



**Arab American University**  
**Faculty of Graduate Studies**

**CLINICAL CHARACTERISTICS OF HOSPITALIZED  
COVID -19 INFECTED PATIENTS IN PALESTINE  
(Descriptive Analytics and Data Mining Approach)**

By

**Ibrahim M.M. Sabbah**

Supervisor

**Prof. Dr. Mohammed Awad**

Co- supervisor

**Dr. Shahenaz Najjar**

**This thesis will submit in partial fulfillment of the  
requirements for Master's degree in Health informatics**

**© Arab American University – 2020 All rights reserved**

I

**CLINICAL CHARACTERISTICS OF HOSPITALIZED COVID -19  
INFECTED PATIENTS IN PALESTINE  
(Descriptive Analytics and Data Mining Approach)**

By  
Ibrahim M.M. Sabbah

Committee Members

Signature

1. Supervisor: Prof. Mohammed Awad

.....  


2. Co-Supervisor: Dr. Shahenaz Najjar

.....  


3. Internal Examiner: Dr. Yousef Mimi

.....  


4. External Examiner: Dr. Amira Shaheen

.....  


## الإهداء

الى حضرة رسول الله وقدوتنا ... خير معلم واول باحث و قارئ... سيدنا محمد صل

الله عليه وسلم

الى قدوتي، ومثلي الأعلى في الحياة، ومثال التقاني والعطاء ،,,, أبي العطوف

الى من لا أجد كلمات يمكن أن تمنحها حقها، ملحمة الحب وفرحة العمر ،,,, أمي

الحبيبة

الى سندي وعضدي ومشاطري أفراحي وأحزاني ،,,, أخوتي وأخواتي

الى رفيقة الدرب أسمى رموز الإخلاص والوفاء ،,,, زوجتي

الى روعي وقُرّة عيني ونبض فؤادي ،,,, أبنائي

الى من سجلنا المثابرة والنجاح بفضلهم ،,,, أساتذتي

الى من عرفتمكم في الشدائد وكنتم خير من سار معي،,,, أصدقائي وزملائي

إليكم جميعاً أهدي هذا العمل المتواضع ...

## **Abstract**

**Background:** Palestine is one of the countries that has been affected by COVID-19 pandemic. This research study aims to understand and improve the clinical and diagnostic knowledge of COVID-19 patients in Palestine using conventional statistical analysis and data mining tools to develop model can classify and predict patient future health situation.

**Method:** Quantitative research design using descriptive analysis was carried out to obtain an insight into 132 hospitalized patients. A retrospective review of patient's medical records was conducted. Data was collected from two designated hospitals in West Bank of Palestine for a period of 19 March to 20 July 2020. The clinical data includes follow up laboratory tests, clinical observation, and treatments plan during their first two days of stay in the hospital.

**Results:** The most common symptoms on admission were cough (51.5%), fever (41.7%), and shortness of breath (25.8%). Numerous differences were reported between severe and not severe cases, including higher White blood cells (WBC), neutrophil, Lactate dehydrogenase (LDH), ferritin, BUN, and creatine ( $P < .05$ ), and lower lymphocytes percentage and SPO<sub>2</sub> ( $P < .001$ ). CRP, Ferritin level, and Monocytes tests were statistically significant in their association with severity level ( $P$ -value  $< 0.05$ ). Having diabetes, hypertension and Cerebrovascular Disease is significantly associated with the severity level of COVID-19. The most common treatments given for COVID-19 patients were Antipyretic (75.0%), Supplementary drugs (55.3%), Anticoagulant (47.0%), and Antibiotic (46.2%). Four popular artificial intelligent tools were used to develop a prediction model for COVID-19 severity level; Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Networks (ANNs) with accuracy (71%, 71.5%, 71.5%, and 74%) respectively.

**Conclusion:** This study was conducted in two phases; First: a comprehensive descriptive analysis of the patient's demographic, comorbidities, complaints, and laboratory findings, examined the relationship between each variable with patient's health status and the severity of the symptoms experienced (severe, non-severe), This phase showed that the most common symptoms on admission were cough (51.5%), fever (41.7%), and shortness of breath (25.8%). Numerous differences were reported between severe and non-severe cases, including higher WBC, neutrophil, LDH, ferritin, BUN, and creatinine ( $P < .05$ ), and lower lymphocytes percentage and SPO<sub>2</sub> ( $P < .001$ ). Relationships between CRP and severity level on the one hand and Ferritin level, and Monocytes on the other were statistically significant ( $P$ -value  $< 0.05$ ). Having diabetes, hypertension and Cerebrovascular Disease was significantly associated with the higher severity level of COVID-19 ( $P < .001$ ). Second: The disease severity data from the first phase with artificial intelligence (AI) algorithms were used to develop an algorithm that can predict the patients' future condition. Four popular machine learning tools were used; Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Networks (ANNs). Models' accuracy levels were; 71%, 71.5%, 71.5%, and 74%, respectively, all considered high accuracy levels the models registered in spite of the limited available data used in models testing.

