical decisions, this study [27] addressed this gap in medical education (H.R.) which developed a new teaching tool that has the advantage of helping clinicians based on in the neurobiology of empathy, they used the acronym E.M.P.A.T.H.Y as the key components of assessing nonverbal behaviors:.—E: eye contact; M: muscles of facial expression; P: posture; A: affect; T: tone of voice; H: hearing the whole patient; Y: your response.

2.3.3 Nonverbal Communication in Law and Politics

Distinct verbal language is being used a lot by the most popular politicians, in political elections, people pay a lot of attention to candidates' verbal communication and speech, but also nonverbal communication, such as facial expressions, physical appearance, and eye contact play a crucial role in elections and politics in general. Nonverbal communication sends information about people's behavior in their performance. Common gestures, that researchers have picked up in the politicians who succeed in accomplishing to attain the sympathy of the crowd. Many studies investigated voters' abilities to assess the credibility of politicians, they classified the cues from non-verbal behaviors such as [28] and [29].

The fact that people pay more attention to candidates' verbal communication, and nonverbal communication, such as physical appearance, facial expressions, and eye contact, plays a decisive role in elections and politics in general. Nonverbal communication, to begin, according to [30] nonverbal communication is defined be "the intentional or unintentional transmission of meaning through non-spoken physical and behavioral cues", and it has different means for transmitting information nonverbally, such as facial expressions, eye contact, gestures, and body postures. Moreover, according to [31] nonverbal communication conveys additional information about the behavior being performed, and it can be performed with other behaviors to reinforce the meanings of those behaviors or contradict them. For instance, nonverbal communication can inform others whether a person is performing a behavior earnestly with a smile or unwillingly with a grim face.

2.3.4 Nonverbal Communication Business and Education

Non-verbal communication is important for small and medium-sized business institutions, master the art of non-verbal communication, critical relationships with business partners, clients, and staff members, and during online meetings can effectively navigate using many nonverbal cues in businesses [32], today's business environment where so many employees work remotely and team members companies managements always ensure that it's staff members have the tools to communicate with each other to coordinate projects, interacting with customers to coordinate efforts between employees and negotiating with vendors [33]. A lot of information can be transferred through body language than words. Such as having control over slight facial movements like a slight crease in the lips or a raised eyebrow is a major asset in maintaining eye contact telling customers they have their full attention and confidence, and making hand gestures deliberate and meaningful. They must avoid cues that make people seem nervous or anxious such as some Unconscious movements like handwringing, fidgeting, or scratching an itch [34].

School is one of the initial places a child learns, inculcating skills to communicate effectively should be the principal aim of the school so that individuals assist society effectively. It is the superior responsibility of the stakeholders of the school to create a conducive environment to communicate.

The teacher always uses nonverbal communication, they start the class through a simple greeting with students. In the field of teaching, certainly one of the main characteristics of a good teacher is his good communication skills in the classroom. The students unconsciously receive nonverbal signals sent from the teacher to draw the students' motivation to more understanding and attention [35].

Teachers use non-verbal communication skills in the classroom and various activities can act as a catalyst for encapsulating learners of different backgrounds, abilities, cultures, communication styles, etc. under an umbrella. Nonverbal or non-linguistic communication includes facial expression, eye contact, body movement and touching, vocalization, object communication, picture communication, and animal communication.

2.4 Non-Verbal Cue Detection vs Physiological Responses

Three known principles have been used to detect deception: 1) the analysis of nonverbal cues, including movement, sweating, facial expressions such as smiling or disgusted glances, vocalizations, speech pronunciation rate, and metaphor 2) analyzing discourse content and 3) measuring physiological responses [36].

Many procedures have been developed to measure physiological responses (electrodermal, cardiovascular measures, Rate, and depth of respiration, and skin conductivity), such as the Control Question Test [37] which the first used in the U.S. Army Criminal Investigation Division in the 1950s [38]. This method concerned on the structure of the questions, the pre-test interview, the review and development of the questions to be asked, the gradation of questions presentations that suspects are asked relevant, and irrelevant questions; relevant questions is dealt with the crime such as “did you steal Mrs. Jamila's jewels on September 20, 2020? on the other hand, irrelevant questions are used to obtain a baseline such as were you in Ramallah last week? Control questions tests are structured to be not concerned with a guilty witness to respond and arouse more vigorously to relevant questions because their attention remains focused on his actual crime, innocents are subjected and expected to be reactive to control than to relevant questions, they show no difference in relevant or irrelevant questions.

The CQT (GKT) also is a used technique developed as a part of a polygraph examination, it is based on measuring physiological responses to a series of multiple choice questions to assess whether suspects conceal “guilty knowledge” to differentiate between guilty and innocent participants, The idea is that each question prepared to focus on a detail that only the person who involved in with the target event, The correct details were included among several incorrect details, called “distractors,” or “controls.” [39].

Many studies have used various methods to detect deception, but they have common limitations. Results are analyzed by experienced researchers Both researchers are susceptible to influences from contextual factors, and the knowledge and training of position change researchers can affect the results , overconfidence can lead to hasty decisions that can be harmful if held unreasonably. By examining factors such as mental state, intelligence, and predictability of questions, these gender differences are also an important factor affecting the accuracy of polygraph results, as most studies draw the focus is mainly on male subjects ignoring the possibility of independent variation.

In addition, certain activities such as physical exertion during a Covert Information Test (CQT) polygraph examination, use of sedation techniques, etc. can yield unobservable information Research shows sedation may make no noticeable difference between responses to relevant questions and control questions. Finally, the performance of polygraph machines is hampered by their slowness, making them unsuitable for use in large groups of people in settings such as airports. [40] [41].

On the other hand, since nonverbal cues observing addressed the previously mentioned limitations. The first published research on this area was Darwin’s [42], which investigated behavioral gestures and facial expressions with the use of photography and close observational techniques, this study put the scientific footing in emotional expression field area, however in the modern era, the systematic study of nonverbal cues in communication commenced in the midtwentieth century [43], and the early seminal works (e.g., [44] were researchers published about nonverbal communication and the study attracted enormous interest in this discipline mainly during the 1960s and 1970s). Lately, many studies were published about using behavioral observation to understand the process of communication, physiological functions are altered to a diverse group of researchers, practitioners, and students from a variety of disciplines including communication, health care, psychology, political science, law enforcement, sociology, education, business, and management.

2.5 Abnormal Facial Movements

Facial Micromovements often express emotions, but also there are abnormal facial movements that they are a sign of the psychological health disorder movements that can be associated with neurologic disorders such as brainstem tumor, multiple sclerosis, peripheral neuropathy, brainstem tumor, and Guillain-Barré syndrome [45].

Functional movement disorders (FMDs) affects facial muscles, such as the lips, eyelids, tongue, and other are difficult to recognize because of their phenomenology [46]. This chapter will discuss the movement disorders in facial landmarks FMDs that could be considered by the combination of features, when a patient exhibits symptoms e.g. fixed unilateral facial contractions, paralysis on one side of the face can be recognized in droopy in eyelids, and lower lip, and other inconsistent features happen during examinations; reduction or abolition of facial spasm with distraction, rapid onset and/or spontaneous remissions and response to suggestion or psychotherapy normal neurologic examination. Supportive features are young age, female gender, and associated medical conditions such as depression, headaches, facial pain, fibromyalgia, or bowel syndrome.

2.5.1 Facial Nerve and Muscles

Facial muscles with the facial nerve work together to make facial expressions. The facial nerve extends from its origin in the brainstem to the innervation of the facial muscles. Disorder movement happens when facial muscles can’t receive signals properly. which would happen for many reasons discussed in the next section causing use Drooling, paralysis facial palsy or weakness in facial muscles indicate a serious medical problem.

Damage to the facial muscles and facial nerve can be caused by:

- Bell's palsy: is the case when the nerves that control facial muscles have been injured or stopped working properly. Which leads inability to wrinkle eyebrows paralysis on mouth, or weakness or paralysis on one side of the face [47].
- Autoimmune disease: that is described diseases that are caused by accidental attacks of the immune system in the body. There are more than 80 autoimmune diseases. Types include Guillain-Barré syndrome or multiple sclerosis can cause facial muscle weakness palsy over time [48].
- Infection: sometimes a viral or bacterial infection such as ear infections, Lyme disease, or Ramsay-Hunt syndrome can cause problems in the muscles and inflammation of the facial nerve of the face [49].
- Stroke: stroke occurs when the blood vessel that supplies a part of the brain is reduced or interrupted, which reduces nutrients and oxygen sent to the brain tissues. Brain cells begin to die in minutes which damage Which causes paralysis or facial weakness [50].
- Head and neck cancer: facial muscle is affected because of growing tumor which caused by head and neck cancer [51].
- Head or face injuries: injuries caused by accidents damage muscles and nerves of the face age the facial nerve and facial muscles [51].

2.5.2 The Facial Muscles

All the muscles of the face are stimulated and invigorated by seventh nerve branches for example, eyelid closure is innervated by the orbicularis muscle. The orbicularis muscle is divided into three regions: The orbital orbicularis causes eyes forceful squeezing, and the perceptual and pretarsal orbicularis muscle parts lubricate the cornea are integral to involuntary eyelid closure (blinking) that the facial nerve when one of the reasons discussed in the previous section may cause a significant paralysis of the orbicularis muscle. These medical problems are diagnosed by asking the patient to forcefully close his eyes or to make a broad smile. For example, in idiopathic Bell palsy in a normal situation as can be noticed that slight drooping of the corner of the mouth, where the area of the left brow is higher the brow and eye if the eye is open more, and when the doctor asks the patient to close their eyes focally, he can notice the asymmetry of the closure that the patient eyelashes cannot be not buried on his left eye (weak or incomplete blink) and the mouth is drawn or upward.

2.6 Overactivity of the Facial Muscles

Facial movements are taken for granted. Blinks of the eyelids when they are quick, and frequent they keep the cornea healthy. Narrowing them protects an irritated eye. However, both underactivity and overactivity of the facial muscles are common problems. This section is related to overactivity of the facial muscle conditions [52].

2.6.1 Orbicularis Myokymia

Orbicularis myokymia is diagnosed with abnormal disorder movements of the upper or lower eyelid. The movements are quick and last a second or less. Patients themselves can easily notice these movements but doctors and experts need close observation to see them. Bundles of muscle fibers Orbicularis myokymia is related to fatigue, stress, excessive caffeine, or alcohol [53].

2.6.2 Facial Tics

Facial tics are uncontrollable spasms of facial muscles that last for a short period of time. The condition may be a unilateral movement of the eyelids such as rapid eye blinking, opening the mouth, and raising eyebrows and the side of the face. there are facial tics that are more complex such as clicking the tongue, flaring nostrils, clearing the throat, grunting [54].

2.6.3 Chronic Motor Tic Disorder Facial Tics Tourette Syndrome

Chronic motor tic disorder is more common than Tourette syndrome and less common than transient tic disorder, this type of tic disorder may occur during sleep, and requires qualified experts to diagnose it these symptoms may include detection of the following symptoms, excessive blinking, twitching, and grimacing, are common tics associated with chronic motor tic disorder.

2.6.4 Tourette Syndrome

People with Tourette syndrome is born with it, it is a genetic disorder. it affects the brain and nervous system, patients with Tourette syndrome have at least one vocal tic such as sounds (throat clearing, humming, or grunting. vocalizing curse words) and multiple motor tics (flapping arms, shrugging shoulders, sticking the tongue out, or exaggerated blinking of the eyes) [55].

2.7 Nonverbal Cues and Facial Tics Detection Using Machine Learning

Over the past decade, it has been noticed a rise in Machine Learning (ML) based techniques, including and impacting industries area including manufacturing, healthcare, autonomous driving, finance, and more. The general goal of ML is for the extraction of relevant knowledge, signal processing statistics (Bayesian methods, bootstrapping, Montecarlo method, etc.), to recognize patterns in data using (support vector machines, neural networks, reinforcement learning, decision trees, etc.).

Scientists are becoming more and more interested in the potential of ML for fundamental research, and physics. This interest is concurrent with the rise of ML approaches in industrial applications. This is somewhat expected given that ML and physics both share some techniques and objectives. Image recognition is one of the fields that have been improved and demonstrated the impact of ML methods, the success of ML in recent times has been marked at first by significant improvements in some existing technologies, for example in the field of image recognition [56].

The common method was to extract a feature vector (local features) from the image to perform image classification and recognition. Supervised machine learning approaches were used to classify images, this method requires a large amount of class-labeled training dataset samples, but it does not require some rules as in the case of rule-based methods that are required to be designed by researchers. So, different image recognition can be realized. In the 2000s, handcrafted features such as histogram of oriented gradients (HOG) and scale-invariant feature transform (SIFT) as image local features, were designed based on the knowledge of researchers. image local features with machine learning were combined, and practical applications and achievements of image recognition technology have advanced, as represented by face detection. Next, in the late 2010s , deep learning to perform the feature extraction process through learning has come under the spotlight. A handcrafted feature is not necessarily optimal because it extracts and expresses feature values using a designed algorithm based on the knowledge of researchers. Deep learning is an approach that can automate the feature extraction process and is effective for image recognition [57]. To a large extent, these advances constituted the first demonstrations of ML methods' impact on specialized tasks. The use of reinforcement learning methods in gameplay, Deep learning was the core theme of machine learning and convolutional neural networks are one of the most important Well-known approaches. Convolutional neural networks won Many competitions in recent years. Get excellent results with image recognition [58].

prior developing and get into used methods, it is important to define the task at hand. Whether the task involves virtual image processing, image classification, object tracking, object-specific detection, object recognition, or visual perception, broad definitions and techniques are defined which sticks clearly below, comprehensive definitions for each of these tasks and the corresponding approaches are elucidated below:

2.7.1 Image Verification

Image verification is an answer to the question of whether the object in the image is the same as the reference pattern. Features extraction is based on an element's visual rendering, the distance between the feature vector of the input image and the feature vector of the reference pattern is calculated. pixel-by-pixel visual verifications If the distance value is less than a threshold value, the images are determined as identical, otherwise, the images are determined as different. this approach is often used in identity verification that this approach whether an actual person is another person, face, fingerprint, iris recognition, or handwritten signature verification. In deep learning, designing a loss function (triplet loss function) is used to achieve identity verification by calculating the value of the distance between two images of the same person’s biometric image as small, and the value of distance with another person's biometric image as large [59].

2.7.2 Image Classification

Image classification is the task that solves the problem of finding out the class or the category to which an object in an image belongs, among predefined classes. There are many images classification machine learning and deep learning algorithms, previous studies researchers used SVM or KNN algorithms, and techniques that used an approach called bag of visual features have been used: (BoF) based on vector quantifying and expressing the local features of the whole image as a histogram [60]. However, deep learning is perfectly adapted to the task of classifying images, and it gained popularity in 2015 when it exceeded human recognition in the challenge of classifying images into 1000 different categories [61].

Deep learning highlights the importance of features in image learning for the pattern recognition system to be classified, in this study [62] they implemented a model that extracts image features in the input image and found out that the recognition rate of the system is affected by the quality of feature extraction. Through layer-by-layer feature space conversion, in-depth learning can get the most excellent expression of features.

2.7.3 Specific Object Recognition

Specific object recognition focuses on identifying and labeling objects within images, and videos, it solves the problem of finding a specific object in images. Object recognition algorithms are trained by giving attributes to objects and annotating them with proper nouns, and then the model can carry out new data. The key component is the object recognition bounding box to identify the edges of the object tagged with either a square or rectangle. They are annotated by a label of the object, whether it be a person, a car, or a dog to describe the target object [63].

2.7.4 Object Detection

While an image classification task can find whether an object is contained in the image or not, object detection is the problem of finding where in the image the object is located in addition to the image classification. Object detection is assumed as a supervised machine learning problem, and that model should be trained with labeled images dataset. images in the training dataset accompanied with a file that includes the boundaries and classes of the objects, to learn the features of the objects of interest that it contains. Many tools such as CNN, YOLO v2, and RNN create object detection annotations. fact, most object detection networks use an image classification CNN and repurpose it for object detection [64].

2.7.5 Object Tracking

Object tracking, in this task the algorithm problem of determining (estimating) the location and information of moving objects in image sequences this task includes: defining the object of interest, Visual representation, Statistical modeling, object localization, motion segmentation Three-dimensional shape from motion: also called structure from motion. similar problems as in a stereo vision, target initialization, motion estimation, target positioning, many approaches are used for Object Tracking: MDNet, GOTURN, ROLO—Recurrent YOLO, DeepSORT, SiamMask, Vitpose, etc. [65].

2.7.6 Scene Understanding (Semantic Segmentation)

Image segmentation uses simple object properties such as shape, size, or color to understand the scene structure in an image and make a prediction for a whole input, approaches that work in this task find object categories in each pixel in an image it provides the spatial location of those classes (localization/detection)[48], this type of problems cannot be saved using CNN, many approaches were invented to accomplish this task such as AlexNet, VGG-16, GoogLeNet, ResNet, all those approaches work with two network the encoder network which followed by a decoder network [66].

2.7.7 Classical Approaches to 2D Human Pose Estimation

The most classical used algorithms used in this application are the algorithms based on the mean shift technique or on over probability distribution techniques, this technique uses the mode value of the provided distribution to make useful clustered data without training, and the random forest framework is the most used algorithms that used to predict joints in the human body. classical approaches always start with selecting the target in the first frame manually or automatically in some cases as a template, the process locates the target object in each iteration, and the new target position is found by finding the highest correlation score. Different correlation metrics, e.g., standard correlation, phase correlation (PC), normalized correlation, and normalized cross correlation are usually used as a similarity measure in tracking applications. However, this technique is not preferred and doesn’t work successfully with unpredicted object motion and against illumination changes. Moreover, color histogram (HOG) used in those approaches may cause spatial information lost due to precise spatial relationships between pixels and normalization, and also the used coefficients aren’t considered a strong discriminative measure [67].

Object tracking detection approaches used in facial and body detection, first scientists used statistical-based algorithms and regression trees to predict body and facial deformation coefficient, such as belief propagation (BP) algorithms, and Montecarlo methods, but statistical approaches were unsuitable for unrestricted environments such as unusual illumination, different positions, large poses, occlusion variations, impractical for the very complicated position to be developed, those approaches shows poor detection quality [68], so neural network was developed. More input parameters can achieve better expressiveness and accuracy.

2.8 Nonverbal Cues Detection Using Machine Learning

Early research about facial expression recognition established quantitative classification. In 1972-1978 Ekman inspired and developed Facial Action Coding System(FACS) to measure the facial movements to taxonomize emotions. In the “Facial Blueprint Photographs," the author and other researchers studied the shown emotions of the same person under restricted photographic conditions [43].

Facial landmark detection is a computer vision task, used to detect the key points inside images for the human face processing pipeline, this process is being used for many applications such as emotion recognition, drowsiness recognition, and deception detection, each image is represented in a form of W*H*C, where W= width, H= height, and C= color, considering that color images are used 3 channels (red, green, and blue). There are common datasets that are used in facial landmark studies for detection e.g. (“300W”, “AFLW”, “COFW”, “WFLW”, in the wild”) these datasets include photos were taken in both controlled environments and unrestricted environments; 2D or 3D face landmarks, with different photo shooting conditions such as large pose, the presence of face occlusion, make-up, etc.; real images or synthetic (when faces are generated with an algorithm).

Facial landmark tracking, Facial landmark detection algorithms are generally the simplest way is tracking by detection, where facial landmark detection is applied. There are three types of works that perform facial landmark tracking: joint tracking methods, probabilistic graphical model-based methods, and tracker-based independent tracking methods.

The tracker-based independent tracking methods are based on performing the individual points in each frame on facial landmarks using object trackers, such as Kanade-Lucas-Tomasti tracker (KLT) and the Kalman Filter, the face shape model is applied to restrict the independently tracked points so that the face mesh in each frame is obtained, many applications require accurate in-the-wild facial landmark detection, the most used tool in these applications is Dlib [28], which is machine learning open-source library, gradient boosting based facial landmark detection algorithms, and Ensemble of Regression Trees (ERT), but the limitations of this method is not preferred for faces with large pose [69].