AI can be very effective to describe behavior of COVID-19 and inform clinical decisions and treatment plans,

## Table of Contents

|  |     |
|--|-----|
| <b>Abstract</b> .....  | III |
| <b>Table of Contents</b> .....                                 | V   |
| <b>List of Figures</b> .....                                   | IX  |
| <b>List of Tables</b> .....                                    | X   |
| <b>List of appendices</b> .....                                | XI  |
| <b>List of Acronyms and Abbreviations</b> .....                | XII |
| <b>Acknowledgments</b> .....                                   | XIV |
| <b>Statement of Original Authorship</b> .....                  | XV  |
| <b>1. Chapter 1: Introduction</b> .....                        | 1   |
| <b>1.1. Background</b> .....                                   | 1   |
| <b>1.2. Problem statement</b> .....                            | 2   |
| <b>1.3. Study objectives</b> .....                             | 3   |
| 1.3.1. Study specific objectives.....                          | 4   |
| 1.3.2. Secondary objectives .....                              | 4   |
| <b>1.4. Study justification</b> .....                          | 4   |
| <b>1.5. Study outcome</b> .....                                | 5   |
| <b>1.6. Thesis chapter outlines</b> .....                      | 5   |
| <b>1.7. Summary</b> .....                                      | 6   |
| <b>2. Chapter 2: Literature Review</b> .....                   | 7   |
| <b>2.1. Introduction</b> .....                                 | 7   |
| <b>2.2. Diagnosis</b> .....                                    | 8   |
| <b>2.3. Prevalence of COVID 19</b> .....                       | 9   |
| 2.3.1. Global Situation .....                                  | 9   |
| 2.3.2. Palestinian situation .....                             | 10  |
| <b>2.4. Infection transmission and incubation period</b> ..... | 11  |
| <b>2.5. Symptoms</b> .....                                     | 12  |
| 2.5.1. Symptom severity .....                                  | 13  |

|   |    |
|---|----|
| <b>2.6. Clinical finding</b> .....                              | 14 |
| <b>2.7. Laboratory finding</b> .....                            | 15 |
| <b>2.8. Radiological finding</b> .....                          | 16 |
| <b>2.9. Co-morbidities</b> .....                                | 16 |
| 2.9.1. Hypertension .....                                       | 16 |
| 2.9.2. Diabetes Mellitus .....                                  | 17 |
| 2.9.3. Malignancy .....   | 18 |
| 2.9.4. Cerebrovascular diseases .....                           | 18 |
| 2.9.5. Chronic kidney disease .....                             | 18 |
| 2.9.6. Chronic liver disease.....                               | 19 |
| 2.9.7. Neurological diseases.....                               | 19 |
| 2.9.8. Cardiovascular Disease (CVD).....                        | 20 |
| 2.9.9. Respiratory disease .....                                | 20 |
| <b>2.10. Treatment of COVID-19</b> .....                        | 20 |
| <b>2.11. Data mining and machine learning in COVID-19</b> ..... | 22 |
| 2.11.1. Data mining techniques.....                             | 23 |
| 2.11.1.1 Logistic regression (LR).....                          | 23 |
| 2.11.1.2 Decision tree (DT).....                                | 23 |
| 2.11.1.3 Support vector machine (SVM) .....                     | 24 |
| 2.11.1.4 Artificial neural networks (ANN's).....                | 24 |
| 2.11.2. Model evaluation .....                                  | 24 |
| 2.11.3. Relevant studies .....                                  | 26 |
| <b>2. Chapter 3: Conceptual framework for COVID-19</b> .....    | 28 |
| <b>2.1. Introduction</b> .....                                  | 28 |
| <b>2.2. Variables</b> .....                                     | 29 |
| 2.2.1. Demographic variables .....                              | 29 |
| 2.2.2. Medical history .....                                    | 29 |
| 2.2.3. Laboratory tests.....                                    | 29 |
| 2.2.4. Complaining.....   | 30 |
| <b>2.3. WHO criteria for severity level staging</b> .....       | 30 |
| <b>3. Chapter 4: Study Methodology</b> .....                    | 33 |
| <b>3.1. Introduction</b> .....                                  | 33 |