Methods based on probabilistic graphical models use visual representations to identify distributions with dependent variables, by creating dynamic systems that track facial landmark points by considering two main factors : their temporal (time-related), and spatial (location-related) relationships. The two most commonly used models for this purpose are the Markov Random Field (MRF) model, seen in [68], and the Dynamic Bayesian Network. [70].

The dynamic probabilistic models capture both shape dependencies among landmark points and the temporal coherence, this type is not preferred for the complex process because the probability of the final stage cannot be differed so the outcome results from the simulation of the process cannot be predicted [71].

In [72] authors introduced a computational methodology based on computer vision techniques that automatically detect and classify visual content of human communication from diverse facial expressions and body gestures visual data. They tested their created approach to presidential debates between Donald Trump and Hillary Clinton videos. They used Open Face to detect faces and classified facial action units (AUs). they retrieved (binary) prediction and its intensity (numeric). After the feature extraction step, obtained features from the frame of the videos a recurrent neural network (RNN) classification module with long short-term memory (LSTM) units, they divided the dataset into an (80%) training set and (20%) as validation a performed 10-fold cross-validation for reliable accuracy, and achieved the accuracies average 0.825, they also used CNN approach and trained it in Expression in- the-Wild dataset and then applied the model on the frames extracted from videos and achieved approximately the same accuracy.

Facial landmark detection also used in Drowsiness detection, many studies focused on this area of study due to associated health and safety risks for individuals, especially in activities that require constant attention such as driving, in general, they focused on yawning and excessive eye blinking detection and other drowsy behavior indicators.

In [73] authors focused on driver drowsiness detection, to detect whether the driver shows symptoms of drowsiness to avoid road traffic accidents, they proposed a system that alerts drivers, alert drivers of their drowsy state, to extract numeric features from images they developed two alternative solutions, they used recurrent and convolutional neural network(R-NN), and deep learning techniques. they obtained 93% in true negative in the resulting confusion matrix, but the accuracy obtained does not achieve very satisfactory rates; they achieved 65% over the training dataset and 60% for the test data, while in [74] authors achieved high accuracy 96.42% they used SoftMax layer in CNN classifier based on Eye state while driving the vehicle. those two studies showed a higher accuracy using deep learning than [75] in this study authors used a real drowsy dataset and proposed a technique that was based on hybrid features (eye state and body motion), they used histogram of oriented gradients (HoG) descriptors to describe the eye region from each frame using facial landmarks and used frame difference For body motion description, after that they used principal component analysis for dimensionality reduction, then they used SVM algorithm to detect drowsiness, and achieved 90%.

Many studies only focused only on the eyes as an interesting area in the face specially in such as human-computer interaction, diagnosis of diseases, assistive devices for motor-disabled persons, or biometrics such as [76] used Nonverbal communication for social robots, authors investigated the power of non-verbal communication, focused their study on Human-Robot Interaction for robots that cooperate with humans, used machine learning techniques, that work on human-understandable natural language symbols as an input data, the system they created needs support of the human teacher to identifies the need to acquire a new skill, then adds a symbolic interpretation to this learned skill, finally provides an online evaluation of the robot performance which consists at least of a good or no good feedback. Therefore, machine learning in multimodal interaction of nonverbal communication aims to improve the machine’s comprehensive learning ability of both rational intelligence and emotional intelligence. In [77] researchers used Nonverbal communication for social robots to take advantage of communicative behavior that can be used to augment and support real-time intelligence and improve the fluency of human-robot, using natural language processing, that learns from plan motions, and demonstrations.

In [78] authors proposed a real-time framework to estimate the eye accessing cues and to classify the eye gaze direction. they used Viola-Jones algorithm to detect faces found in Still Eye Chimera dataset. They used geometric relations and facial landmarks to obtain A rough eye region to be then used to obtain the direction of the eye gaze using a convolutional neural network. The proposed algorithm was tested on the Eye Chimera database and found to outperform state of the art methods. The computational complexity of the algorithm is much less in the testing phase. The algorithm achieved 86.81 % accuracy for 7 classes and for 3 classes.

In this study [79] scientists proposed a model that estimates the eye gaze direction from detected eyes using Dlib-ml to be used in a real-time eye gaze controlled robotic car application they used Convolutional Neural Network (CNN) architecture that has two identical streams for right and left eye image they used Eye-Chimera dataset and to train the model. They reached accuracies up to 90.21% and 99.19% for the datasets respectively. This paper also introduces a new dataset EGDC and the proposed algorithm finds 86.93% accuracy.

In [80] researchers used NIR camera, NIR camera is used for better accuracy in capturing image views, and used Dlib facial feature tracker to obtain the left and right eye they used CNN to extract features values that are normalized, and distances were calculated some studies used (SURF) descriptor to extract features of iris images that work efficiently on real-time applications and large datasets. [65] authors worked on; UBIRIS, MMU, and UPOL databases, each image created a vector with 64 dimensions and used LA classifier in MATLAB for separating decision boundaries reached a higher rate of recognition than previous studies that worked on the same areas and datasets such as [81] and [82]. A recent study proposed “Ize-Net” they trained on the size of the image 128 × 128 × 3 of the entire that they collected from YouTube videos, to classify the position of the pupils in the eye socket (left, right, or center) they used calculations by utilizing the relative position and used Circular Hough Transformation (CHT), and OTSU thresholding to localize the pupil-center in images, and the algorithm they proposed reached accuracy to 91.5% but when they used the same algorithm on CAVE dataset the accuracy was 82.8% [83].

2.8.1 Facial Disorder Movements Detection

There are limited studies that focus on facial tics detection, but there are some studies that implemented their image classification approaches based on visual feature learning such as [84] where in this article researchers attempted an automatic method to assist in diagnosis and evaluation by proposing unsupervised and supervised learning deep learning architecture to detect tic movement. Based on real clinical data, they trained the model using leave-one-subject-out cross-validation for multiclass classification tasks. And achieved an average recognition precision of 86% and recalls of 78%, respectively. And [85] where authors developed different approaches to improve PD detection and approached to 78.4%, the dataset they created records for seventeen Tourette syndrome TS patients’ electromyogram and acceleration data from, and then calculated captured the dominant tic. used as features in a (SVM) model to detect and classify movements

2.8.2 Machine Learning in Body Movements Tracking

Human Pose Estimation (HPE) is a task in computer vision to identify and classify the joints in the human body. Most of the HPE research scientists detected the body parts were based on recording an RGB (red, green, and blue) image with the optical sensor.

Detecting points of interest in the human body such as facial landmarks, limbs, and joints as a key point to describe the pose of a person to produce a 2D or 3D representation of a human body model. The output of the model is a skeleton-like representation of a human body to be processed for task-specific applications.

The models (skeleton-based model, contour-based, and volume-based) used are basically based on creating a map of body joints or face landmarks to be tracked during the movement. This technology tells the movement performed correctly with specific calculations of the angle of flexion in a specific joint, besides the classification of people's, actions in their life such as running, smoking, eating, or other actions. The most famous and flexible used model is a skeleton-based model. It consists of a set of joints like shoulders, knees, ankles, elbows, limbs, and wrists.

The approaches used in modeling human pose estimation used to understand the geometric and the motion of the human body can be divided into classical approaches or deep learning approaches. This section will explore both of these approach types and discuss how deep learning overcomes the limitations of the classical approaches.

Recently Human pose estimation has received a lot of attention, for its various applications. so many studies were focused on state-of-the-art human pose estimation, this task algorithm should locate the human parts, such as the head, the neck, the left/ right elbows, and the right/left shoulders such as [84], however human poses are variant, which makes it difficult to create a dataset for this tracking task, some datasets were created such as Buffy, FLIC, Parse but those datasets were not suitable for deep learning algorithms because of their limited number such as AI Challenger, PoseTrack, MSCOCO, MPII, LSP.

In [86] the authors proposed a ConvNet model to predict 2D human body poses in images. They used a heatmap representation model regression for each body key point and can learn and represent the appearances and the configuration context of the parts. they made the following three contributions: intermediate feature representations, for performance improvement, in the second stage they trained the model end-to-end and from scratch, with auxiliary losses; and in the final stage, they investigated whether key point visibility can also be predicted. The model is evaluated on two benchmark datasets. The result is a simple architecture that achieves performance on par with the state of the art, but without the complexity of a graphical model stage (or layers). where in this work [87] they did not use convolutional network as in previous studies in the area, Pose Former was presented for 3D human pose estimation in videos., they modeled the human joints relations within frames that are extracted from datasets Human3.6M and MPI-INF-3DHP using a spatial-temporal transformer structure of 9 frames as an input of Cascaded Pyramid Network (CPN-detected) 2D poses (J = 17) and then to take the output to predict the 3D pose. They fixed all the parameters to compare the impact of each module fairly; the number of spatial transformer encoder layers is 4 and the spatial transformer embedding dimension is 17 × 32 = 544, they used. MPJPE (Mean Per Joint Position Error) for evaluating the average score and results showed that MPJPE is improved by approximately 2% compared with SRNet in [88].

In [89] conducted an extensive experimental study, this paper addressed and introduced the PoseTrack benchmark for video-based human pose estimation and articulated tracking. They implemented modules that accomplish many tasks focusing on estimating a multi-person pose from a single frame, multi-person pose estimation extracted from videos, and tracking multi-person from videos. The dataset they used is collected, annotated, and labeled. Furthermore, First, they relied on a person detector in the image person in the bounding box. They reprocessed images before it is entered into the models; they are cropped and rescaled. As a second simplification, they applied the RCNN model from the TensorFlow Object Detection on the level of full body poses. They evaluated the model using MOTA evaluation metric and achieved a good MOTA score this, in general, degrades pose estimation performance. [90] Used in assistive service robot applications, even though current solutions provide high accuracy results in controlled environments, they tried to fill the gap in tracking initialization and failure, large object handling and partial-view body part tracking, and body part intersection they used `Make Human’ open source tool, their implemented framework based on 3DSDF data representation model, and finally they evaluated the model in three datasets SMMC-10, EVAL and PDT. The framework makes 3D projecting 9 skeleton joints: head, elbows, hands, knees, ankles. And performed annotation every 10 frames. achieved promising results: ADE 0.075, mAP 0.825.

2.8.3 Rule-Based Modeling integrated with Mediapipe

Recently, MediaPipe has been used to detect hand movements in many applications and experiments. For example [91], researchers developed real-time, accurate visual recognition using MediaPipe to facilitate sign language communication for the deaf, among others. There were three steps in this process:

First: hand searching in the current frame using an encoder-decoder to extract important visual information by training the finger to estimate the bounding box around the area of interest. Then, extracting x and y coordinates using a Hand Landmark model, cleaning points to remove bias, removing null entries, and normalizing the x and y point, and finally they employed supported vector machine (SVM) to accurately classify data points in a data set divided into training and validation sets (80:20%). A RBF (radial basis function) kernel was used to segment several sign language letters and numbers. Finally, the performance of the model was evaluated using performance measures and an accuracy of 99% was obtained on most sign language data. And in the established study [92] for detecting one of the common Parkinson’s diseases (PD), researchers aimed to quantitative asses hand movement using MediaPipe, on two types of data frame-rate videos and accelerometer data that were recorded for 11 patients, they investigated the frequency and amplitude relationship between accelerometer data and video recordings and achieved an automatic estimation of the movement frequency.

Also, the experimental study [93] used MediaPipe’s opensource framework to demonstrate Real-time accurate detection with a methodology that simplified Sign Language Recognition to help communication problems for the deaf-mute community, they used Multiple sign language datasets to analyze the capability of the framework the created model is efficient with an average accuracy of 99%.

The majority of the time, multiple streams of incoming image or video data are analyzed using neural networks like TensorFlow, PyTorch, CNTK, or MXNet. When handling data with such models, one input results in one output, allowing for highly effective processing execution.

On the other hand, MediaPipe operates at a much higher semantic level and supports more complex and dynamic behavior. For instance, neural networks are unable to describe situations when a single input might produce zero, one, or numerous outputs. AI perception and video processing require streaming processing as opposed to batching methods. Due to its intrinsic capability for streaming time-series data and operations on any data format, MediaPipe is substantially more suited for processing audio and sensor data.

Humans often perceive their emotions through Facial expressions, Facial expressions, account for up to 30% of nonverbal expressions, and the most easily can be proceeded by visual recognition Regardless of language, culture, or personal background, Become versed in reading faces, learning to recognize when an emotional response is beginning.

Table 2.1: Differences Between Mediapipe, YOLO, Blazepose, Openpose, Tensorflow, And Dlib Based on Various Based on Various Aspect

Aspect	MediaPipe (Pose)	YOLO	BlazePose	OpenPose	TensorFlow	Dlib
Tasks	Pose, Face, Hand Tracking	Object Detection	Pose Estimation	Multi-Person Pose	General ML, CV tasks	Face Recognition, CV
Key Points - Head	468	-	190	70	-	68 (Facial Landmarks)
Key Points - Body	33	-	-	25	-	-
Key Points - Hands	21	-	21	-	-	-

2d/3d	Both	Mostly 2D	Both	Both	Depends	2D (Landmarks)
Multi-Person	Yes	Yes	Yes	Yes	Depends	Yes
Framework	Google	Custom	Google	Custom	Google	C++
Applications	Various	Object Detection	Fitness, Animation	Sports, Dance, Medical	Various	Face Recognition, CV
Speed	Fast	Fast	Fast	Moderate	Variable	Moderate
Community	Active	Active	Active	Active	Active	Active
Input Resolution	Flexible	Fixed	Flexible	Flexible	Flexible	Flexible
Detecting Far-Away	Yes	Yes	Yes	No	Yes	Yes
GPU Power Required	Low to Moderate	High	Low	Moderate	High	Moderate to High

2.8.4 Rule-Based Modeling integrated with Mediapipe

Mediapipe and role based works together. Developers can integrate these models recognize specific facial expressions, body poses, or both of them , enabling the system to make informed predictions about the roles individuals play within a given context [94].

In a role-based model, key is assigned specific roles or tasks based on their observed actions or positions. Thresholds and roles translate and specify functions and status performed by participants in an image or video stream. MediaPipe's use of facial and body landmark information makes it easy to build models that can detect activities in a landscape. MediaPipe, a powerful computer vision and machine learning platform, enables usage-based modeling of understanding human interaction and behavior through faces and bodies.

2.9 Chapter Summary

In this chapter, an overview of facial landmarks movements, nonverbal cue types, recognizing applications, overactivity of the facial muscle types, and their symptoms were discussed, and then the basic tasks of algorithms with images were mentioned and clarified, Moreover, several previous approaches to for movements detections were reviewed. In addition, irregularities and expressions detection were stated. the classical approaches were compared with machine learning approaches, and they discussed machine learning approaches their advantages and disadvantages in the applications, and how each approach overcame the obstacles and limitations for each. As a result, the gathered literature review in the previous chapter inspired the researchers to invent, design, and propose a novel application that detects irregularities to develop and present a computational framework for the automatic detection of multiple nonverbal cues simultaneously. Face expressions, tics, and head bending those features are analyzed and to be detected and to be used in several applications (clinical, psychiatry, educational, and organizational psychology some of which include military psychiatric evaluation, social skills training), Table (2.1) shows that MediaPipe model is the best choice to do this mission.

MediaPipe's user-based modeling capabilities enable developers to create advanced applications that sense, interpret and respond to human behaviors, thereby providing computing and information relevant improvements Role-based systems are faster, easier to implement, and machine learning increases productivity, effectivity and efficiency. The proposed framework is presented in the next chapter.

Chapter 3: Technical Requirements and Used methods

Introduction to Methods

This chapter outlines the method for detecting facial movement irregularities macro movements and micro movements in real time. This chapter explains the technical requirements, dataset used, equations and mathematical models, which aims to detect irregularities in face movements based on ML algorithms in real time. It begins with describing the requirements to implement the system. Then the MediaPipe model and its configuration parameters will be explained which is the main component of the proposed model.

MediaPipe Face mesh model will be used as the basis of the face landmarks mesh tracker, to improve the key point accuracy, and to detect irregularities in face landmarks. Focused on important regions of the face (head, eyelids, iris, and mouth), MediaPipe is used to localize the face features, then results will be defined by the decision maker as will be discussed in the next chapter.

3.1 Technical Requirements

To design the system, the following requirements are needed as a guideline during the implementation of the facial disorder movements detection system, the target of the methodology the implemented system should be able to classify facial movements as accurately as possible and the number of misclassifications has to be minimal, and it must be able to handle a variety of data inputs (Realtime video, recorded videos, and images). Therefore, the robustness of the method must be tested for the properties.

3.1.1 Hardware Requirements:

- CPU/Memory usage: desktop computer with an Intel Core i7-3770K CPU @ 3.50 GHz, 16 GB memory, and a NVIDIA GeForce GTX 1070, When the disorder movements detection system is used, it shouldn't take up too many resources of the computer it's running on. In this way, the system can be embedded in an application and run on any modern computer.

- Single RGB-Depth camera: the pre-processing and classification process must be real-time in order to have a practical application that feeds a webcam's stream into the system. Even though this project does not involve creating a software application that utilizes a webcam stream, it's important to remember this requirement for future research.

3.1.2 Software Requirements:

The system will be developed using Python, leveraging libraries used to read images, extract features, preprocessing image processing and implementing machine learning algorithms, the most important libraries installed:

- Anaconda -Python is used, and installed important libraries such as NumPy and, MediaPipe, pyttsx3, SciPy, Spatial libraries.
- OpenCV: is an open-source cross-platform library that is used to develop real-time computer vision applications. It is usually used to process images, and videos, to analyze captured features like face detection and object detection.
- MediaPipe: is a Framework for building machine learning pipelines for processing time-series data like video, this is used for facemesh MediaPipe function; face mesh consists of two deep neural network models that work together that work in real-time: first works as a detector for the face locations from the full image and computes face locations and the second model that operates the 3D face landmark on those coordinates locations and make a regression to predict the approximate 3D surface. Where the face is cropped drastically which reduces the need for common data processing and augmentations like translation, scale changes, and rotations. In video when the model could not identify face presence in a frame, it crops the face landmarks based on landmarks identified in the previous frame.
- NumPy: its name stands for Numerical Python, it is an open-source library that is highly optimized library for numerical operations, media pipe uses NumPy to make

the image frame; pixels of image frames because pixel data of images will be Reorganized to be contiguous in python.

- Pyttsx3: a library that converts the entered text into speech.
- SciPy. Spatial: this library is used to calculate Euclidean distance between detected points.

3.2 Detection Algorithm

The system leveraging the MediaPipe model for facial gestures detection", the MediaPipe face Detector task uses a machine learning (ML) model to locate faces and facial nodes within a single image or a continuous stream of image frames. MediaPipe framework enables developers to implement multi-modal (video, audio, times series data). It also provides a collection of human body tracking and detection models which are trained on a massive dataset and various data of Google. MediaPipe framework is written in Java, C++, and Obj-C with the components (Calculator API that written in (C++), Graph Construction API which is written using (Protobuf) and Graph Execution API that written using (C++, Java, and Obj-C) as shown in the Figure (3.1) [98].

The model detects key points on different parts of the body as nodes, edges, or landmarks, the graph is fed as a collection of nodes joined with directed connections, and all three-dimension coordinate points are normalized. MediaPipe enables developers to implement a pipeline incrementally, MediaPipe pipelines are composed of nodes that are generally specified and connected to C++ files. A vision pipeline according to is defined as a directed graph of a collection of nodes (“Calculator”) joined with directed connections (“Streams”), Each stream represents a time series of data “Packets”. Together, the calculators and streams define a data-flow graph. The time stamps of the packets that traverse the graph are used to collate them together. To enable the receiving node to utilize the packets at its own rate, each input stream maintains its own queue. calculators can be replaced, customed inserted incrementally to refine the pipeline in the graph [99].

3.2.1 MediaPipe Toolkit

MediaPipe Toolkit comprises the framework and the solutions as shown in the block diagram in Figure (3.1) that works with Face Landmarked task enables the user to detect face landmarks, once when the data is fed into the system the pipeline nodes are connected, and the task outputs an estimate of 468 3-dimensional face landmarks and 52 blend shape scores in Figure (3.2) the landmarks of the human body were detected to analyze the posture of the body landmarks.

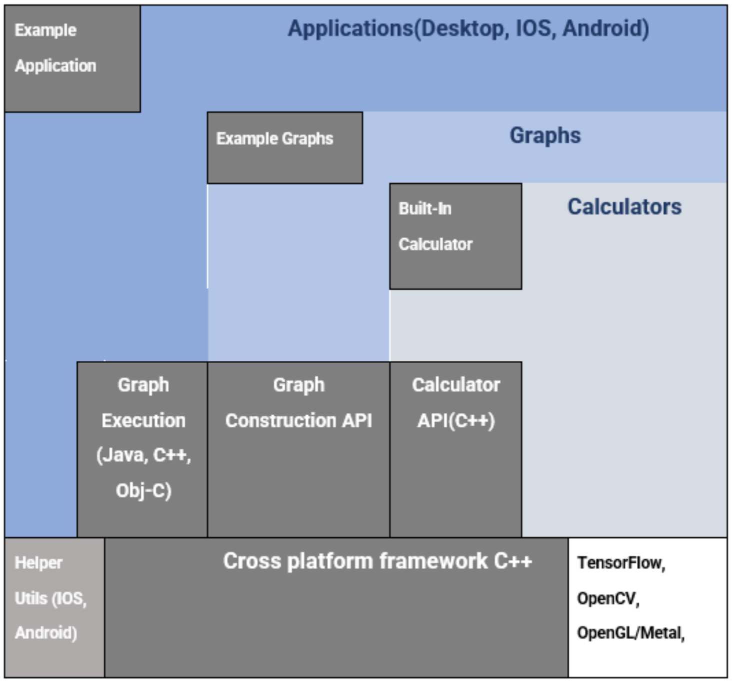

Example Application
Applications(Desktop, IOS, Android)
Example Graphs
Graphs
Built-In Calculator
Calculators
Graph Execution (Java, C++, Obj-C)
Graph Construction API
Calculator API(C++)
Helper Utils (IOS, Android)
Cross platform framework C++
TensorFlow, OpenCV, OpenGL/Metal,

Figure 3.1: Mediapipe Toolkit

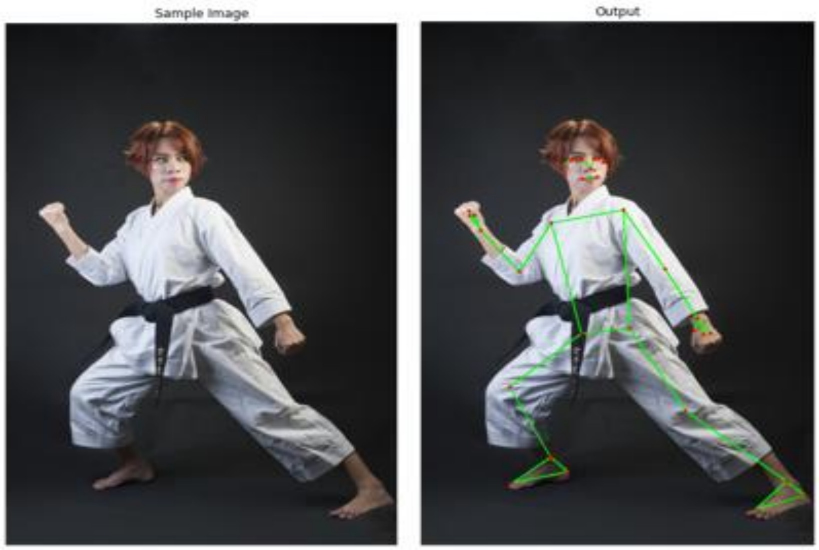


Figure 3.2: Detecting Landmarks of the Human Body using Mediapipe.

3.2.2 Model Configuration

There are two main steps to accomplish the required task that are discussed as following input image processing - includes image resizing, normalization, rotation, and color space conversion.

Score threshold – results will be filtered based on prediction scores. The model was configured as the following:

Table 3.1: Task Input and Task Out in Mediapipe

Task inputs	Task outputs
The input that Face Detector accepts: • Live video feed • Decoded video frames • Still images	The results of the Face Detector output: • Bounding boxes around detected faces in an image frame. • 3D dimensions Coordinate for 6 face landmarks for each detected face.

After Reading the input image or video frames, MediaPipe detector was built. In this stage, the input frames will be passed to the detector. Then an object called “results” to store information about landmarks will be created, facial area coordinates such as eyes, mouth, and nose, and the confidence score.

3.3 Features

Key features of a person’s face will be identified using MediaPipe facial mark recognition, providing a wealth of data for activity-based modeling The system enables developers to create emotional, gesture, or interactive indicators of certain functions through facial marks exploring the spatial relationships of species and forward

exclusion of relevant facial features identified in the literature review that might suggest facial abnormalities. This includes head positioning, monitoring rapid eye blinks, and other involuntary contraction of facial muscles, I have identified the same measures used in [100] and described them in the following points:

- Head position: this feature is chosen to be detected because it helps in emotion recognition. For instance, it can be used for safety driving monitoring to recognize whether the driver pays attention to the road or not, the developed interface can easily classify each side the person is looking at (left, right, down, or forward).
- Frame counter: Micro-expression happens in a short duration and subtle movement amplitude, so a new parameter was added as a threshold to determine if the face movements are irregularities, the application should compare each detection duration with the threshold, for example, mouth opening cannot be considered as yawning only if the duration time of the mouth opening detection exceeds the predetermined threshold it will be considered as a facial irregular movement.
- Eye tics: in facial tics described in the section (2.7), shows that the occurrence of these eye tics is common in many physiological cases or mental cases, our development focused on detecting the abnormal movements that happen in the eye region, and focused on the most occurring tics in the eye region, which are detected by eyelid and iris movement:

- Eyelid movement: detecting eyes lid movement has been of great interest observations by researchers There are a wide of applications that concern eyelid movement detection, such as human activity recognition, emotion recognition, visual behavior change, mental illness diagnosis, eyelid closure for fatigue detection, emotion recognition, and other researches focused on the blinks detection; that muscles at are responsible for the lid movements are led by electrooculogram activity, blinking is a natural defense system from exposure in the environment. many applications in the fields of human-computer interaction and medical diagnostics, But the involuntary irregular eye blink is an indicator for fatigue diagnostics, on this line, the status of eye blink is detected and the number of blinks can be counted in the developed interface.
- Iris position: Eye-gaze mapping or localizing the pupil could be used in many applications that are related to driving alert systems, marketing (identifying individuals of interest in public spaces), training, and it helps in developing assistive technology for individuals with disabilities, allowing them to control computers investigation, the methodology was described to detect and classify iris position in three positions (right, center, and left) is described in section 4.2.

- Mouth tics: lips can express a wide range of emotions, and moods or it could be an indicator of an underlying condition. The symptom has several different causes (such as (Bell’s palsy, Hemifacial spasm, Trauma, etc.), our built system focused on the following features:
  - Yawn: yawning is a common reflex where someone opens his jaw wide, which is accompanied by a long inspiration, Yawning detection has a variety of important applications in driver fatigue detection, well-being assessment of humans, driving behavior monitoring, however also yawning is associated with some released hormones that briefly increase the alertness and heart rate.
  - Lips pulling or puckering: this feature is chosen to be detected as an important feature, that people with facial nerve impairments cannot blow out a candle, this feature is used to distinguish between facial nerve impairments cases and facial

movement disorders, and also detects mouth puckering can help diagnose conditions like temporomandibular joint (TMJ) disorders.

- Mouth droopy corner: mouth droopy corner or mouth deviation can be related to both mental status such as feeling sad, angry, or tired, or a medical problem such as Bell's palsy, Facial paralysis, and strokes in the application we developed the detected deviation and in which side of the face (left or right).

3.4 Equations and Mathematical Models

For every video frame, the landmarks are detected. The aspect ratio (AR) between points of the area of interest is computed. This technique will be used for Eyes and mouth opening detection since they are partially person and head pose insensitive.

The points in the face detected using MediaPipe will be used to find the single scalar quantity named Aspect Ratio (AR) formula to obtain the level of eyes or mouth opening as illustrated in Equation (1).

$$\mathbf{AR} = \frac{\|\boldsymbol{P_2} - \boldsymbol{P_6}\| + \|\boldsymbol{P_3} - \boldsymbol{P_5}\|}{2\|\boldsymbol{P_1} - \boldsymbol{P_4}\|} \qquad \mathbf{(1)}$$

where P1, P2, P3, P4, P5, P6 are the 2D landmark locations.

For eye irregularities detection it will be needed to find the average ratio If the eyes have been closed exceeded the predefined threshold, and the alarm will set off.” The Eye average aspect ratio EAR, for both eyes EAR value is calculated as in Equation (2):

$$\mathbf{Avg.\,EAR} = \frac{\mathrm{AR}_R + \mathrm{AR}_L}{2} \qquad \mathbf{(2)}$$

3.5 Frames Counter and Wait Time

This project focuses on detecting the irregularities in facial landmarks, frames counter is the aspect that will be considered whether the movement will consider whether the movement a disorder or not , that can consider if the mouth open during talking or it is a yawning is frame counters and thresholds, the application implemented system will compare each detection lasts number of frames with the declared threshold value.

Also, time aspect feature will be used; the wait time threshold to consider the movement as a disorder when the passed time is less the predefined exceeds the permissible limit for example, mouth opening cannot be considered as yawning only if the duration time of the mouth opening detection exceeds the predetermined threshold, then movement can be considered as a facial irregular movement.

3.6 Evaluation Metrics

The model should be evaluated to ensure that the model functions effectively and reliably; when evaluating the performance of the model correct and wrong predictions should be explained.

Correct predictions include the number of correct (true positives and true negatives) predictions for each class.

- True positive (TP): for correctly predicted event instances values.
- True negative (TN): for correctly predicted negative instances values.

Model errors include false positives and false negatives.

- False Positive (FP): incorrectly predicted instances values.
- False Negative (FN): for incorrectly predicted negative instances values.

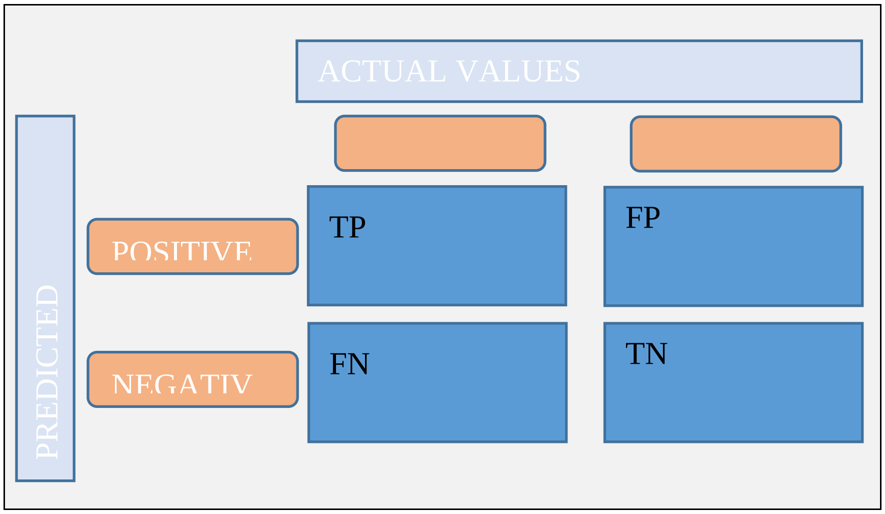


Figure 3.3: Confusion Matrix for the Binary Classification

Here's is a summary to describe these evaluation metrics

1. Accuracy: this metric measures the overall correctness of the model's predictions by finding the ratio of correct predictions total number to the number of predictions and the formula is obtained in Equation (3):

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{correct\ prediction}}{\boldsymbol{all\ predictions}} \qquad (3)$$

2. Precision: this metric measures the ability of the model to correctly predict the positive class among a total number of predicted positives (both true and false positives) the formula is obtained in Equation (4).

$$\textbf{Precision} = \frac{\textbf{True Positives}}{\boldsymbol{True\ Positives + False\ Positives})} \qquad (4)$$

3. Recall (Sensitivity or True Positive Rate: measures the ability of the model to correctly predict positive samples (true positives) among all the actual positive samples in the dataset the formula is obtained in Equation (5)

$$\textbf{Recall} = \frac{\textbf{True Positives}}{(\textbf{True Positives + False Negatives}).} \qquad (5)$$

4. The F1 score: it is a harmonic mean of precision and recall, the F1 score will be high if both recall and precision are high and is calculated using the following formula in Equation (6):

$$\mathbf{F1\ score} = \frac{\mathbf{2 * (Precision * Recall)}}{\mathbf{(Precision + Recall)}}. \qquad (6)$$

3.7 Web Graphical User Interface (GUI)

Streamlit library is used to create a user-friendly GUI, this library allows user real-time interaction with the detection system. Using open-source Streamlit Python library. Streamlit-Webrtc library is used to create the Web Graphical User Interface (GUI).

Streamlit was founded by Google engineers to overcome challenges faced while deploying and developing machine learning applications and dashboards. It supports mainstream Python libraries such as pandas, matplotlib, and plotly and It allows developers to transmit and handle real-time video/audio streams over a secure network.

The built web application will work on three types of input data phases as following data input from a single RGB camera, to detect nonverbal cues in real-time (in this project, the laptop camera is being used), so there is no need for specialized hardware, the second phase, Data input from video uploaded by the user, and uploaded Images.

3.8 Chapter Summary

This chapter highlighted the needed requirements, and an overview of the importance of facial movement detection in various applications and the contribution of the chosen feature, a an explanation in details of the used methods and techniques used to achieve the aim to detect irregularities in face movements based on ML algorithms in real time and covered the required components for creating facial disorder movements detection applications, metrics used to evaluate the performance of the implemented model, and finally talked about the used library to create a user-friendly GUI. Next chapter will discuss the implementation and the experimental work.

Chapter 4: Methodology and Implementation

4.1 Introduction

The methodology employed in this research is crucial to the successful implementation of the proposed facial analysis system. This chapter outlines the steps and processes undertaken,. The methodology is structured into distinct sections, each addressing a key aspect of the research process.

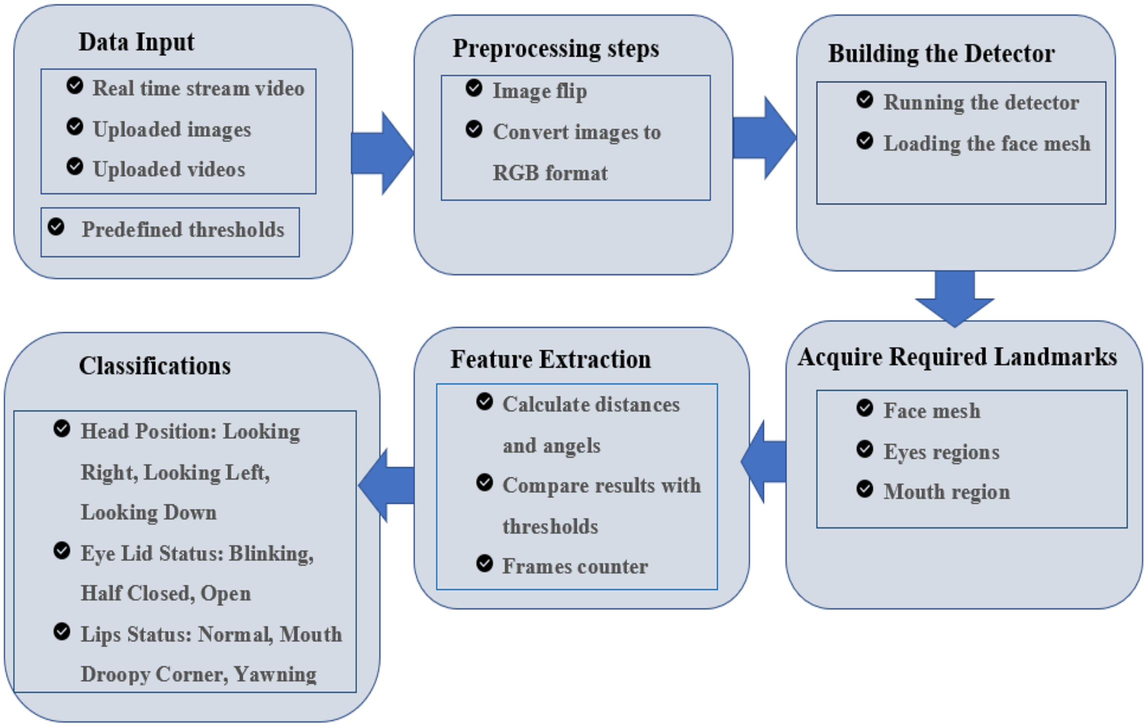

Data Input
Real time stream video
Uploaded images
Uploaded videos
Predefined thresholds
Preprocessing steps
Image flip
Convert images to RGB format
Building the Detector
Running the detector
Loading the face mesh
Acquire Required Landmarks
Face mesh
Eyes regions
Mouth region
Feature Extraction
Calculate distances and angels
Compare results with thresholds
Frames counter
Classifications
Head Position: Looking Right, Looking Left, Looking Down
Eye Lid Status: Blinking, Half Closed, Open
Lips Status: Normal, Mouth Droopy Corner, Yawning

Figure 4.1: Methodlogy Flow Used for Classifying Facelandmarks Movements Tracking Flow Chart In This Thesis

4.2 Dataset

The system will be designed to handle various types of images and they are described as the following:

- Real-time video input involves processing a continuous stream of frames from a from a webcam camera in real-time, the model will process each frame independently.
- pre-recorded video: the system will be designed to handle with user with pre-recorded video data in different types of formats (MP4, AVI,etc.), the video file is read frame by frame.
- images: the system also should work with individual images, images can be uploaded in various formats such as (JPEG, PNG) and to be analyzed.

4.3 Preprocessing Steps

Image preprocessing is an essential step in preparing data before being passed in MediaPipe model or any computer vision model, it helps in improving model performance and reducing computational demands.

4.3.1 Image Flip

Since the input image is taken with a front-facing/selfie camera, the model assumes it is mirrored with images flipped horizontally, so images need to be flipped before feeding the model for a more natural video feed.

4.3.2 Convert the image to RGB format

The format from when OpenCV library is used is BGR, but MediaPipe model but the (algorithm) should have an input image with RGB format to make it better so for the end-user.