|   |    |
|---|----|
| <b>3.2. Study settings and population characteristics</b> .....   | 33 |
| 3.2.1. Study population .....                                     | 33 |
| 3.2.2. Study setting and participants .....                       | 34 |
| <b>3.3. Study design:</b> .....                                   | 34 |
| <b>3.4. Study sampling and sampling techniques</b> .....          | 34 |
| <b>3.5. Data collection</b> .....                                 | 35 |
| 3.5.1. Inclusion criteria .....                                   | 35 |
| 3.5.2. Exclusion criteria .....                                   | 36 |
| <b>3.6. Study tools</b> .....                                     | 36 |
| <b>3.7. Study variables</b> .....                                 | 37 |
| <b>3.8. Data analysis</b> .....                                   | 38 |
| <b>3.9. Work pathway</b> .....                                    | 39 |
| <b>3.10. Pathway model</b> .....                                  | 40 |
| <b>3.11. Ethical consideration</b> .....                          | 41 |
| <b>3.12. Summary</b> .....  | 42 |
| <b>4. Chapter 5: Data analysis and research findings</b> .....    | 43 |
| <b>4.1. Introduction</b> .....                                    | 43 |
| <b>4.2. Data Preparation and cleaning</b> .....                   | 44 |
| <b>4.3. Socio-demographic characteristics of the sample</b> ..... | 44 |
| <b>4.4. Co-morbidities in the study sample</b> .....              | 48 |
| <b>4.6. Laboratory findings</b> .....                             | 53 |
| <b>4.7. Treatments provided to COVID-19 patients</b> .....        | 58 |
| <b>4.8. Prediction models</b> .....                               | 60 |
| <b>4.9. Linear regression</b> .....                               | 60 |
| <b>4.10. Data mining techniques</b> .....                         | 64 |
| 4.10.1. Logistic regression (LR) .....                            | 66 |
| 4.10.2. Decision tree (DT) .....                                  | 67 |
| 4.10.3. Support vector machine (SVM) .....                        | 68 |
| 4.10.4. Artificial neural networks (ANNs) .....                   | 70 |
| 4.10.5. Model comparison .....                                    | 71 |
| <b>4.11. Summary</b> .....  | 73 |
| <b>5. Chapter 6: Conclusion and Recommendation</b> .....          | 74 |

|   |     |
|---|-----|
| 5.3. Recommendations and further research ..... | 76  |
| 2.1. Study limitation .....                     | 77  |
| 2.2. Strength of the study .....                | 78  |
| <b>Bibliography</b> .....                       | 79  |
| <b>Appendices</b> .....                         | 99  |
| ملخص الدراسة.....                               | 101 |

## List of figures

|   |    |
|---|----|
| Figure 1: COVID-19 cases according WHO regions. ....              | 10 |
| Figure 2: COVID-19 cases in Palestinian governorates .....        | 11 |
| Figure 3: COVID-19 transmission (Tahamtan & Ardebili, 2020).....  | 12 |
| Figure 4: Represent the research pathway model .....              | 41 |
| Figure 5: Socio-demographic Characteristics of the Sample.....    | 45 |
| Figure 6: Sample Co-morbidities.....                              | 48 |
| Figure 7: severity level compareson according Co-morbidities..... | 51 |
| Figure 8: complaining according to severity level.....            | 53 |
| Figure 9:Treatments provided to COVID-19 patients .....           | 58 |
| Figure 10: linear regression correlating Age with Spo2 .....      | 61 |
| Figure 11:: Linear regression correlating BUN with SPO2.....      | 62 |
| Figure 12: Linear regression correlating CRP with SPO2.....       | 62 |
| Figure 13: Linear regression correlating LDH with SPO2 .....      | 63 |
| Figure 14: Linear regression correlating Ferritin with SPO2 ..... | 63 |
| Figure 15: Logistic regression Area Under Curve .....             | 66 |
| Figure 16: Logistic regression Confusion Matrix .....             | 67 |
| Figure 17: Decision Tree Area Under Curve .....                   | 68 |
| Figure 18: Decision Tree Confusion Matrix .....                   | 68 |
| Figure 19:Support victor machine Area Under Curve .....           | 69 |
| Figure 20: Support victor machine Confusion Matrix .....          | 70 |
| Figure 21: Artificial neural network Area Under Curve.....        | 71 |
| Figure 22: Artificial neural network Confusion Matrix.....        | 71 |
| Figure 23: model evaaluation compareson .....                     | 72 |

**List of tables**

|  |    |
|--|----|
| Table 1: summary of the effects of the drug interventions on the outcomes .....              | 22 |
| Table 2: COVID-19 symptoms classification.....   | 31 |
| Table 3: Data collection categories.....   | 36 |
| Table 4: Study Population Characteristics .....  | 46 |
| Table 5: demographic Characteristics correlation with a severity level .....                 | 47 |
| Table 6:Co-morbidities correlation with a severity level .....                               | 49 |
| Table 7: patients complying relationship with a severity level.....                          | 52 |
| Table 8: Lab Tests Findings of Patients Infected with COVID-19 on Admission to Hospital .... | 53 |
| Table 9: Diabetes relationships with Ferritin and Severity.....                              | 55 |
| Table 10: features in utilizing prediction model .....                                       | 64 |
| Table 11: Data mining model evaluation.....  | 72 |

**List of appendices**

Appendices 1: Data collection form ..... 99

Appendices 2: Ethical Committee of the Palestinian Ministry of Health Approval..... 100