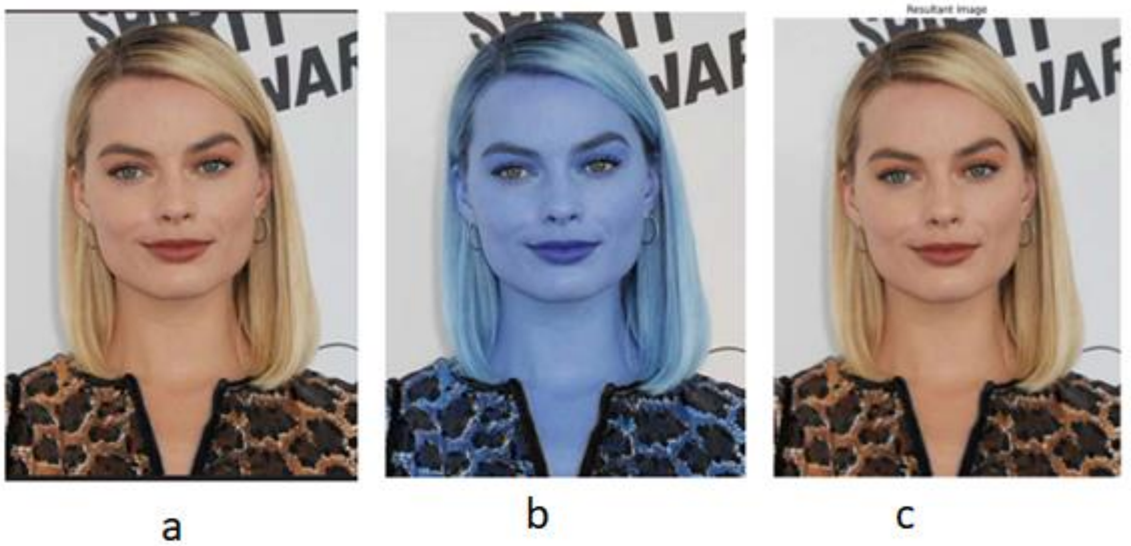


Figure 4.2: (a) Shows the Input Image With Its Original Color (b) Shows When The Image Is Read Using OpenCV Library and (c) When The Function COLOR_BGR2RGB Is Used

4.4 Building the Detector

MediaPipe solutions already have a built-in Face detection module. The detector is built with the following parameters:

- max_num_faces: that describes how many faces the system will be used to detect, in our case, the task is to detect only one face to detect.
- refine_landmarks: this function refines the detected Eyes and Lips landmarks and adds the additional landmarks for Irises of Eye so in our model this value was set to True.
- min_detection_confidence: here the function describes the minimum detection confidence of the detection it’s value usually in the range (0.0, 1) for the face detection model and set to 0.7.min_tracking_confidence: it describes the minimum

confidence for landmarks tracking and, in our implemented system, user allowed to determine this value in the GUI.

4.4.1 Running The Detector

MediaPipe contains functions to trigger inference using code that executes the processing with the task model, using the load_detector_process() method to pass the input.

4.4.2 Loading the Face Mesh model.

MediaPipe library contains two different models, the first is a face detector that operates on the full image and locates the face in the image. The second model is the face landmarks detector that predicts the 3D facial landmarks, Figure (4.3) shows 468 facial landmarks are detected in the face; each landmark has a key number.

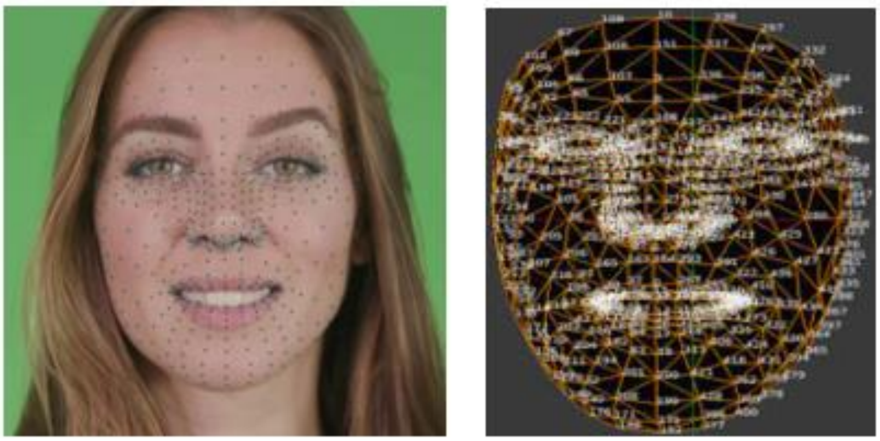


Figure 4.3: Loading 3D Facial Landmarks using MediaPipe

4.5 Acquire Required Landmarks

Relevant facial features are extracted to capture face micromovements and macromovements. Distances and angles between facial landmarks, and changes in facial contour, are computed. This step aims to extract the spatial information obtained from MediaPipe into quantitative metrics for further analysis.

Table (4.1) is describing a chosen 27 selected corresponding MediaPipe landmark IDs, and their locations on a sample face image are shown in Figure (4.4) selected key landmarks (vertices) with the corresponding MediaPipe landmarks.

Table 4.1: Selected Key Landmarks (Vertices) and the Corresponding Mediapipe Landmarks.

Key landmark ID	MediaPipe landmark	Description
0	336	Left eyebrow (inner)
1	276	Left eyebrow (outer)
2	334	Left eyebrow (middle)
3	46	Right eyebrow (outer)
4	105	Right eyebrow (middle)
5	107	Right eyebrow(inner)
6	61	Mouth end (right)
7	308	Mouth end (left)
8	13	Upper lip (middle)
9	14	Lower lip (middle)
10	50	Right cheek
11	280	Left cheek
12	48	Nose right end
13	4	Nose tip
14	289	Nose left end

15	206	Upper jaw (right)
16	426	Upper jaw (left)
17	133	Right eye (inner)
18	130	Right eye (outer)
19	159	Right upper eyelid(middle)
20	145	Right lower eyelid (middle)
21	362	Left eye (inner)
22	359	Left eye (outer)
23	386	Left upper eyelid (middle)
24	374	Left lower eyelid (middle)
25	122	Nose bridge (right)
26	351	Nose bridge (left)

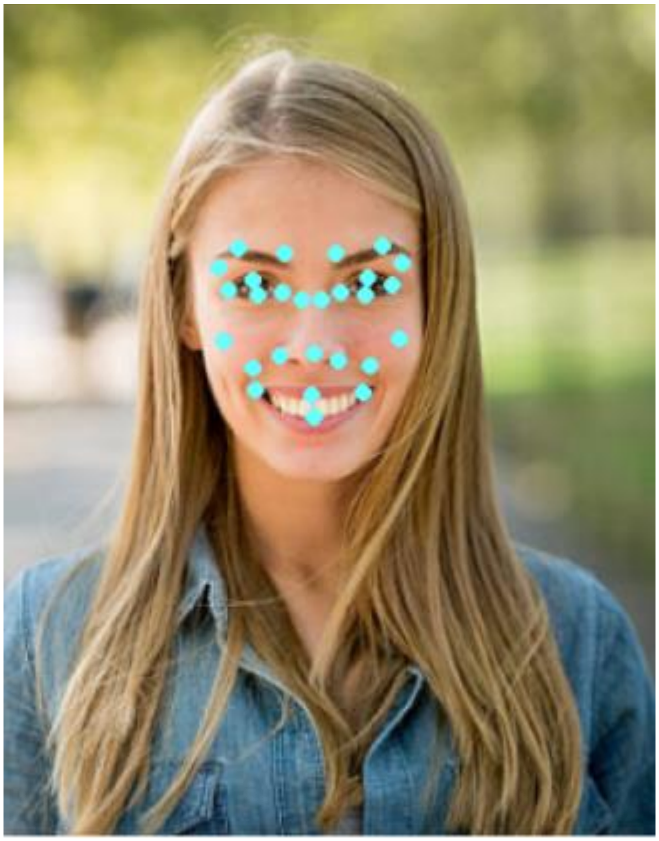


Figure 4.4: The 27 Key Landmarks and Their Locations.

4.5.1 Denormalize Landmarks

MediaPipe detected landmarks are represented as points in 3D space with x, y, and z coordinates. x and y are normalized by the image width and height respectively, while z represents the landmark depth, and the smaller z value the closer the landmark is to the camera.

However, the coordinates should be denormalized to let those scaled points are not relevant for the user and for further processing or analysis. landmark value can easily be obtained by multiplying the x and y of the landmark with the width and the height of the input image.

4.6 Features Extraction

To clarify facial abnormal movement analysis, mathematical modeling comes into play. Aspect ratio equation was used to interpret the extracted features. This stage involves the adjustment of the ranges of movement.

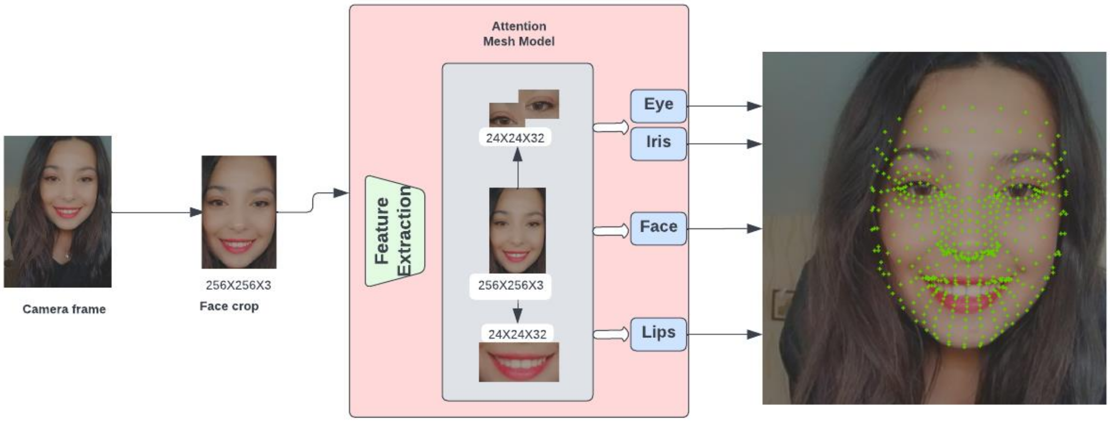


Figure 4.5: Facelandmarks Movements Tracking Flow Chart

We used thresholds are defined to determine accuracy. These thresholds are defined through empirical observations, domain knowledge, trial and error training, fine-tuning, evaluation metrics, and group-specific constraints. Empirical observations help understand the distribution of scores or confidence levels, while domain knowledge guides the selection based on the application. Trial and error allowed to experiment with different threshold values during the testing phase to find the optimal balance between precision and recall.

4.6.1 Head Position Detection

The position of the head yaw, pitch, and roll angles in a fed image was estimated by following these steps:

1. After loading the libraries MediaPipe ‘FaceMesh’ to detect the face landmarks, the next step is to initialize several objects. After processing the image, the next step is to retrieve the key point coordinates. 6 keypoints chosen to represent the face (1, 33, 61, 263, 291, and 199). Those points are on the nose, edge of the eyes, edge of the mouth, and the chin.
2. The function SOLVEPNP_ITERATIVE is a computer vision function used to estimate the position and orientation (pose) of a 3D object, such as a head, in a 2D image or video stream used as a computation method for a pose to produce a rotation

vector which uses a minimization scheme called Levenberg-Marquardt. The function needs two input matrixes, after having the 2D and 3D key points coordinate the camera matrix is gotten. the camera matrix is represented $\begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. Where: (fx and fy): focal point, γ: skew parameter, (u0, v0): the optical center c and added empty distance matrix with the shape of 4x1. This matrix only contains zero.
4. After getting the output two vectors are the translational vector and the rotational vector, so the OpenCV Rodrigues function is used to get rotation matrix.
5. The OpenCV RQDecomp3x3 function for extracting the angles that was used to retrieve the rotational angle on each axis.
6. Finally, classification is applied depending on the output extracted angles' rotation.

4.6.2 Disorders Detected in The Eye Region

Detecting eyeblinks and iris position can be applied in various fields to monitor and address a range of issues. Such as patterns can aid in diagnosing and studying neurological conditions such as Parkinson's disease, Eye-tracking technology aids people with disabilities to control computers and devices using their gaze, provide insights about students' engagement and attention during lectures, And the following sections will describe the methodology used to detect these movements.

Eyelid Movement Detection

In blink count detection, many things should be considered, first of all, to determine whether the eye is open or closed, aspect ratio technique is used and the predefined threshold that eye should be closed at all to be assumed it blinks.

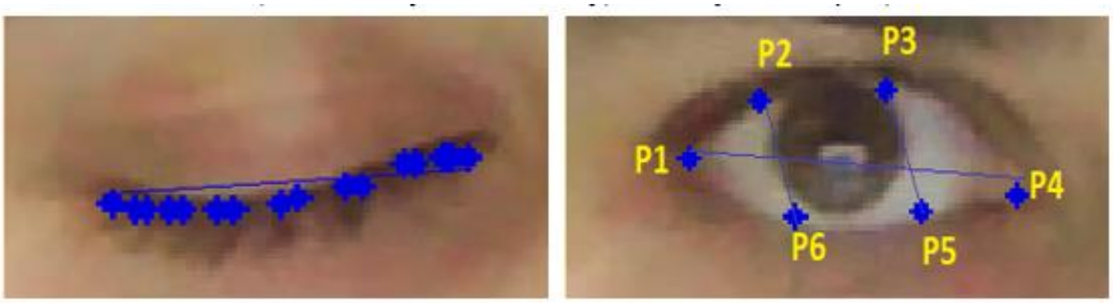


Figure 4.2: Eyelid Movement Detection Using the Aspect Ratio Technique

Figure (4.6) shows two lines drawn on eyes using landmarks, indicating the distance between two landmarks, horizontally and vertically, when eyes are open, the vertical distance reaches it maximum value, while horizontal distance remains constant, on the other hand, when Eyes are closed then vertical distance approaches to it minimum values.

The points in the face detected using MediaPipe will be used, and then, the AR value obtained in Equation (1) formula across sequential frames will be found. If the eyes have been closed exceeded the predefined threshold, we will set off the alarm. “And if the average aspect ratio is more than 4 the counter of blinks will add 1.

Iris Position Detection

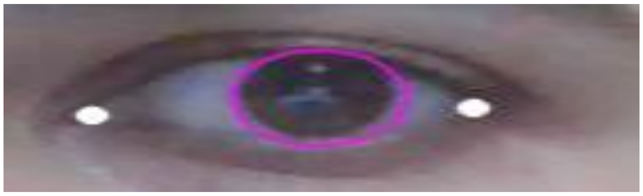


Figure 4.3 : Key Points Used To Detect The Iris Position

MediaPipe Iris is used; MediaPipe Iris function detects 71 eye key points and 5 pupil key points [101]. to recognize where someone is looking landmarks keypoints used were (474, 475, 476, 477) for the left iris and (469, 470, 471,472) for the right iris.

First, a function called minEnclosingCircle in OpenCV library is used to plot a contour polyline around the iris as shown in Figure 4.7), center point coordinates (x, y) of the plotted circles are the assumed center point of the iris, the return values have been in floating type, they should be converted to integer, because pixels cannot be float.

Two reference points are plotted on the right side and left side of each eye and then the aspect ratio is calculated using Equation (2) in chapter 3 with points obtained using the following Equation.

$$\boldsymbol{Aspect\ ratio} = \frac{(Euclidean\ distance(right\ point\ reference, iris\ center\ point)}{Euclidean\ distance(right\ point\ reference, left\ point\ refernce)} \qquad (1)$$

A range was set of the aspect ratio to determine the position of the iris left, right, and center of both eyes’ landmarks and iris landmarks, the iris model takes an image patch of the eye region and estimates both the eye landmarks (along the eyelid) and iris landmarks (along the iris contour).

4.6.3 Lips Disorder Movements Detection

Detecting disorders in lips such as (mouth deviation, mouth puckering, or yawning) can have applications in various applications; such as mouth deviation can be crucial in diagnosing and monitoring conditions like stroke, and Bell's palsy, puckering and jaw movements can aid in orthodontic treatment planning and evaluating jaw disorders Here are some areas where these technologies can be applied, yawning detecting s can be integrated into vehicles to monitor driver fatigue and alertness.

Mouth Open and Puckering Lips Detection

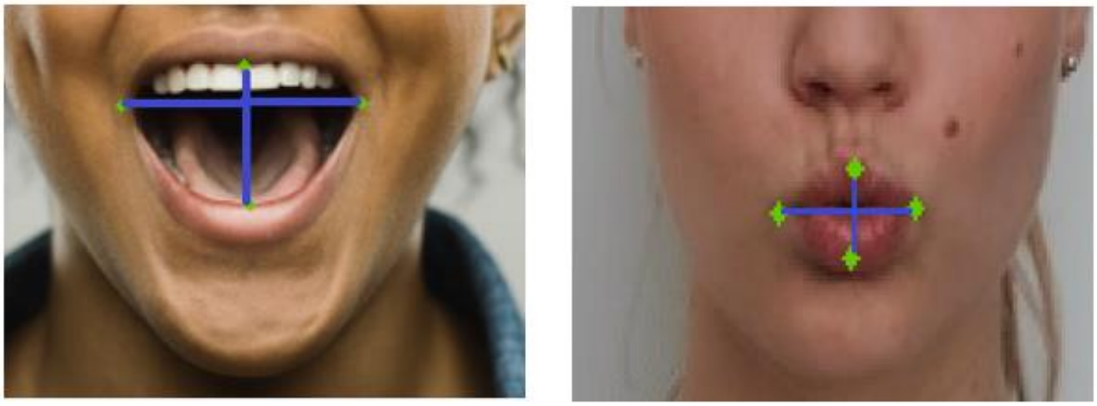


Figure 4.4 : Finding Mouth Aspect Ratio to Detect Pulling and Yawning

To detect mouth irregularities movements, the region of the mouth will be localized, with retrieving keypoints, that be enabled to detect lips status are and used (13, 14, 308, 61) keypoints, where the Euclidean distance between the key points (308, 61) is the horizontal distance and the distance between the keypoints (13, 14) is the vertical distance.

The aspect ratio technique obtained with the points middle point in the upper lip and the middle point in the lower lip as obtained in Equation (2)

$$Aspect\ ratio = \frac{Euclidean\ distance(\text{Upper lip (middle),Lower lip(middle)})}{Euclidean\ distance(\text{Right lip corner,Left lip corner})} \qquad (2)$$

Figure (4.8) shows two lines draw on eyes using landmarks, indicating the distance between two landmarks, horizontally and vertically, when lips are open, the vertical distance reaches its maximum value, while the horizontal distance remains constant, and when lips are closed then Vertical distance approaches to it minimum values.

Mouth Droopy Corner Detection

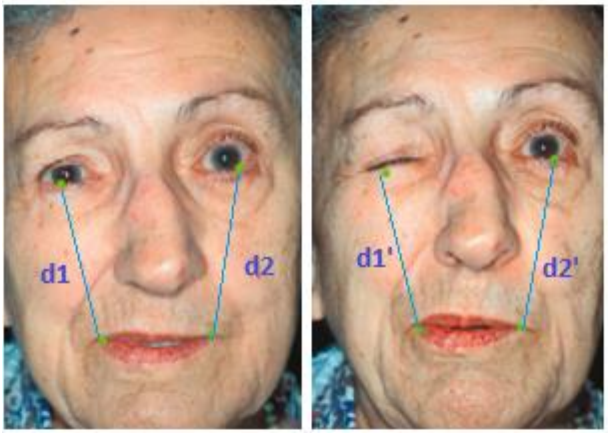


Figure 4.5: Mouth Deviation Detection Technique

In mouth droopy detection, the same technique used in the previous section is being used, the key point coordinates (373, 144, 78, and 308) were retrieved, those points are on the right mouth edge, the right eye lower lid, the left edge of the mouth, and the left eye lower lid respectively, which are shown in Figure (4.9). Then the vertical distance between the points on each side was calculated; the distance between the points on the right side, and the points on the left side.

Figure (4.9) shows the same patient with the palpebral fissure narrows with lip pursing. When the value of the distance between the two values (the distance between the left corner mouth point and left point eyelid point, if the difference between the left vertical distance and the right vertical distance exceeds the predefined threshold value, the system will tell the user that there is a mouth deviation to the left side or the right side.

4.6.4 Drowsy Detection System

To detect drowsy, two facial movements will be detected (eyelid and mouth yawning) besides the waiting time and the threshold values, the same methodology was used in eye blink detection but the threshold average ratio was detected with the user and also the threshold time then the model compares it with the detected AR detection and the threshold time.

This task can be performed by determining two features: the eyes detected remain closed for a continual period of time, and the mouth opens in yawning. First, MediaPipe face mesh solution pipeline will be used to detect eyes and mouth landmarks and demonstrate the Eye Aspect Ratio (EAR) technique to detect drowsiness, then Streamlit library to create a drowsiness detection web application.

Process steps:

1. Predefined thresholds and counters will be set to values, EAR Threshold for eyes and mouth to check whether the real-time evaluated EAR value is within range or not, also Duration Time counter is set to 0 where this is a counter value that tracks the amount of time spent with current EAR < EAR Threshold, and finally the Wait Time value that is set to the value which where the permissible limit passed time where EAR calculated value is less than EAR threshold.
2. The application will record the current time (t1) and start reading the frames.
3. Frames will be passed through MediaPipe's Face Mesh pipeline.
4. Relevant face landmark points (eyes, and mouth landmark points) if any landmark detections are available. Otherwise, reset t1 and also tracked time (Duration-Time) which is the time that passed less than the threshold time.
5. When detections are available, the average EAR value will be calculated for both eyes and mouth using the retrieved landmarks.
6. When the current EAR less than the predefined EAR_THRESH, the difference between (time t2 and t1) to the to D_Time. The model will reset t1 for the next frame as t2.
7. When the D_TIME is more than or equal to the time, the alarm will be on or move on to the next frame.

- **Python Decorator**

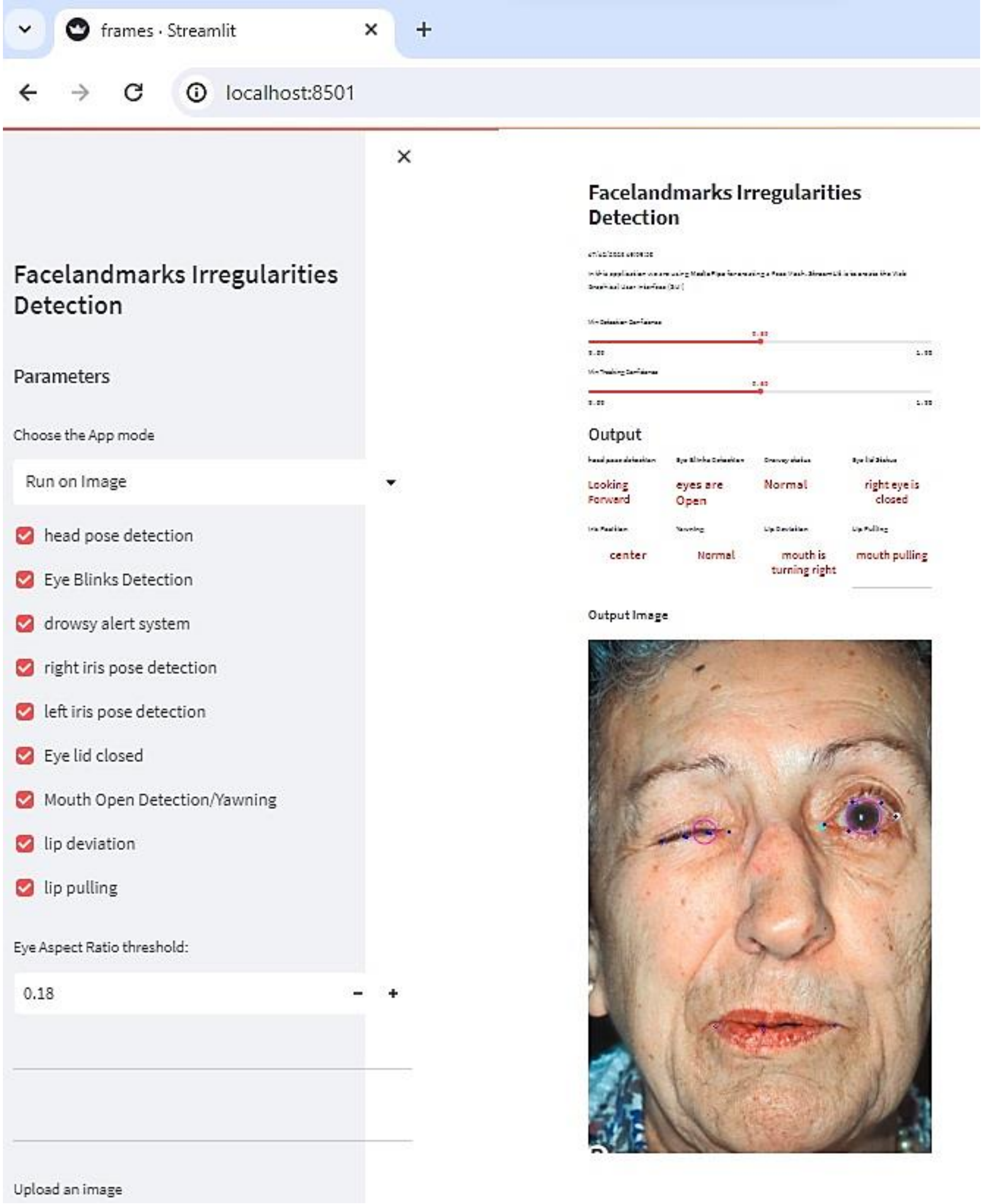
The detector @st.cache was used to make the Streamlit application more efficient; this marking function tells Streamlit to cache the result data and return it to the caller and stops running again as the user interacts with the application as entering a new value.

There are no parameters in our functions, so they will only ever be called once. This is exactly needed; the data is not going to change, so it will only be fetched the first time the functions are called. Thereafter, the cached data will be used.

For example this approach is applied to implement a drowsy detection system application, it can be an alert system used to measure the concentration while driving, to make an alert voice message for the driver if he is fully conscious while driving by detecting the irregularities (yawning or blinking a lot or looking up and down); the proposed system will indicate whether the driver is sleepy upon the detected features, focusing on the eyes if they are approached to be closed, also it will detect yawning, however, the implemented system will detect whether the person doesn't pay attention while driving or the conversation by detecting the iris position, also system detects the eyeblinks and count it, if counts exceed the threshold also it can be marked as an irregular cue.

4.7 Prototype Development User Interface

As mentioned in section (3.7) Streamlit library is loaded to implement the graphical user interface, Streamlit has effective components that allowed us to develop an easy, effective, and attractive dashboard, in the following sections the layout and user interface features and the available options that help the user to detect the facial disorder movements will be introduced,



frames · Streamlit

localhost:8501

Facelandmarks Irregularities Detection

Parameters

Choose the App mode

Run on Image

☑ head pose detection

☑ Eye Blinks Detection

☑ drowsy alert system

☑ right iris pose detection

☑ left iris pose detection

☑ Eye lid closed

☑ Mouth Open Detection/Yawning

☑ lip deviation

☑ lip pulling

Eye Aspect Ratio threshold:

0.18

Upload an image

Facelandmarks Irregularities Detection

[illegible]

In this application we are using MediaPipe for creating a Face Mesh. StreamLit is to create the Web Graphical User Interface (GUI)

Min Detection Confidence

0.50

0.00 1.00

Min Tracking Confidence

0.50

0.00 1.00

Output

head pose detection	Eye Blinks Detection	Drowsy status	Eye lid Status
Looking Forward	eyes are Open	Normal	right eye is closed
Iris Position	Yawning	Lip Deviation	Lip Pulling
center	Normal	mouth is turning right	mouth pulling

Output Image

Figure 4.6: Facelandmarks Irregularities Detection Interface

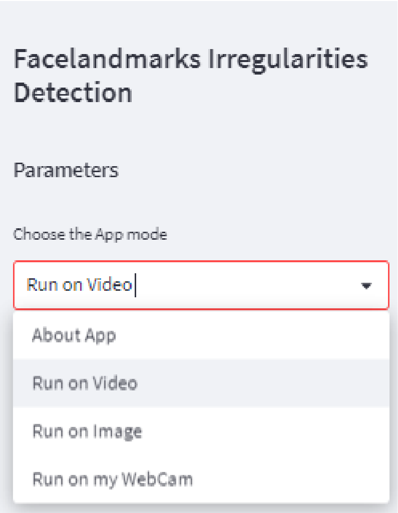
4.7.1 Layout and Components

The components of our interface were allocated in two columns, side by side on the screen, as shown in Figure (4.10), the column on the left side contains the parameters and buttons, and the column on the right side is wider which contains the output widgets.

- **Drop down list to choose the phase:**

The user first has to choose from a drop-down list:

- About app: in this choice, the user will read about the application and can see a short YouTube video about MediaPipe library.
- Run on video: in this choice, the user can upload and run the application with uploaded video to the system to be processed.
- Run on image: in this choice, the user can upload images to be processed.
- Run on my webcam: in this choice, detections will work on the camera, and the resulting detections will be real-time.



Facelandmarks Irregularities Detection

Parameters

Choose the App mode

Run on Video

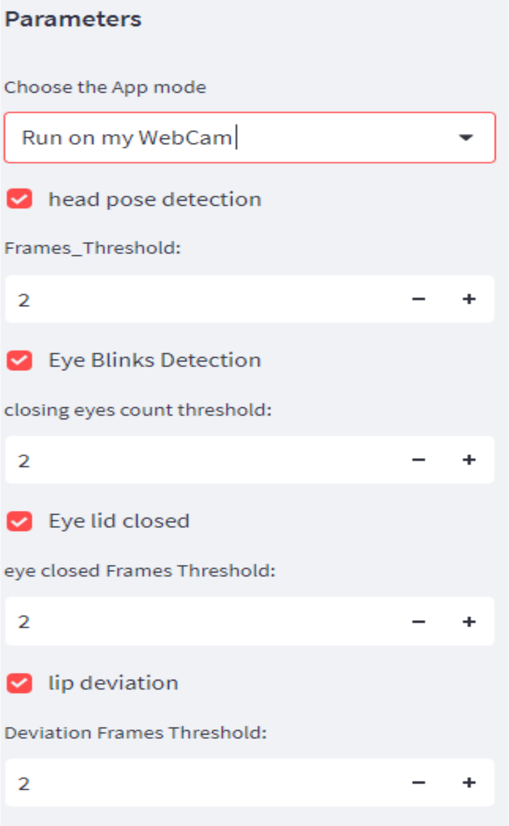
Figure 4.7: The Phases User Can Feed the Input

- **Checkboxes to choose the features:**

After choosing the phase that the application will work on, the user will see a list of checkboxes to choose which disorder movements will be detected, when the user chooses the required detections by checking the checkboxes each checkbox invokes developed functions in the python code.

The features that can be detected are:

- Head pose detection: this choice detects whether the person in the output frame is looking left, looking right, looking down, or looking forward.
- Eyeblink: in this choice, the user will detect blinks for both eyes
- Left eyelid closure: when this checkbox is set will detect the left eye status (open Left eye, half closed Left eye, closed Left eye)
- Right eye lid when this checkbox is set will detect the right eye status (open right eye, half closed eye, closed eye)
- Lip deviation: when this checkbox is set will detect the droopy in mouth corners it will classify whether (the lips are in the normal state, mouth deviation to the right, or mouth deviation to the left).
- Lip pulling: when this checkbox is set will detect whether the mouth is puckered
- Yawning detection: when this checkbox is set will detect whether the person yawned during the video.
- Left iris position: when this checkbox is set will classify the position of the left iris (in the left, in the right, and center).
- Right iris position: when this checkbox is set will classify the position of the right iris (in the left, in the right, and center).
- Drowsy alert system: when this checkbox is checked, the system will detect whether the person feels sleepy or not and the features detected are discussed in section (4.5.4).
- Sidebar Input Number to Choose the Threshold Number of Frames:
- Sidebar input number using the function st.number_input" has been chosen. as an interface component to allow the user to input numerical entries within the sidebar and that number will be the threshold number of frames, and another sidebar to detect the seconds threshold that is used in the drowsy detection system that was described in section (4.4), and based on this number the disorder movement will be decided as a detected irregularity as discussed in Chapter (4).



Parameters

Choose the App mode

Run on my WebCam

☑ head pose detection

Frames_Threshold:

2

☑ Eye Blinks Detection

closing eyes count threshold:

2

☑ Eye lid closed

eye closed Frames Threshold:

2

☑ lip deviation

Deviation Frames Threshold:

2

Figure 4.8: Parameters to Fed into the System

4.7.2 Buttons

The buttons were used to invoke two types of functions:

- File Picker

the button with the upload function st.file_uploader was added to provide a file uploader, which is accessible, when the user clicks the button to upload an image or video, the upload function will run.

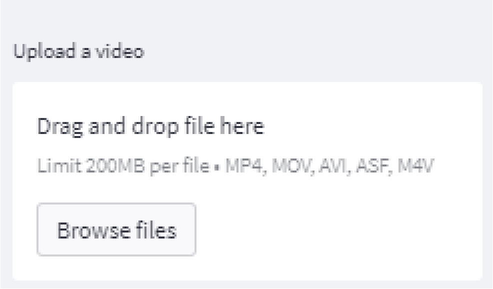


Figure 4.9: Upload Video Button

- **Run the system**

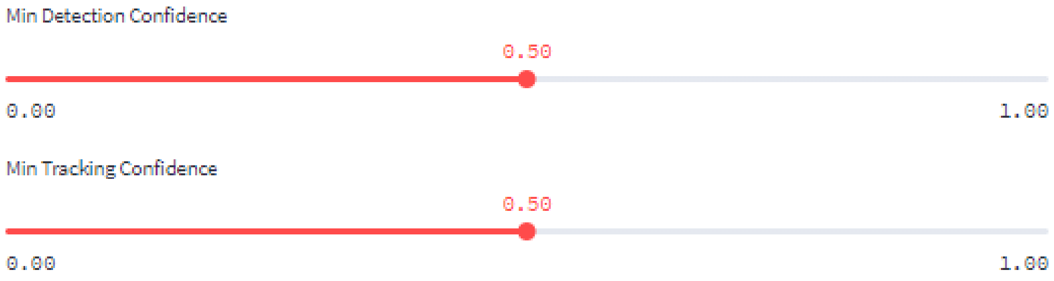
Button to run the system and to trigger the model were created, and to have results, as the user presses on the run button, he can see the output image and the output sections and detections classifications in the second column that will be discussed in the next section.

Figure 4.10: Button to Trigger the Model

○ **Sliders to Enter the detection Confidence**

The slider is an easy component that allows the user to choose a value by sliding the slider to the desired value in a defined range between two elements passed by the developer.

In this application, the sliders were used to allow the user to choose critical values in the model configuration Minimum Detection Confidence, and Minimum Tracking Confidence, both sliders allowed to choose between 0 and 1, and the system will work on the entered values.



Min Detection Confidence

0.50

0.00 1.00

Min Tracking Confidence

0.50

0.00 1.00

Figure 4.11: Slider to Choose Mediapipe Configuration Parameters

4.7.3 The Output Results

To print the output results, side-by-side eight columns were inserted as a container, eight labeled multi-element containers were inserted, to print out the results in the returned containers.

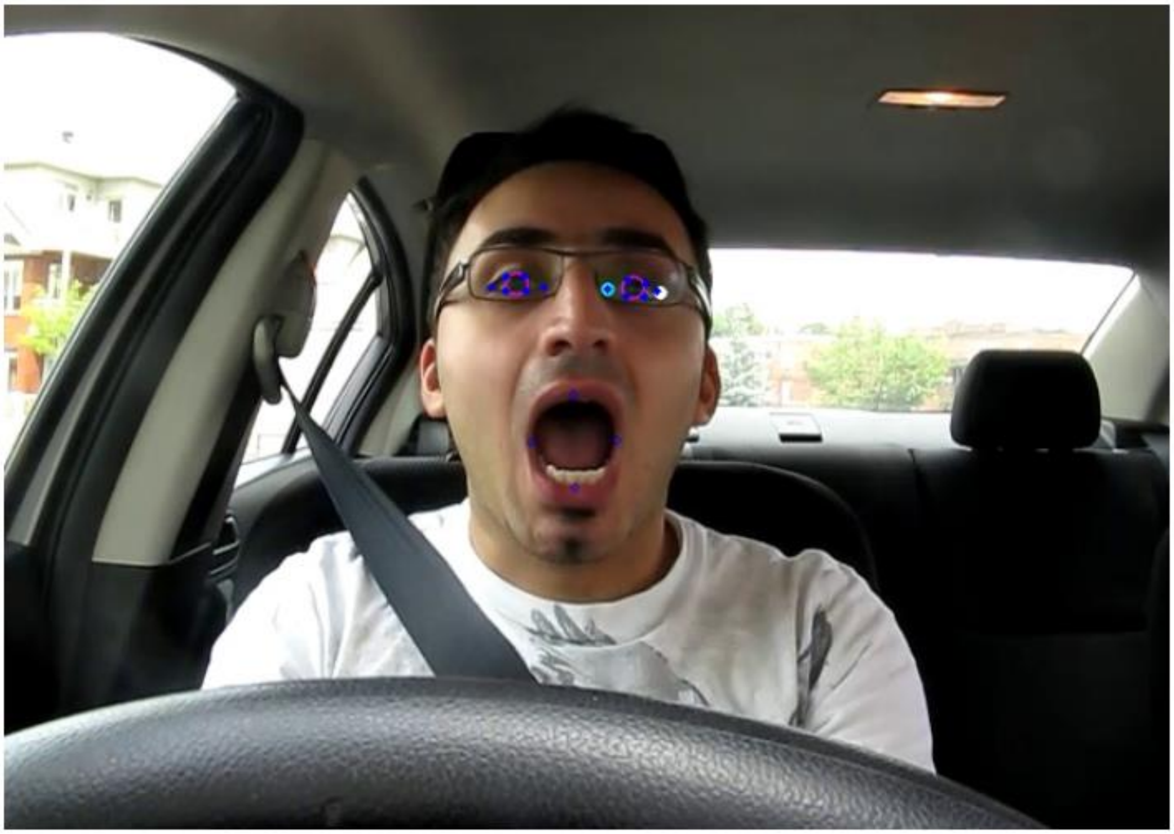
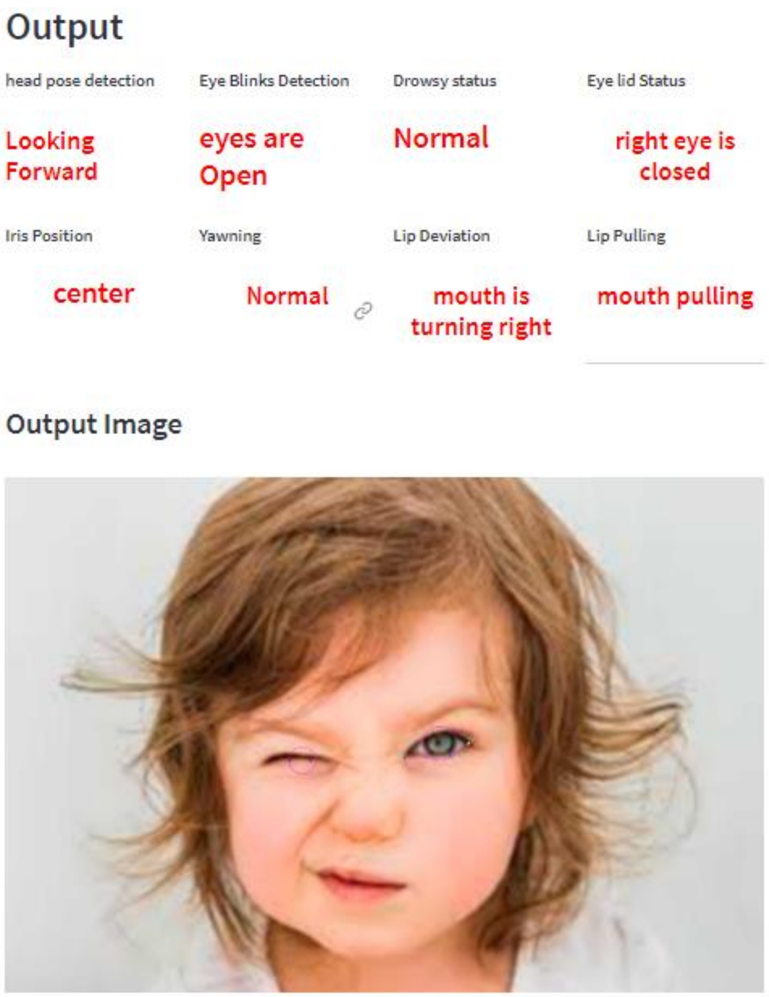


Figure 4.12: Output With Visualized Facial Landmarks On an Input Image.

When the user presses the run button shown in Figure (4.16), he will see the frames fed by the web camera, uploaded video, or uploaded images with the annotated points uploaded video or image as in Figure (4.17).