**List of acronyms and abbreviations**

|          |                                       |
|----------|---------------------------------------|
| AAUP     | Arab American University- Palestine   |
| ACC      | American College of Cardiology        |
| AHA      | American Heart Association            |
| ANN      | Artificial Neural Network             |
| AUC      | Area Under the Curve                  |
| BP       | Blood Pressure                        |
| CA       | Malignancy                            |
| CI       | Confidence Interval                   |
| CKD      | Chronic Kidney Disease                |
| COPD     | Chronic Obstructive Pulmonary Disease |
| COVID-19 | Coronavirus Disease 2019              |
| CRP      | C-Reactive Protein                    |
| CT       | Computed Tomography                   |
| CVD      | Cardiovascular Disease                |
| DM       | Diabetes Mellitus                     |
| DT       | Decision Tree                         |
| FN       | False Negative                        |
| FP       | False-Positive                        |
| GLMs     | Generalized Linear Model              |
| HGB      | Hemoglobin Levels                     |
| HTN      | Hypertension                          |
| IGG      | Immunoglobulin G                      |
| IGM      | Immunoglobulin M                      |
| LDH      | Lactate Dehydrogenase                 |
| LR       | Logistic Regression                   |
| ML       | Machine Learning                      |
| MOH      | Ministry of Health                    |
| NP       | Nasopharyngeal                        |
| OP       | Oropharyngeal                         |

|        |                                  |
|--------|----------------------------------|
| PPIs   | Proton Pump Inhibitors           |
| RA     | Rheumatoid                       |
| RT-PCR | Real-Time Reverse Transcription  |
| SA     | Statistical Analysis             |
| SOB    | Shortness of Breath              |
| SVM    | Support Vector Machine           |
| TESSy  | The European Surveillance System |
| TN     | True Negative                    |
| TP     | True Positive                    |
| WB     | West Bank                        |
| WBCs   | White Blood Cells                |
| WHO    | World Health Organization        |

## **Acknowledgments**

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my master's degree!

The Health Informatics team at Arab American University, especially to my supervisor Prof. Dr Mohammed Awad, and Dr Shahenaz Najjar, whom insight and knowledge into the subject matter steered me through this research.

And my biggest thanks to my family (Father, Mother, brothers, sisters) for all the support you have shown me through this research, the culmination of three years of distance learning. (I love You)

For my kids, sorry for being even grumpier than normal whilst I wrote this thesis! And for my wife Sabrin, thanks for all your support, without which I would have stopped these studies a long time ago, You have been amazing.

And special thanks to my colleagues at Palestinian Ministry of Health, whom allowed my studies to go the extra mile. Mghanamin & Basim (You have been the best supporters)

My colleagues at CasaMed Medical Supplier Co., who have supported me and had to put up with my stresses and moans for the past three years of study! (sorry for all the extra work Hazem!).

### Statement of original authorship

I certify that this thesis content has not been submitted to fulfill the requirement for academic degree, it also includes no previously published material except where due reference is made.

Student Name: Ibrahim M.M. Sabbah

Signature:



30 / 8 / 2021

## **1. Chapter 1: Introduction**

---

### **1.1. Background**

Coronavirus COVID-19 became the first human tragedy nowadays affecting all people of the world with an increasing number of deaths every day. COVID-19 has a great impact not only on the population health but also on whole life aspects economically and socially (Response & Lives, 2020).

Globally, more than 121 million confirmed cases of COVID-19 were reported, with more than a 2.5 million deaths in approximately 219 Countries, areas, or territories (Cucinotta & Vanelli, 2020). World Health Organization (WHO) sensed the gravity of the situation and declared the COVID-19 outbreak as a pandemic in March11-2020, whereas the first diagnosed case in Palestine; seven confirmed cases of COVID-19 had been detected at a hotel in the middle of Bethlehem city on 5-March after a group of tourists had visited the city and stay at that hotel in late February (WHO, 2020) (Zachariah et al., 2020). On 6-March the Palestinian Government declared a state of emergency in all countries and closed mosques, churches, and educational institutions (universities and schools), prohibiting of gathering in large numbers in markets, streets, and public places. The number of cases continues to rise slowly to reach 422 confirmed cases on 23 May 2020 (WHO, 2020) (Response & Lives, 2020). After that, the Palestinian government has eased the lockdown in the country by keeping the closure of educational institutions and maintaining public safety procedures. The smooth rising in the number of new confirmed cases is continue till 19 June 2020, which is what the Palestinian Ministry of Health (MOH) thought to be

the second wave of covid-19 strikes in the country, while a dramatic increase of cases starts to reported in Hebron to reach more than 3000 confirmed cases and 8 deaths within two weeks (AlKhalidi et al., 2020). Several emerging hot spots of a pandemic were determined across the country with the rapid spread of disease to reach about 9744 confirmed cases and 67 deaths by 24 July 2020.

Many studies showed that most cases develop symptoms during 4-6 days after receiving the infection. The virus may remain dormant for a period that may exceed 14 days so that the average incubation period for the covid-19 virus is approximately 5 days. Relevant studies of hospitalized cases have shown that the most common symptoms associated with COVID-19 are cough, fever, fatigue, and myalgia (Cao et al., 2020) (L. Zhang et al., 2020).

This study will depend on two stages, the first stage is to determine possible correlations between the main category's variables (Patient demographics, clinical finding, Laboratory tests, Co-morbidity, patient's health status, and severity level by using the SPSS as statistical analysis tool. The second stage depends on using machine learning (ML) techniques to classify the infected health status if it Severe or Non-Severe.

## **1.2. Problem statement**

The Coronavirus Disease known as (COVID-19) pandemic continue to have a mortifying impact on the well-being and health of the global population. A persistent increase in the number of patients testing positive for COVID-19 has generated a lot of stress on government across the globe finding it complicated to tackle the situation. Dramatic increase of COVID-19 cases continues to

reported in Palestine to reach more than 9744 confirmed cases and 67 deaths by 24-Jul-2020. (WHO, 2020).

To our knowledge, limited studies have been made in Palestine to evaluate and describe the current situation of COVID-19, in term of describing its characteristics and link it with the severity level of hospitalized patients. Also, the Palestinian healthcare system frequently can't be dealing with increasing numbers of COVID-19 patients. However, conducting an analytics study that deals with most aspects of this disease and relates it to the patient's condition and symptoms severity, using machine learning, will support in preparing the infrastructure for a strong and integrated health system needs and provide a scientific evidence to understand the development, characteristics, conditions, and symptoms of COVID-19 patients. In addition, there is, up to now, no research on the utilize of machine learning in Palestine to determine the symptoms severity level of COVID-19 patients in the future, which would lead decision-makers to understand and realize the needs of the healthcare sectors and resources allocation.

### **1.3. Study objectives**

The main objective of this study is to evaluate and describe the current situation of COVID-19 in Palestine, in terms of describing its characteristics and link it with the severity level of hospitalized patients depends on statistical analysis and machine learning techniques.

### **1.3.1. Study specific objectives**

The study specific objectives are:

1. To determine the association between the epidemiological variables, laboratory tests, comorbidities, and clinical features with the severity level of the disease.
2. Using statistical analysis and machine learning techniques to laying down the practical foundations for developing a model capable of classifying the severity level of the disease.

### **1.3.2. Secondary objectives**

Provide a prompt understanding of the disease behavior that will enable policymakers to real-time actions and providing better planning, severity detection, complication prevention, and resource allocations in future similar situations.

## **1.4. Study justification**

Coronavirus COVID-19 became the first human tragedy nowadays affecting all people globally with an increasing number of deaths every day, to our knowledge, this study is the first of its kind at the national level, as it contributes significantly to develop a scientific theory to confront this epidemic, due to its role in understanding all disease aspects.

Conducting analytics research using computational and data mining techniques will help to provide a panoramic insight into the disease datasets pulled from Palestinian MOH, driving to

better understanding the disease behavior, risk factors, limitations, which will contribute with policy extension, as it supports policymakers to effectively tackle the epidemic situation, better resource allocations and prepare control plans and needs of the healthcare system to save lives. Techniques like descriptive analytics and machine learning can help front line medical staff to analyze a massive amount of data in a short period and discover the most influential features for providing medical decisions regarding the patient's health situation.

### **1.5. Study outcome**

The study results will be laying down the practical and scientific foundations for developing a model capable to develop a prediction model for categorizing groups of patients that undergoing the risk of developing severe symptoms in the future.

Using data mining technique and exploring the risk factor affecting hospitalized patient's outcome depending on descriptive data analysis and results visualization of correlation between laboratory tests, comorbidities and clinical features with severity level of the disease.

### **1.6. Thesis chapter outlines**

In this chapter of the thesis, we illustrate the study background, study justification, problem statement, study aim and objectives, and study outcomes for Coronavirus clinical characteristics.

The rest of the thesis is organized as follows. **Chapter two** includes outlines of the scientific work from previous contributors to the field. **Chapter three** is used to explain our research methodology and provides a detailed account of our experiment, includes the study area and setting, study

subjects, study design overview, sample size, data collection methods and measurements, data management, analysis plan, and ethical consideration, then **chapter four** provide the study conceptual framework and pathway followed to achieve the study goals. **Chapter five** of the thesis consists of the data analysis and research findings, includes actual results with findings and a discussion of the results, as we presented the results of the analysis and discuss them step-by-step in the same section. **Chapter six** provides a general conclusion that describes the results of the studies presented in this thesis. Also, the strengths and limitations of this thesis are considered, and recommendations are given for further research.

## **1.7. Summary**

In this chapter of the study, we start by talking about the main terms used in the study to give the reader an overview of the study, where we have defined and clarified some terms like Coronavirus, clinical characteristics of COVID-19, and its symptoms. Then we presented the significance of the study and problem statement. Also, the study objectives were covered, and the content of each chapter of the study has been explained.