Output

head pose detection: Looking Forward

Eye Blinks Detection: eyes are Open

Drowsy status: Normal

Eye lid Status: right eye is closed

Iris Position: center

Yawning: Normal

Lip Deviation: mouth is turning right

Lip Pulling: mouth pulling

Output Image

Figure 4.13: Output Results of the Implemented System

4.8 Chapter summary

The chapter delves into the methodology I employed for detecting head position and disorders detected in eyes and mouth region, using and advanced technologies such as MediaPipe and aspect ratio technique. The key steps in the methodology included the type of dataset that model can handle with it, and the preprocessing steps (image flip, and converting to RGB), Mediapipe configuration, choosing the acquiring landmarks, movement analysis techniques and finally the implemented system and its components was described.

Chapter 5: Model Evaluation

5.1 Introduction

The performance and effectiveness model should be evaluated to ensure that the model functions effectively and, there is a need to prepare the dataset and aggregate the evaluation metrics across all samples to get an overall assessment of the model's performance. Report the evaluation findings, including accuracy and other relevant metrics, to summarize the model's classification capabilities.

5.2 Evaluation Criteria

To comprehensively assess the reliability of the system, we focus on three key criteria:

- Robustness: The system must be able to handle a variety of data inputs (Realtime video, recorded videos, and images). Therefore, the robustness of the method must be tested for the properties.
- Accuracy: precision, recall, and F1-score metrics were employed to measure the system's ability to correctly detect facial macromovements or micromovements.
- Performance: performance is the most important requirement of the system. The implemented system should be able to classify facial movements as accurately as possible and the number of mis-classifications has to be minimal. while accuracy is a crucial indicator, a model's real-time performance is determined by a wider range of factors, including speed, resource efficiency, flexibility, and the model's capacity to operate well under changing and dynamic circumstances. These aspects must be considered in addition to accuracy when determining whether a model is appropriate for real-time applications.
-

5.3 Dataset Used in The Evaluation of The Detection

To evaluate the implemented model, evaluation dataset was prepared; many images were collected from several datasets, and images from (AFLW2000-HeadPose, Helen, AFW, 300-W, NITYMED, Driver Drowsiness Dataset (DDD)) were chosen ensuring a wide range of facial movements. These datasets were described in the following points:

- Annotated Facial Landmarks in the Wild (AFLW2000-HeadPose): this dataset contains 2000 images for evaluation of 3D facial landmark detection models with annotation files that contain 68-point 3D facial landmarks this dataset was used for evaluating the head position feature in the implemented model [100].
- Helen: This dataset is composed of 400×400 pixels labeled 2330 face images of through manually-annotated contours along eyebrows, eyes, nose, lips and jawline; this dataset is used for evaluating the iris position feature [101].
- Annotated Faces in the Wild (AFW): this dataset is used often for face detection models performance evaluation; it contains labeled 205 images with 468 faces [102].
- 300-W: It consists of a large collection of facial images focus on facial landmark detection, with ground truth annotations of 68 facial landmarks. It is always used to understand and analyze facial geometry and expressions.
- Driver Drowsiness Dataset (DDD): This dataset consists of 227 x 227 pixels labeled 41,790 RGB images that were extracted as frames from the videos of the Real-Life Drowsiness Dataset (RLDD). images were split into directories (Drowsy & Non-Drowsy) [103].
- NITYMED: This dataset consists of 130 videos, captured in Greece and Patras, it displays drivers (10 females and 11 males with different features (glasses, hair color, beard, etc.)) in real cars, The videos were split into two categories: Yawning: (107 videos) the drivers yawn many times in each video, look around and have and microsleep (21 videos) [104].

5.4 Evaluation Methodology and Performance Metrics

To evaluate the detection and classification MediaPipe model, the following general steps were followed: the trained MediaPipe classification model was loaded. After that the images in the directory will be passed through the MediaPipe model to obtain the predicted class feature. to assess the model's accuracy, the predicted Facial Action Units will be compared with the ground truth Facial Action Units.

The evaluation metrics have been used to provide insights into the model's performance such as accuracy, precision, recall, and F1 they are calculated using a ground truth dataset and the predicted labels from our model, also considered the duration between the input of a video frame and the output of tic detection was used to calculate latency.

5.5 Results and Discussion

In this section, the prepared created and labeled dataset in each directory used for multiple evaluation tasks (head position, Eye lid movement detection, iris position, yawning, drowsiness, mouth deviation (mouth droopy corners)).

- Head Position Evaluation Results

Figure 5.1: Examples Dataset Used for Head Position Detection

Head position MediaPipe classification model was loaded and configured. After that images were passed in the directory through the MediaPipe model to obtain the predicted class feature (looking forward, looking left, looking right).

the evaluation metrics results were found as in the Table (5.1), and confusion matrix was visualized is obtained in Figure(5.2).

Table 5.1:Head Position Classification Report

Class	precision	recall	f1-score	accuracy
Looking Forward	1.00	1.00	1.00	100%
Looking Left	1.00	1.00	1.00	
Looking Right	1.00	1.00	1.00	

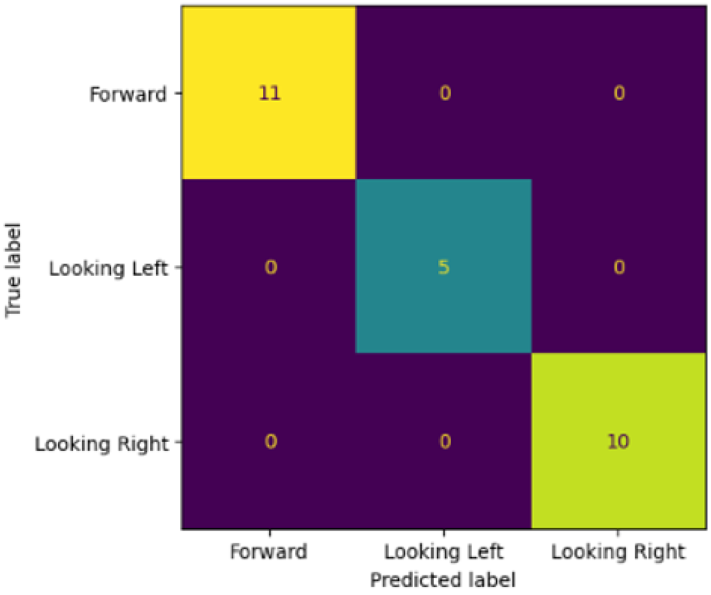


Figure 5.2: Confusion Matrix of Head Position

- Eye Lid Status

MediaPipe classification model was loaded and configured. and predefined the thresholds EAR ratio and after those images were passed in the directory through the MediaPipe model to obtain the predicted class feature (both eyes are in the same status, the right eye is closed, left eye is closed). The images shown in Figure 5.3 obtain examples.

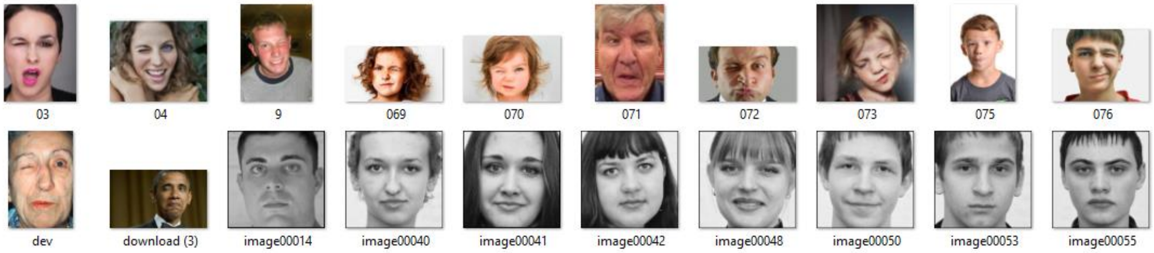


Figure 5.3: Examples Dataset Used for Eye Lid Detection

Evaluation metrics results are summarized in Table (5.2) and confusion matrix was visualized and is obtained in Figure (5.5).

Table 5.2: Eye led movements detection Classification Report

Class	Precision	recall	f1-score	accuracy
both eyes are in the same status	0.86	0.86	0.86	86%
left eye is closed	0.80	0.80	0.80	
right eye is closed	0.90	0.90	0.90	

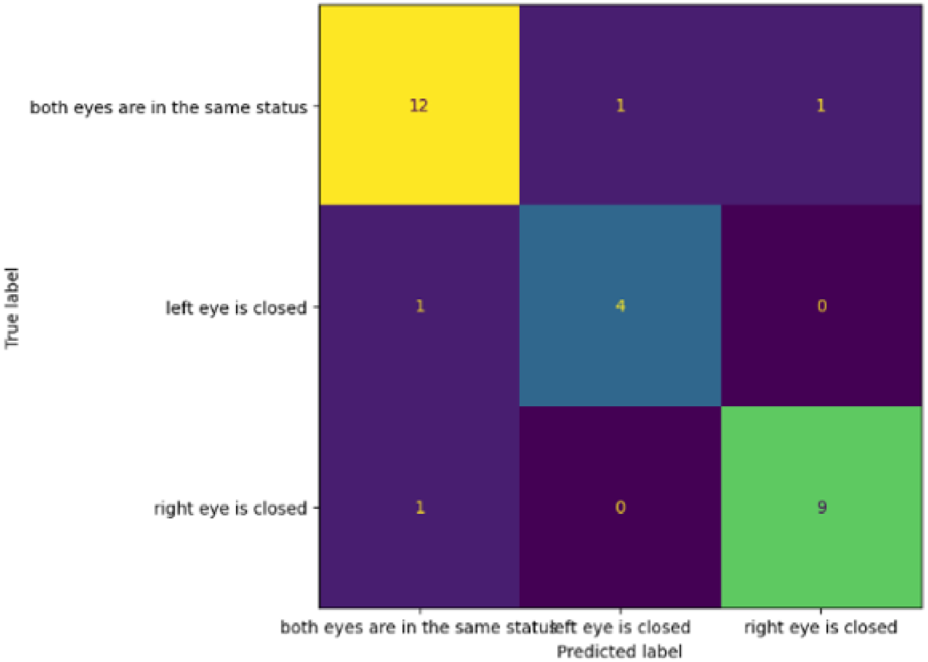


Figure 5.4: Confusion Matrix Results for Eyelid Status

- Iris Position

MediaPipe classification model was loaded. Two reference points were plotted on the right side and left side of each eye and then the aspect ratio is calculated using Equation (3) (Center, Left, and Right) the images shown in Figure 5.4 obtains examples.

Figure 5.5: Iris Position Samples for Evaluating Dataset

Evaluation metrics results are summarized in Table (5.3) and visualized confusion matrix and is obtained in Figure (5.6).

Table 5.3: Evaluation Results of Iris Position Status

Detection	precision	Recall	f1-score	accuracy
Center	0.90	1.00	0.95	96%
Left	1.00	1.00	1.00	
Right	1.00	0.88	0.93	

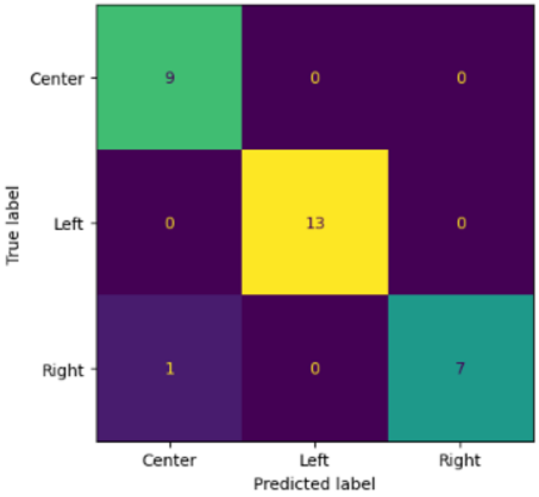


Figure 5.6: Evaluation Results of Iris Position Status

- Yawning

MediaPipe classification model was loaded. and predefined the thresholds EAR ratio in of the mouth as discussed in chapter (4) to two classes (yawning, not yawning) after that, images were passed in the directory through the MediaPipe model to obtain the predicted class feature yawning. The images shown in Figure (5.7) obtains examples.

Figure 5.7: Yawning and Not Yawning Samples for Evaluating Dataset

Evaluation metrics results are summarized in Table (5.4) and visualized confusion matrix and is obtained in Figure (5.8).

Table 5.4: Evaluation Results for Yawning Detection

Detection	precision	Recall	f1-score	Accuracy
Yawning	1.00	0.95	0.97	96%
Not Yawning	0.88	1.00	0.93	

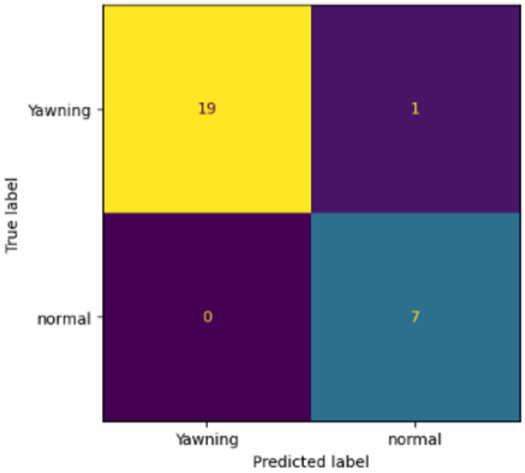


Figure 5.8: Confusion Matrix Results for Eyelid Status

- Mouth Deviation or Droopy Corner

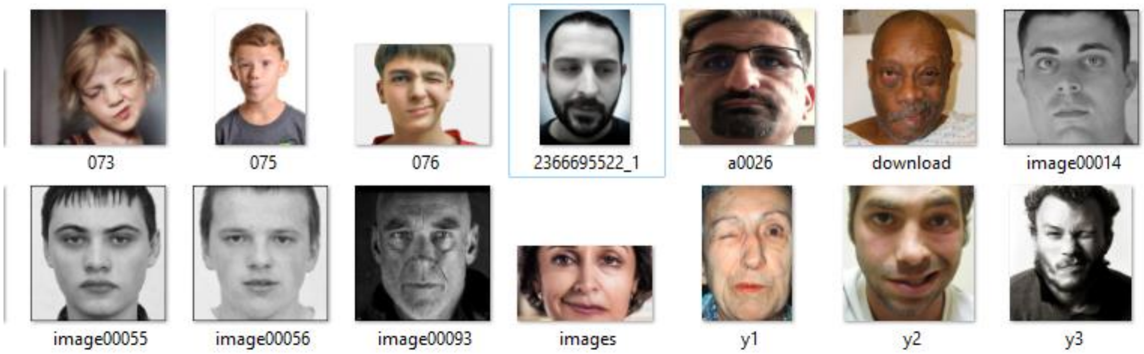


Figure 5.9: Mouth Deviation and Normal Status Samples for Evaluating Dataset

MediaPipe classification model was loaded. And predefined the thresholds difference of the vertical distance between the middle point on lower lip and the middle point on the upper and the horizontal distance between mouth corner points as discussed in section (4.5.3) after that images were passed in the directory through the MediaPipe model to obtain and predict labels (Normal status, mouth is turned to the left, and mouth is turned to right) the examples of images shown in Figure (5.9) ,and evaluation metrics results are summarized in Table (5.5) and visualized confusion matrix and is obtained in Figure (5.10).

Table 5.5: Evaluation Results of Mouth Deviation Status

Detection	precision	recall	f1-score	accuracy
Normal	1.00	0.79	0.88	88%
mouth is turned the left,	0.75	1.00	0.86	
mouth is turned right	0.86	1.00	0.92	

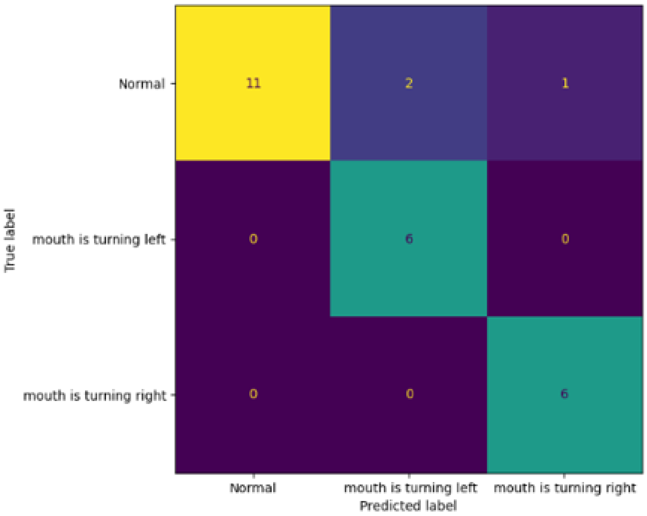


Figure 5.10: Confusion Matrix Results for Mouth Deviation Status

- Drowsy detection

MediaPipe classification model was loaded. and predefined the threshold average EAR of two eyes as discussed in chapter (4) to three classes after that images in the directory were passed through the MediaPipe model to obtain the predicted class feature (Normal status, mouth is turned to the left, and mouth is turned to right). The images shown in Figure (5.11) and evaluation metrics results are summarized in Table (5.6) and visualized confusion matrix and is obtained in Figure (5.12).

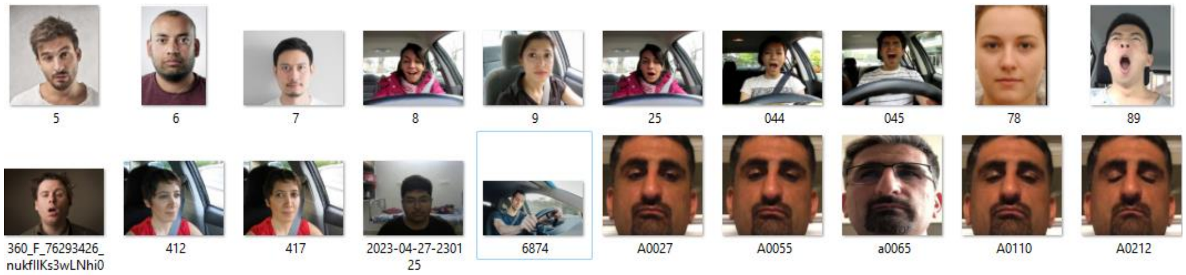


Figure 5.11: Sleepy Eyes and Normal Status Samples of Evaluating Dataset

Table 5.6: Evaluation Results of Sleepy Eyes Status

Detection	precision	recall	f1-score	accuracy
Normal	0.92	1.00	0.96	97%
Sleepy	1.00	0.95	0.98	

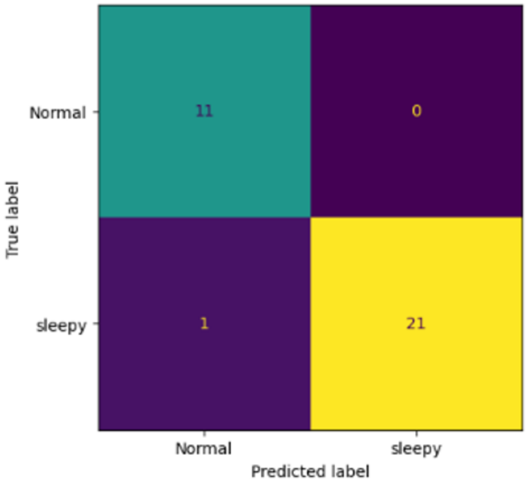


Figure 5.12: Confusion Matrix Results for Mouth Drowsy Status

- Mouth Pulling Detection Evaluation

MediaPipe classification model was loaded. and predefined the threshold ratio of the vertical and horizontal distances as discussed in section (4.5.3) and Equation (4) and then classified it to two classes (Normal, mouth pulling) after that image were passed in the directory through the MediaPipe model to obtain the predicted class feature. The images shown in Figure (5.13) obtains the dataset used for the evaluation of this feature. and evaluation metrics results are summarized in Table (5.7) and visualized confusion matrix and is obtained in Figure (5.14).