## 2. Chapter 2: Literature review

---

### 2.1. Introduction

Coronaviruses (CoV) are a big virus's family that causes various illnesses like SARS-CoV (severe acute respiratory syndrome) and MERS-CoV (Middle East Respiratory Syndrome). A new strain of Coronaviruses was discovered for the first time identified in humans in December 2019 in Wuhan, China. Where it was called a novel (new) coronavirus (COVID-19) by the World Health Organization (WHO). COVID-19 became a human tragedy affecting around 47 million people with more than a million deaths (WHO, 2020). The common symptoms of the disease are fever, dry cough, and fatigue. At the beginning of infection, the symptoms seem like seasonal flu with more severity attack. Some cases show additional symptoms like nasal congestion, aches, sore throat, and diarrhea. Usually, these symptoms start mild and then begin to intensifies gradually (Y. P. Zhang, 2020) (WHO, 2020). Many infected cases develop no symptoms and don't complain unwell. 80% of infected cases can recover without needing medical intervention. And around 16% of cases becomes seriously ill and develops an advanced symptom like difficulty breathing. Old age infected cases and those who have another co-morbidity like diabetes, heart problems or hypertension, are more likely to develop advanced symptoms (Heymann & Shindo, 2020) (Ayttey et al., 2020) (Guan et al., 2020)

COVID-19 is also having a great impact on the global economy (Matt Craven , Linda Liu is a, n.d.2020). Supply chains and global trade have been interrupted by the chaos of this suddenly occurred situation (Matt Craven , Linda Liu is a, n.d.). Under these circumstances, economic implications are thus prejudicial impact not only to global trade but agriculture industries, oil

prices, travel, food productions, and various market types (Evans, 2020). The uncertainty and fear situation created by the Coronavirus epidemic amplifying the volatility of the financial market, and seriously hitting the global economy (Albulescu, 2020). In addition to the Coronavirus epidemic impact on health systems and the global economy, it can have a psychiatric impact on persons with confirmed or suspected cases, that these persons may experience fear of the consequences of being infected with a fatal virus. Even those who in quarantine might experience anger, boredom, and loneliness. Timely psychiatric treatments may need to provide in some cases with severe mental health problems (Carvalho et al., 2020).

In the following sections we will discuss COVID-19 diagnostics tests, the global and Palestinian prevalence rates, infection transmission and incubation period, symptoms, clinical, laboratory, and radiological finding, co-morbidities, and treatment of covid-19. Then we'll talk about data mining and machine learning in covid-19

## **2.2. Diagnosis**

To diagnose COVID-19, several tests can be done, Real-time reverse transcription-PCR (RT-PCR) test remains the golden standard exam for etiologic diagnosis of COVID-19, either by nasopharyngeal (NP) swab or oropharyngeal (OP) swab to collect materials, and physical pieces of coronavirus. however, NP swabs are more sensitive than OP swabs for monitoring and diagnosis of COVID-19 and virus load. A positive result means that there is the genetic material of the virus in the spacemen and the person is infected, while a negative result indicates that the person was not infected or have a very low level of the virus (Tahamtan & Ardebili, 2020) (Salameh et al., 2020).

COVID-19 Antibody (Serology) Test also can be done as a supplementary tool, which can help to detect immunoglobulin's, IgG, and IgM (antibodies), the immune system produced antibodies as a response to an infection, that will be specific to this particular infection and can be found in blood specimens (serum or plasma) (Salameh et al., 2020)(Tahamtan & Ardebili, 2020).

Chest imaging for diagnosing COVID-19 includes chest X-ray radiography, computed tomography (CT), and ultrasound. Chest x-ray produces a 2-D image of organs and structures of chest area (lungs, airways, and heart), usually carried out in hospitals and medical centers to detect inflammation, blockages, and lung effusion. CT scan also uses x-ray radiation from different angles to produce axial 2-D images that can be constructed to produce 3-D images. Ultrasound scans use sound waves (high frequency) to produce 2-D and 3-D images (Tahamtan & Ardebili, 2020) (Cleverley et al., 2020). High-resolution lung CT scan is sensitive in the diagnosis of COVID-19 in suspected cases but not specific, where it could incorrectly identify COVID-19 of cases who were not infected as is not susceptible to differentiating between COVID-19 from other respiratory illness (false positive). There is limited data for evaluating the accuracy of chest x-ray and ultrasound in COVID-19 diagnosis but there is no adoption about their important role in following up COVID-19 patients status during the disease course (Tahamtan & Ardebili, 2020) (Salameh et al., 2020) (Cleverley et al., 2020).

## **2.3. Prevalence of COVID 19**

### **2.3.1. Global Situation**

According to World Health Organization (WHO) surveillance systems, to identify risks for infectious diseases like COVID-19, the total confirmed cases around the world have been about

121,969,223 cases (Last update: 19 October 2020, 02:00 EET), and around 2,694,094 confirmed deaths globally (Last update: 20 March 2021, 02:00 EET). (WHO, 2020). The cases are divided by WHO regions as shown in Figure (1)

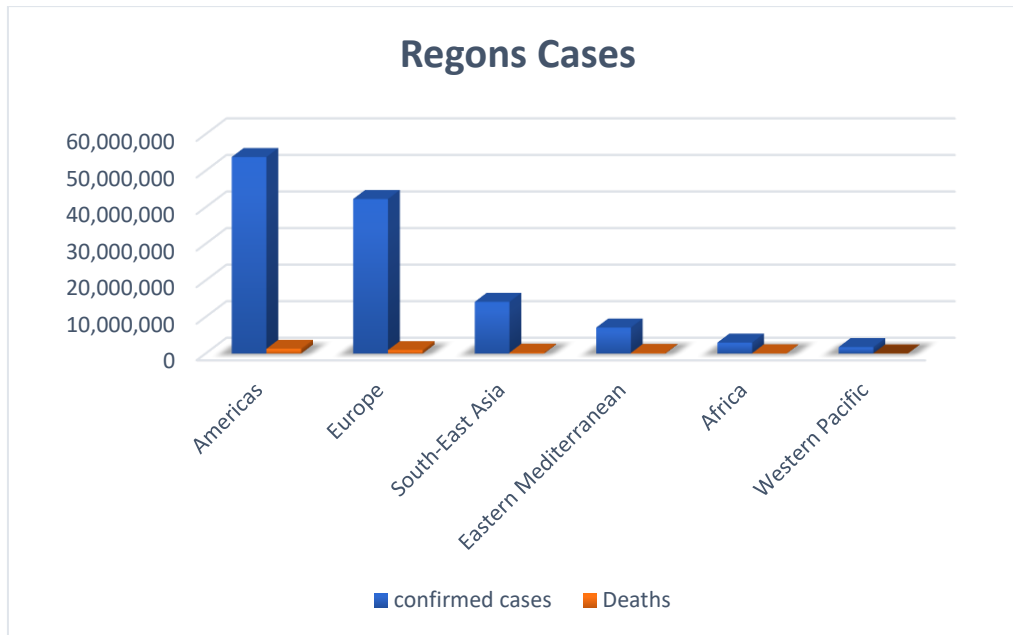


Figure 1: COVID-19 cases according WHO regions. (WHO, 2020)

### 2.3.2. Palestinian situation

Palestine is one of the affected countries by the COVID-19 pandemic. According to the Palestinian Ministry of Health and its CORONAVIRUS - COVID19 surveillance system the total confirmed cases in Palestine is about 251611cases, and around 2677 confirmed deaths, (WHO, 2020). As shown in figure (2)

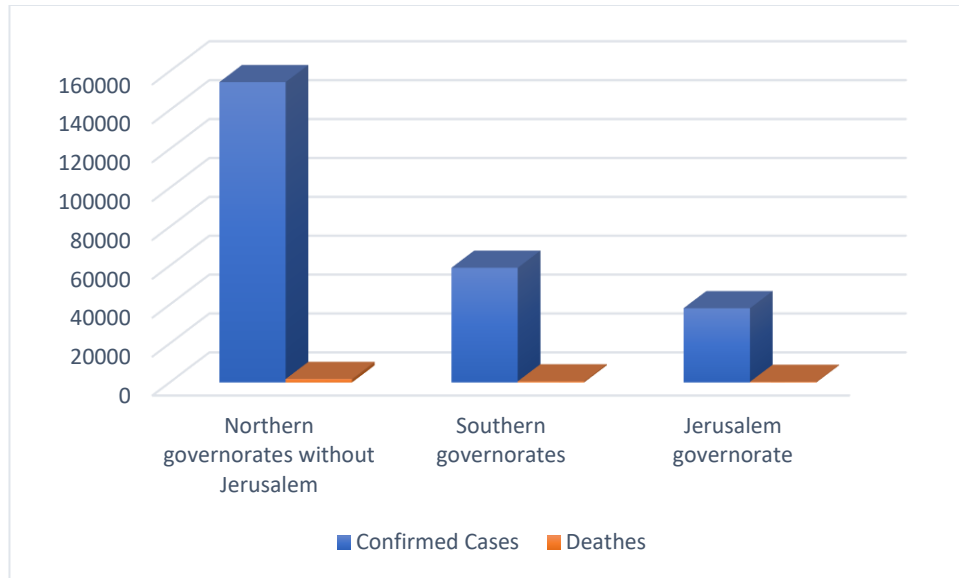


Figure 2: COVID-19 cases in Palestinian governorates. (WHO, 2020)

#### 2.4. Infection transmission and incubation period

The spread of disease from one person to another mainly can occur through the spread of nose or mouth droplets when an infected person exhales or coughs. The droplets from an infected person can be dropped down on surfaces around, which can transmit the disease when others touching these surfaces or objects by hands and then touching their face (mouth, nose, and eyes). People also can have infected through direct face to face breathing these droplets. That explains the importance of the recommendation of having a social distance for at least 1.5 meters as well as wearing mask (Guan et al., 2020) (WHO, 2020). Figure (3) Abbreviates the methods of transmission of infection.