Figure 5.13: Mouth Puckering and Normal Status Samples of Evaluating Dataset

Table 5.7: Evaluation Results of Mouth Puckering Status

Detection	Precision	Recall	F1-Score	Accuracy
Normal	1.00	0.78	0.88	91%
Mouth Pulling	0.88	1.00	0.93	

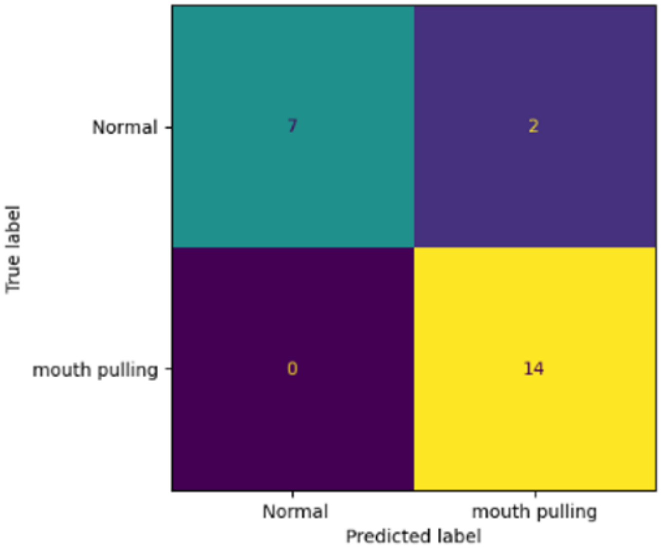


Figure 5.14: Confusion Matrix Results for Mouth Pulling

5.5.1 Conclusion

Table 5.8:Results Summary

Detection	Class	Precision	Recall	F1-Score	Accuracy
Head Position	Looking Forward	1.00	1.00	1.00	100%
	Looking Left	1.00	1.00	1.00	
	Looking Right	1.00	1.00	1.00	
Mouth Puckering Detection	Normal	1.00	0.78	0.88	91%
	Mouth Pulling	0.88	1.00	0.93	
Drowsy Eyes Detection	Normal	0.92	1.00	0.96	97%
	Sleepy	1.00	0.95	0.98	
Yawning Detection	Yawning	1.00	0.95	0.97	96%
	Not Yawning	0.88	1.00	0.93	
Each Eyelid Status	Both Eyes are in the Same Status	0.86	0.86	0.86	86%
	Left Eye Is Closed	0.80	0.80	0.80	
	Right Eye is Closed	0.90	0.90	0.90	
Iris Position	Center	0.90	1.00	0.95	96%
	Left	1.00	1.00	1.00	
	Right	1.00	0.88	0.93	
Mouth Deviation or Droopy Corner Detection	Normal	1.00	0.79	0.88	88%
	Mouth is Turned to the Left	0.75	1.00	0.86	
	Mouth is Turned Right	.0.86	1.00	0.92	

The model performed well in head position classification, iris position, and eyelid status evaluation, with 96% accuracy. However, it missed one classification due to an optimal threshold value. The system also detected micromovements, but errors occurred when one eye was closed due to congenital differences in eye size. Mouth deviation detection also had errors, with 94% accuracy for mouth pulling. Drowsy eyes detection had 97% accuracy, but the model missed detections when the threshold was too close to the correct classification. Yawn detection had a high accuracy of 96%, but errors occurred when the threshold did not fit the face.

5.6 MediaPipe Landmarks Performance Evaluation Case Study

Dlib library is used widely in computer vision solutions vision. it is based on the C++ language. It enables the user to find 68 landmark points that are in the range (0,67) that are described in Table (4.1).

In this case study, both Dlib library and MediaPipe were loaded on the same dataset Images;

Table 5.9: Dlib Key Points and Facial Landmarks

Dlib key points	**facial landmarks provided by Dlib face landmark model**
0 to 16	Jawline
17 to 21	Right eyebrow landmarks
22 to 26	Left Eyebrow
27 to 35	Nose landmarks
36 to 41	Right eye landmarks
42 to 47	Left Eye landmarks
48 to 60	Outline of the Mouth
61 to 67	Inner line of the Mouth

On the other hand, MediaPipe detects 469 facial landmark points and 72 points are for face oval that are:

(0, 7, 10, 13, 14, 17, 21, 33, 37, 39, 40, 46, 52, 53, 54, 55, 58, 61, 63, 65, 66, 67, 70, 78, 80, 81, 82, 84, 87, 88, 91, 93, 95, 103, 105, 107, 109, 127, 132, 133, 136, 144, 145, 146, 148, 149, 150, 152, 153, 154, 155, 157, 158, 159, 160, 161, 162, 163, 172, 173, 176, 178, 181, 185, 191, 234, 246, 249, 251, 263, 267, 269, 270, 276, 282, 283, 284, 285, 288, 291, 293, 295, 296, 297, 300, 308, 310, 311, 312, 314, 317, 318, 321, 323, 324, 332, 334, 336, 338, 356, 361, 362, 365, 373, 374, 375, 377, 378, 379, 380, 381, 382, 384, 385, 386, 387, 388, 389, 390, 397, 398, 400, 402, 405, 409, 415, 454, 466), many of these keypoints were described in them in Table(5.10), but only 68 keypoints of them in our evaluation were chosen that are

(21, 162, 93, 132, 172, 136, 149, 148, 152, 377, 378, 365, 397, 288, 323, 454, 389, 70, 63, 105, 66, 107, 336, 296, 334, 293, 301, 168, 197, 5, 4, 75, 97, 2, 326, 305, 33, 160, 158, 133, 153, 144, 362, 385, 387, 263, 373, 380, 61, 39, 37, 0, 267, 269, 291, 405, 314, 17, 84, 181, 78, 82, 13, 312, 308, 317, 14, 87)

In this experiment MediaPipe Face Landmark model will be evaluated on a labeled dataset that includes ground truth data, then I will Compare the output prediction generated by the MediaPipe model with the ground truth data to determine if the prediction is correct or not.

5.6.1 Dataset Used for Evaluating Detecting 68 Keypoints Face Oval Using Mediapipe

To evaluate the performance of the MediaPipe model we used 300-W dataset, this dataset is widely used for face landmarks. It contains a large collection of images with PNG extension and each image has a file with pts extension contains annotations for 68 keypoints, that annotate the eyes, nose, mouth, and other facial features for each image, the landmarks are in the range (0,67). It covers a large variation of pose, occlusion, identity, expression, face size, and illumination conditions. In this experiment 87 indoor images were carefully selected, with an overall mean size is 85k (about 292 × 292) pixels.

5.6.2 Methodology used for MediaPipe performance evaluation:

To evaluate a MediaPipe face landmark detection model using the 300-W dataset, the following steps were followed:

1- dataset preparation by extracting the files and organizing them by putting the images in one directory and the annotation files in another directory.

2. dataset loading (images and annotations into Python script. Using OpenCV library to read the images and to process the annotation files.

3. creating and initiating the MediaPipe face landmark model using the class mp.solutions.face_mesh.FaceMesh()

4. dataset iterating through each image in the directory of images.

5-Reading and processing the images Read the image and convert it to the RGB format expected.

6. Passing the image through the MediaPipe face landmark model using the process () method.

7. Retrieving and extracting the predicted landmarks from the MediaPipe model's output using the attribute multi_face_landmarks.

8. Comparing the corresponding ground truth landmarks in the dataset annotations with the predicted landmarks using MediaPipe.

9. Calculate evaluation metrics (accuracy, precision, and recall) based on the distance between predicted and ground truth landmarks.

5.6.3 Results and Discussion

The annotated landmarks were printed on the images of the 300W dataset, Figure 5.7(A) shows a sample that is annotated with landmarks in the pts file in with blue color, wherearea the image in the Figure (5.15 B) shows the annotations of the points tried to a rough mapping based on the general facial landmarks the points keypoints used are

(21, 162, 93, 132, 172, 136, 149, 148, 152, 377, 378, 365, 397, 288, 323, 454, 389, 70, 63, 105, 66, 107, 336, 296, 334, 293, 301, 168, 197, 5, 4, 75, 97, 2, 326, 305, 33, 160, 158, 133, 153, 144, 362, 385, 387, 263, 373, 380, 61, 39, 37, 0, 267, 269, 291, 405, 314, 17, 84, 181, 78, 82, 13, 312, 308, 317, 14, 87)

Figure (5.15) shows landmarks on a sample image presented in yellow color and chosen predicted landmarks in MediaPipe) that drew the ground truth landmarks in the dataset with the blue color and the third image both landmarks of two libraries are visualized on the same sample.

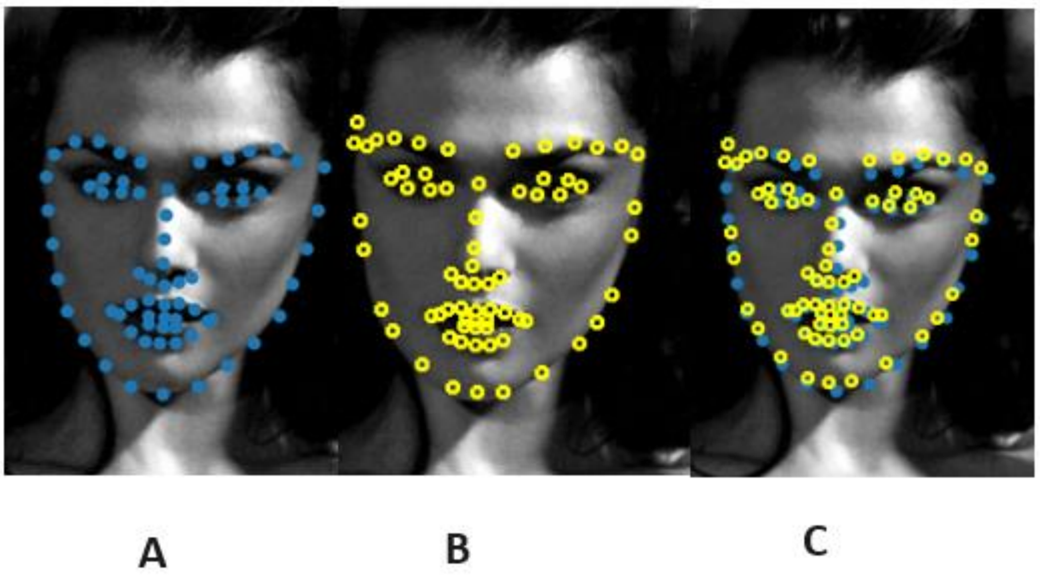


Figure 5.8: Mediapipe, Dlib Facial Landmarks Visualized on a Sample of the Dataset

The face oval key points that can be detected in MediaPipe library are

(0, 7, 10, 13, 14, 17, 21, 33, 37, 39, 40, 46, 52, 53, 54, 55, 58, 61, 63, 65, 66, 67, 70, 78, 80, 81, 82, 84, 87, 88, 91, 93, 95, 103, 105, 107, 109, 127, 132, 133, 136, 144, 145, 146, 148, 149, 150, 152, 153, 154, 155, 157, 158, 159, 160, 161, 162, 163, 172, 173, 176, 178, 181, 185, 191, 234, 246, 249, 251, 263, 267, 269, 270, 276, 282, 283, 284, 285, 288, 291, 293, 295, 296, 297, 300, 308, 310, 311, 312, 314, 317, 318, 321, 323, 324, 332, 334, 336, 338, 356, 361, 362, 365, 373, 374, 375, 377, 378, 379, 380, 381, 382, 384, 385, 386, 387, 388, 389, 390, 397, 398, 400, 402, 405, 409, 415, 454, 466)

Results show that there are variations in the positions; specific mapping between the MediaPipe face landmark 468 key points and the Dlib library 68 key points are not exact and may not be perfect, and that due to the difference in the number of key points of face oval key points and variations in their positions.

5.7 Summary

Walkthrough of face landmark detection MediaPipe library and compared it with 68 landmark ideal positions with Dlib key points. MediaPipe is easier than other libraries to install and does not need a dependency, unlike Dlib. MediaPipe can detect the 3D coordinates of the key points while Dlib library can detect only the 2D coordinates of the key points which makes MediaPipe can be used for head pose estimation. Finally, with MediaPipe the facial area and landmarks can be extracted much more sensitive than Dlib.

Chapter 6: Conclusion and Future Work

Machine Learning (ML) techniques were employed as a tool for irregularities detection in non-verbal cues. The proposed system can detect multiple non-verbal cues in real-time and it works to reduce the false alert times during the process of detection. It is useful to create a system that can be used to estimate patient's status or to interpret and express their attitudes through their non-verbal cues; and the motivation for this approach is the that we manually defined the model using the facial landmark distances ratio with no need for the dataset to be trained, and the proposed system it can be applied widely in various applications, as a new framework for real-time behavioral analysis.

6.1 Conclusion

In this thesis, a general focus on nonverbal cues was discussed.

Chapter 2 literature review has presented an overview of verbal and nonverbal cues and then focused on kinesics cues which are visible body movements and posture micromovements, that send messages about Attitude towards the other person, Emotional states, and Desire to control the environment, and we discussed Detecting and identifying facial expressions is necessary in several fields, such as health, politics, business, marketing, airport security, criminal investigations, and education. And also discussed used techniques and the types of Abnormal facial movements (facial tics and nonverbal cues symptoms. the relevant works related to machine learning approaches used in facial disorder movement detection. And found that MediaPipe model supports more complicated behavior and different types of data than neural network models.

In Chapter 3: A novel facial framework named "facial irregularities movements detection" was proposed. We set advantages of the proposed framework, then the detected cues were defined, the system implementation requirements were discussed, The MediaPipe Face

Mesh, which creates a mesh of the approximate face geometry, is applied in the pipeline's initial stage choosing a specific keypoints on the original image may utilize it facial landmarks tracking. The pipeline is constructed as a MediaPipe graph that makes use of a specific keypoints-and-depth renderer subgraph, the specific facial landmarks subgraph from the iris landmark module, and a face landmark subgraph from the face landmark module, then detects irregularities depending on (the distance between the chosen points, classify the movement among the distance or angles and the predefined thresholds and the number of frames.

In Chapter 4: The methodology and steps for each type of detection were described, in this chapter we described the methodology and the way we used the techniques were described in Chapter 3 for each feature (head position, eyelid movement detection, iris position, yawning, drowsiness, mouth deviation(mouth droopy corners)) and applied as real-time assisting system for on real-time front face laptop camera, and videos and images that can be uploaded, and described general steps we followed to detect movement. We described the specifications of our implemented Streamlit interface, we described all the components of the interface and all choices, Streamlit is a Python library that allowed us to build an interactive data-driven web app, user can make multiple types of detections in real-time webcam, uploaded videos, and uploaded images we presented widgets as basic input checkboxes, slider bars, etc.

In Chapter 5. We evaluated the implemented model and evaluated a MediaPipe model to identify potential issues, to assess its performance, and to make informed decisions about its improvement and deployment. To evaluate the detection and classification MediaPipe model, we followed these general steps: we loaded the trained MediaPipe classification model. And passed the images in the directory through the MediaPipe model to obtain the predicted class feature. to assess the model's accuracy, we finally compared the predicted features with the ground truth features. Calculated evaluation metrics (accuracy, precision, recall, F1) based on the distance between predicted and ground truth landmarks. And found the accuracy as follows 100% for the head position on, 86% for eye led status, 96% for yawning detection, 96% for mouth deviation, 97% for eyes drowsy detection, 100% mouth pulling.

6.2 Key Challenges:

During the study, we encountered many challenges and we summarized many of them in this section:

- Model Selection and Architecture: Choosing the optimal MediaPipe model architecture for our specific tasks was challenging. However different tasks often require different architectures and models, and selecting the right one involves experimentation and expertise.
- Overfitting and Underfitting: Balancing the model's complexity to avoid overfitting: our model depends and classifies by comparing the Euclidean distance of the threshold with the predicted landmarks using MediaPipe. It was required to be careful to ensure that our model.
- Label noise and annotation errors: Ensuring label quality is very important. Noisy or incorrect labels in our testing data can lead to an incorrect evaluation.
- Interpreting results: Understanding and explaining why the model is making certain predictions can be difficult. Interpretability is crucial for confidence and diagnosing incoming problems.
- Research limitations: Limited researches and experiments that used MediaPipe model as a tool for detections, and also limited researches that automated the detection of irregularities in facial movements.

We overcame these challenges by having a clear plan, and a willingness to iterate and learn from our model's performance, continuous learning, and a deep understanding of specific tasks and data were key factors in successfully implementing our MediaPipe model system.

As a plan for future direction, we do recommend to detect more irregularities and recommended features, and allow the user to change many parameters to reach the best for each case. Also, we will test the models on diverse and challenging datasets across different lighting environments, and conditions. We do recommend to implement techniques to handle occlusions, partial views, or other scenarios that might affect model accuracy. In addition, test the model on more diverse and challenging datasets to assess its robustness across different conditions, and lighting environments. And finally, implement techniques to handle occlusions, partial views, or other scenarios that might affect model accuracy.

References

ZHAO, Guoying, et al. Facial Micro-Expressions: An Overview. Proceedings of the IEEE, 2023.

YANG, Huizhou, et al. Impact of patient, surgical, and implant design factors on predicted tray–bone interface micromotions in cementless total knee arthroplasty. Journal of Orthopaedic Research®, 2023, 41.1: 115-129..

ANNADURAI, Swaminathan; AROCK, Michael; VADIVEL, A. Real and fake emotion detection using enhanced boosted support vector machine algorithm. Multimedia Tools and Applications, 2023, 82.1: 1333-1353..

MURAYAMA, Hirokazu; SUZUKI, Kaiyu; MATSUZAWA, Tomofumi. Evaluation of Accuracy Degradation Resulting from Concept Drift in a Fake News Detection System Using Emotional Expression. Applied Sciences, 2023, 13.10: 6054..

GOSTAND, Reba. Verbal and non-verbal communication: drama as translation. In: The Languages of Theatre. Pergamon, 1980. p. 1-9..

LIU, Meina. Verbal communication styles and culture. In: Oxford research encyclopedia of communication. 2016..

WHARTON, Tim. Pragmatics and non-verbal communication. Cambridge University Press, 2009..

THYSSEN, Ole. Aesthetic communication. Springer, 2010..

PESINA, S.; SOLONCHAK, T. The Sign in the Communication Process. In: International Science Conference: International Conference on Language and Technology (June 19-20). World Academy of Science, Engineering and Technology. International Science Index., 2014. p. 1021-1029.

RICKS, Derek M.; WING, Lorna. Language, communication, and the use of symbols in normal and autistic children. Journal of autism and childhood schizophrenia, 1975, 5.3: 191-221.

GORAWARA-BHAT, Rita; COOK, Mary Ann; SACHS, Greg A. Nonverbal communication in doctor–elderly patient transactions (NDEPT): Development of a tool. Patient education and counseling, 2007, 66.2: 223-234..

EKMAN, Paul; FRIESEN, Wallace V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. semiotica, 1969, 1.1: 49-98.

DAMHORST, Mary Lynn. Meanings of clothing cues in social context. Clothing and Textiles Research Journal, 1985, 3.2: 39-48..

DACY, Jennifer M.; BRODSKY, Stanley L. Effects of therapist attire and gender. Psychotherapy: Theory, Research, Practice, Training, 1992, 29.3: 486..

GREGORY, Stanford; WEBSTER, Stephen; HUANG, Gang. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. Language & Communication, 1993, 13.3: 195-217..

EKMAN, Paul, et al. Kinesic cues: The body, eyes, and face. 1999..

GREGERSEN, Tammy S. Nonverbal cues: Clues to the detection of foreign language anxiety. Foreign language annals, 2005, 38.3: 388-400.

HALL, Judith A.; HORGAN, Terrence G.; MURPHY, Nora A. Nonverbal communication. Annual review of psychology, 2019, 70.1: 271-294..

HANS, Anjali; HANS, Emmanuel. Kinesics, haptics, and proxemics: Aspects of non-verbal communication. IOSR Journal of Humanities and Social Science (IOSR-JHSS), 2015, 20.2: 47-52.

HARRISON, Randall P. Nonverbal communication. Human Communication As a Field of Study: Selected Contemporary Views, 1989, 113: 16..

SUMEISEY, Vivian Savenia; RANGKUTI, Rahmadsyah; GANIE, Rohani. Non-Verbal Communication of the Simpsons Memes in “Memes. Com” Instagram. Language Literacy: Journal of Linguistics, Literature, and Language Teaching, 2019, 3.1: 83-88..

HALL, Judith A.; HORGAN, Terrence G.; MURPHY, Nora A. Nonverbal communication. Annual review of psychology, 2019, 70.1: 271-294.

BREIL, Simon M., et al. 13 Contributions of Nonverbal Cues to the Accurate Judgment of Personality Traits. The Oxford handbook of accurate personality judgment, 2021, 195.

AL-HAMMADI, Dina; MOORE, Roger K. Using Sampling Techniques and Machine Learning Algorithms to Improve Big Five Personality Traits Recognition from Non-verbal Cues. In: 2021 National Computing Colleges Conference (NCCC). IEEE, 2021. p. 1-6.

ELLGRING, Johann Heinrich. The study of nonverbal behavior and its applications: State of the art in Europe. 1984.

ETHERINGTON, Cole, et al. Interprofessional communication in the operating room: a narrative review to advance research and practice..

RIESS, Helen; KRAFT-TODD, Gordon. EMPATHY: a tool to enhance nonverbal communication between clinicians and their patients. Academic Medicine, 2014, 89.8: 1108-1112..

TROTTA, Daniela; GUARASCI, Raffaele. How are gestures used by politicians? A multimodal co-gesture analysis. IJCoL. Italian Journal of Computational Linguistics, 2021, 7.7-1, 2: 45-66..

BETA, Annisa R.; NEYAZI, Taberez Ahmed. Celebrity Politicians, Digital Campaigns, and Performances of Political Legitimacy in Indonesia’s 2019 Elections. International Journal of Communication, 2022, 16: 331-355.

GRYSEN, Brian, et al. Social Competence and Language Ability in School-Age Children. 2012..

ONWUEGBUZIE, Anthony J.; ABRAMS, Sandra Schamroth. Nonverbal communication analysis as mixed analysis. In: The Routledge Reviewer's Guide to Mixed Methods Analysis. Routledge, 2021..

BAILEY, Britton. The importance of nonverbal communication in business and how professors at the University of North Georgia train students on the subject. 2018.

KIESNERE, Aisma Linda; BAUMGARTNER, Rupert J. Sustainability management emergence and integration on different management levels in smaller large-sized companies in Austria. Corporate Social Responsibility and Environmental Management, 2019, 26.6: 1607-16.

HAENLEIN, Michael, et al. Navigating the New Era of Influencer Marketing: How to be Successful on Instagram, TikTok, & Co. California management review, 2020, 63.1: 5-25.

FERGUSON-PATRICK, Kate. Cooperative learning in Swedish classrooms: Engagement and relationships as a focus for culturally diverse students. Education sciences, 2020, 10.11: 312..

HAENLEIN, Michael, et al. Navigating the New Era of Influencer Marketing: How to be Successful on Instagram, TikTok, & Co. California management review, 2020, 63.1: 5-25..

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. Psychological Bulletin, 84, 782–799..

HORVATH, Frank. Chicago: Birthplace of Modern Polygraphy. 2019..

RAJAN, Panthayil Babu. Polygraph Tests-Benefits and Challenges. Academicus, 2019, 2019.19: 146-155..

SCHIESS, Jacob. Potential Jurors' Perceptions of Polygraphs in Court. 2018..

YU, Runxin, et al. Using polygraph to detect passengers carrying illegal items. Frontiers in psychology, 2019, 10: 322..

ELLEGÅRD, Alvar. Darwin and the general reader: the reception of Darwin's theory of evolution in the British periodical press, 1859-1872. University of Chicago Press, 1990.

BUCY, Erik P. Nonverbal Cues. The International Encyclopedia of Media Effects, 2017, 1-11.

GOFFMAN, Erving. The moral career of the mental patient. Psychiatry, 1959, 22.2: 123-142.

MATHUR, Shivang, et al. Visual Analysis of Human-Object Interaction Detection: A Survey. arXiv preprint arXiv:2008.07231, 2020..

V. Kyrkou, N. Tsapatsoulis, “A Survey on Skin Lesion Analysis towards Melanoma Detection,” in ACM Computing Surveys, vol. 53, no. 6, pp. 1-39, Nov. 2020..

. Antanasijević, M. Đorđević, B. Milosevic, D. Milošević, and I. Orovic, "Sentiment Analysis in Healthcare: A Survey," in IEEE Access, vol. 8, pp. 225924-225945, 2020..

SHAHRIZAILA, Nortina; LEHMANN, Helmar C.; KUWABARA, Satoshi. Guillain-Barré syndrome. The lancet, 2021, 397.10280: 1214-1228..

CROUCH, Andrew E., et al. Ramsay Hunt Syndrome. In: StatPearls [Internet]. StatPearls Publishing, 2023..

GUPTA, Ankit. StrokeSave: a novel, high-performance mobile application for stroke diagnosis using deep learning and computer vision. arXiv preprint arXiv:1907.05358, 2019..

DONG, Wanxin, et al. Influence of different measurement methods of arterial input function on quantitative dynamic contrast-enhanced MRI parameters in head and neck cancer. Journal of Magnetic Resonance Imaging, 2023, 58.1: 122-132..

AL ZUBAIDI, Saba H.; ALSULTAN, Mustafa MH; HASAN, Lamiaa A. Stress effect on the mandibular dental arch by mentalis muscle over activity, finite element analysis. Journal of Orthodontic Science, 2023, 12.1: 61..

BLITZER, Andrea L.; PHELPS, Paul O. Facial spasms. Disease-a-Month, 2020, 66.10: 101041..

NILLES, Christelle, et al. Have We Forgotten What Tics Are? A Re-Exploration of Tic Phenomenology in Youth with Primary Tics. Movement Disorders Clinical Practice, 2023..

MAIQUEZ, Barbara Morera, et al. A double-blind, sham-controlled, trial of home-administered rhythmic 10-Hz median nerve stimulation for the reduction of tics, and suppression of the urge-to-tic, in individuals with Tourette syndrome and chronic tics, , and suppression of the urge-to-tic, in individuals with Tourette syndrome and chronic tic disorder. Journal of Neuropsychology, 2023..

BURKOV, Andriy. The hundred-page machine learning book. Quebec City, QC, Canada: Andriy Burkov, 2019..

WANG, Fengyuan, et al. Facial expression recognition from image based on hybrid features understanding. Journal of Visual Communication and Image Representation, 2019, 59: 84-88..

CHOY, Garry, et al. Current applications and future impact of machine learning in radiology. Radiology, 2018, 288.2: 318..

YIN, Mengping, et al. A new approximate image verification mechanism in cloud computing. International Journal of Embedded Systems, 2019, 11.6: 687-697.

WANG, Pin; FAN, En; WANG, Peng. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recognition Letters, 2021, 141: 61-67..

SARWINDA, Devvi, et al. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Computer Science, 2021, 179: 423-431..

WANG, Pin; FAN, En; WANG, Peng. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recognition Letters, 2021, 141: 61-67.

FUNABASHI, Satoshi, et al. Morphology-specific convolutional neural networks for tactile object recognition with a multi-fingered hand. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019. p. 57-63..

DAI, Xiyang, et al. Dynamic head: Unifying object detection heads with attentions. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 7373-7382..

PANG, Ziqi, et al. Standing Between Past and Future: Spatio-Temporal Modeling for Multi-Camera 3D Multi-Object Tracking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. p. 17928-17938..

ZHANG, Jiarui, et al. HALF: A High-performance Automated Large-scale Land Cover Mapping Framework. 2023..

WANG, Jiamei; HU, Xiangdong. Factors Influencing Disease Prevention and Control Behaviours of Hog Farmers. Animals, 2023, 13.5: 787..

KHABARLAK, Kostiantyn; KORIASHKINA, Larysa. Fast facial landmark detection and applications: A survey. arXiv preprint arXiv:2101.10808, 2021..

ARIPIRALA, Gowtham Venkata Sai; RAMALINGAM, Puviarasi. Diabetes mellitus (DM) detection using decision tree and naïve bayes algorithm for accuracy, specificity and sensitivity improvement. In: AIP Conference Proceedings. AIP Publishing, 2023..

KAREEM, Shahab Wahhab; OKUR, Mehmet Cudi. Pigeon inspired optimization of bayesian network structure learning and a comparative evaluation. Journal of Cognitive Science, 2019, 20.4: 535-552..

ZHANG, Yufei, et al. Body Knowledge and Uncertainty Modeling for Monocular 3D Human Body Reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. p. 9020-9032..

JOO, Jungseock; BUCY, Erik P.; SEIDEL, Claudia. Automated Coding of Televised Leader Displays: Detecting Nonverbal Political Behavior with Computer Vision and Deep Learning. International Journal of Communication (19328036), 2019..

M. Ngxande, J. Tapamo and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics, (PRASA-RobMech), 2017, pp. 156-161, doi: 10.1109/RoboMech.2017.8261140.

Magán, E., Sesmero, M. P., Alonso-Weber, J. M., & Sanchis, A. (2022). Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images. Applied Sciences, 12(3), 1145..

SHEIKH, Ali Ayub; MIR, Junaid. Machine learning inspired vision-based drowsiness detection using eye and body motion features. In: 2021 13th International Conference on Information & Communication Technology and System (ICTS). IEEE, 2021. p. 146-150..

GRATCH, Jonathan. The promise and peril of interactive embodied agents for studying non-verbal communication: a machine learning perspective. Philosophical Transactions of the Royal Society B, 2023, 378.1875: 20210475..

REDDY, Tharun Kumar; GUPTA, Vinay; BEHERA, Laxmidhar. Autoencoding convolutional representations for real-time eye-gaze detection. In: Computational Intelligence: Theories, Applications and Future Directions-Volume II. Springer, Singapore, 2019. p. 229-23.

AHN, Hoyeon. Non-contact Real time Eye Gaze Mapping System Based on Deep Convolutional Neural Network. arXiv preprint arXiv:2009.04645, 2020..

SAHA, Dipayan, et al. Deep learning-based eye gaze controlled robotic car. In: 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, 2018. p. 1-6..

NAQVI, Rizwan Ali, et al. Deep learning-based gaze detection system for automobile drivers using a NIR camera sensor. Sensors, 2018, 18.2: 456..

SUNDARAM, R. Meenakshi; DHARA, Bibhas Chandra. Neural network based iris recognition system using Haralick features. In: 2011 3rd International Conference on Electronics Computer Technology. IEEE, 2011. p. 19-23..

TALLAPRAGADA, V. V. S.; RAJAN, E. G. Improved kernel-based IRIS recognition system in the framework of support vector machine and hidden Markov model. IET image processing, 2012, 6.6: 661-667..

X. Chen and A. L. Yuille, Articulated pose estimation by a graphical model with image dependent pairwise relations, in Advances in Neural Information Processing Systems, 2014, pp. 1736–1744..

BELAGIANNIS, Vasileios; ZISSIS, Dimitrios; PEZZIMENTI, Fortunato. A 3D vision-based classifier for head pose estimation of a humanoid robot. Procedia Computer Science, 2018, 130: 1022-1028..

DE LA TORRE, Fernando, et al. Robust 3D face pose estimation from single images or video sequences. International Journal of Computer Vision, 2017, 124.2: 144-167..

DE LA TORRE, Fernando, et al. Robust 3D face pose estimation from single images or video sequences. International Journal of Computer Vision, 2017, 124.2: 144-167.

IANĂŞ, Dacian, et al. Facial landmark detection and head pose estimation for human-robot interaction. In: 2017 16th RoEduNet Conference: Networking in Education and Research. IEEE, 2017. p. 1-6.

DE MELO VENTOLA, Pamela; EISENBERG, Robert; BARRETO, Joao P. Estimating head pose from gaze: A pilot study. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). IEEE, 2018. p. 212-217.

GUO, Qinghua, et al. Enhanced camera-based individual pig detection and tracking for smart pig farms. Computers and Electronics in Agriculture, 2023, 211: 108009.

MARTINELLI, Agostino, et al. Looking at the robot: Visual recognition of fixations and transitions. IEEE Robotics and Automation Letters, 2018, 3.4: 4255-4262.

KAVANA, K. M.; SUMA, N. R. Recognization of hand gestures using mediapipe hands. International Research Journal of Modernization in Engineering Technology and Science, 2022, 4.06.

GOMEZ, Luis F., et al. Exploring facial expressions and action unit domains for Parkinson detection. Plos one, 2023, 18.2: e0281248..

RECALDE, Melchizedek, et al. Creating an Accessible Future: Developing a Sign Language to Speech Translation Mobile Application with MediaPipe Hands Technology. In: 2023 10th International Conference on ICT for Smart Society (ICISS). IEEE, 2023. p. 1-6.

RIEHLE, Dirk; GROSS, Thomas. Role model based framework design and integration. In: Proceedings of the 13th ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications. 1998. p. 117-133.

KIM, Jong-Wook, et al. Human pose estimation using mediapipe pose and optimization method based on a humanoid model. Applied sciences, 2023, 13.4: 2700.

RATHOD, Siddharajsinh, et al. RealD3: A Real-time Driver Drowsiness Detection Scheme Using Machine Learning. In: 2023 IEEE Wireless Antenna and Microwave Symposium (WAMS). IEEE, 2023. p. 1-5..

CARVALHO, Davi R., et al. Head tracker using webcam for auralization. In: Proceedings of the Inter-Noise 2021 Congress, Washington, WA, USA. 2021. p. 1-5.

LIN, Yiqiao; JIAO, Xueyan; ZHAO, Lei. Detection of 3d human posture based on improved mediapipe. Journal of Computer and Communications, 2023, 11.2: 102-121..

DUNNHOFER, Matteo, et al. Visual object tracking in first person vision. International Journal of Computer Vision, 2023, 131.1: 259-283.

LORENZ, Oliver; THOMAS, Ulrike. Real Time Eye Gaze Tracking System using CNN-based Facial Features for Human Attention Measurement. In: VISIGRAPP (5: VISAPP). 2019. p. 598-606..

BRIGHT, Abira, et al. Real-time eye tracking for handheld optical coherence tomography system. In: Women in Optics and Photonics in India 2022. SPIE, 2023. p. 143-147.

Zhu, X., Lei, Z., Liu, X., Shi, H., He, Z., & Li, S. (2016). Face Alignment Across Large Poses: A 3D Solution. Computer Vision - ECCV 2016 Workshops, Springer, 1-16.

http://www.ifp.illinois.edu/~vuongle2/helen/.

X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild, " 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012, pp. 2879-2886, doi: 10.1109/CVPR.2012.6248014.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge. IEEE International Conference on Computer Vision Workshops (ICCVW), 397-403.

Nikos Petrellis, Nikolaos Voros, Christos Antonopoulos, Georgios Keramidas, Panagiotis Christakos, Panagiotis Mousouliotis, July 24, 2022, "NITYMED", IEEE Dataport, doi: https://dx.doi.org/10.21227/85xe-3f88.

DUNNHOFER, Matteo, et al. Visual object tracking in first person vision. International Journal of Computer Vision, 2023, 131.1: 259-283., 2023.

TU, Ching-Ting; LIN, Hwei-Jen; CHANG, Chin-Yu. Lighting-and Personal Characteristic-Aware Markov Random Field Model for Facial Image Relighting System. IEEE Access, 2022, 10: 20432-20444.

GRATCH, Jonathan. The promise and peril of interactive embodied agents for studying non-verbal communication: a machine learning perspective. Philosophical Transactions of the Royal Society B, 2023, 378.1875: 20210475..

GENG, Fudi, et al. Light-efficient channel attention in convolutional neural networks for tic recognition in the children with tic disorders. Frontiers in Computational Neuroscience, 2022, 16: 1047954..

E LA TORRE, Fernando, et al. Robust 3D face pose estimation from single images or video sequences. International Journal of Computer Vision, 2017, 124.2: 144-167..

RODRÍGUEZ MARTÍNEZ, Eder A., et al. DeepSmile: Anomaly Detection Software for Facial Movement Assessment. Diagnostics, 2023, 13.2: 254..

DE LA TORRE, Fernando, et al. Robust 3D face pose estimation from single images or video sequences. International Journal of Computer Vision, 2017, 124.2: 144-167..

DE MELO VENTOLA, Pamela; EISENBERG, Robert; BARRETO, Joao P. Estimating head pose from gaze: A pilot study. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). IEEE, 2018. p. 212-217.

BELAGIANNIS, Vasileios; ZISSIS, Dimitrios; PEZZIMENTI, Fortunato. A 3D vision-based classifier for head pose estimation of a humanoid robot. Procedia Computer Science, 2018, 130: 1022-1028.

الملخص

يعتبر التواصل البشري الفعال بين الناس موضوعا مهما ويلعب دورا أساسيا في الكثير من المجالات مثل الرعاية الصحية والسياسة والقانون والأعمال والتعليم وعلم النفس. تعد الإشارات الغير لفظية عنصرا أساسيا في التفاعلات بين الناس فهي تعطي انطباعات ووصفا عن كل حالة للمشاعر التي يشعر بها الشخص مثل (السعادة، الغضب، الخوف، الحزن، الاشمئزاز، المفاجأة، والازدراء) والحالات الصحية مثل (الضعف العصبي، والاضطرابات، والسكتات الدماغية). ومع ذلك، الحركات اللاإرادية غير المنتظمة مثل تشنجات الوجه تعتبر تحديا في التعرف عليها فهي تحدث في فترات قصيرة، وعدم ملاحظتها يؤدي في كثير من الأحيان إلى فقدان المعلومات.

ثقدم هذه الدراسة إطار عمل مبتكر لاكتشاف مثل هذه الايماءات والتعابير غير اللفظية، بما في ذلك وضع الرأس، حركة الجفون، موقع القزحية، النعاس، وانحراف الفم والقدرة على ضم الشفتين للنفخ. يعمل النظام المبني ويعطي نتائج فورية أي في الوقت الفعلي ويمكن للمستخدم تزويده بالبيانات من خلال كاميرات جهاز الكمبيوتر المحمول ذات الكاميرا الأمامية، ومقاطع الفيديو، والصور المُحمّلة. تتضمن الطريقة تهيئة نموذج نقاط الارتكاز على الوجه من "MediaPipe"، مع معالجة اللقطات أو الصور باستخدام مكتبة "OpenCV" واستخراج نقاط الارتكاز لتحديد نقاط الوجه، ويتم التقاط الحركات والايماءات في ملامح الوجه من خلال تحليل المسافات والزوايا ومدة الثبات على الحركة والحدود المعرفة مسبقا.

تم انشاء واجهة مستخدم ويب سهلة الاستخدام للمستخدمين وتمكنهم من التقاط وملاحظة الحركات البسيطة في الوجه بطريقة سهلة وفعالة.

تم تقييم النظام بدقة وفحص أداؤه باستخدام صور متنوعة تم تجميعها من صور موجودة ومقاطع فيديو. بعد المعالجة المسبقة والمقارنة مع نقاط الارتكاز الصحيحة، تم حساب مقاييس التقييم مثل (الدقة، الدقة المطلقة، الاستدعاء، وقيمةF1). أظهرت النتائج دقة عالية: 100% لوضع الرأس، 86% لحالة الجفون، 96% لالتقاط التثاؤب، 88% لانحراف الفم، 97% لالتقاط النعاس، و100% لحركة الفم.

في الختام، يمكن العمل المقترح إمكانية تعزيز التواصل البشري، ويسهل التقاط الايماءات التي تسهل الحصول على الكثير من المعلومات من خلال فك تشفير التصرفات غير اللفظية المعقدة والحصول

على نتائج في الوقت الفعلي والدقة العالية وهذا يؤكد فعاليته للاستخدام في مجموعة متنوعة من المجالات.