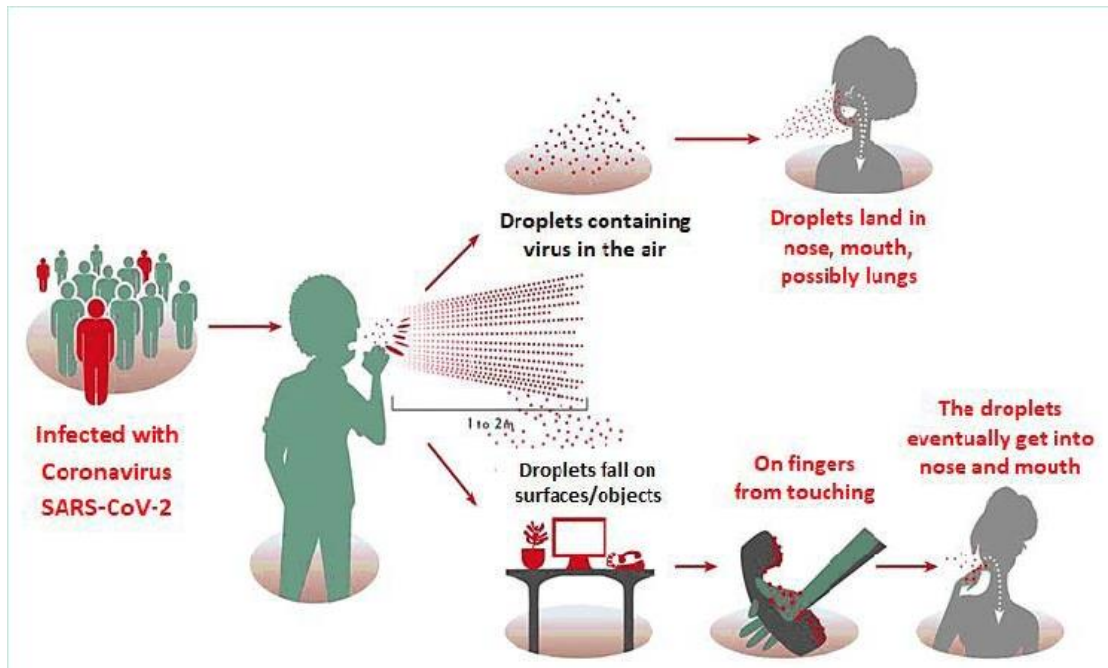


Figure 3: COVID-19 transmission (Tahamtan & Ardebili, 2020)

COVID-19 has an incubation period that ranges from 1-14 days according to most researchers' estimations, most commonly around 5-6 days. The incubation period is the period between the person infected with the virus and the appearance of symptoms of that virus. (WHO, 2020).

## 2.5. Symptoms

The common symptoms of COVID-19 are fever, dry cough, and fatigue. At the beginning of infection, the symptoms seem like seasonal flu with more severity attack. Some cases show additional symptoms like nasal congestion, aches, sore throat, and diarrhea. Usually, these symptoms start mild and then begin intensifies gradually. There is a marked difference between covid-19 confirmed cases in the appearance of symptoms or not, and their severity level. The majority of infected patients may not show any symptoms while the severity varies among others,

knowing that the disease remains contagious even without the appearance of symptoms. (L. Pan et al., 2020). As most of the cases show a moderate degree of symptoms that cause a mild cough, but some symptoms may develop a life-threatening complication, especially in the elderly and those who complain of other chronic diseases (Ali & Gatiti, 2020) (Ayithey et al., 2020). Where 80% of infected cases can recover without needing medical intervention. And around 16% of cases becomes seriously ill and develops an advanced symptom like difficulty breathing. Old age infected cases and those who have another co-morbidity like diabetes, heart problems or hypertension, are more likely to develop advanced symptoms (Guan et al., 2020) (Y. P. Zhang, 2020) (WHO, 2020).

Covid-19 can be presented with various symptoms from one patient to another according to their age group, comorbidities, and other factors. However, the most common general symptoms of covid-19 are fever, along with respiratory signs like cough, dyspnea, gastrointestinal symptoms (dysgeusia, loss of smell and taste), musculoskeletal symptoms (general weakness, fatigue, and joint pain) (L. Pan et al., 2020).

### **2.5.1. Symptom severity**

The symptoms severity level of COVID-19 can range from very mild to critical symptoms, some patients may complain only a limited symptom, and some patients may have no symptoms at all. Others may experience deteriorated symptoms. The severity of symptoms is also classified into four classes; asymptomatic or unknown, Mild-to-moderate symptoms, severe cases, and critical cases (Gao et al., 2020).

1. Asymptomatic patients come with no complaining.
2. Mild-to-moderate symptoms level patients can complain of fever, fatigue, and cough, or other less common symptoms like headache, malaise, sore throat, gastrointestinal, vomiting, and lack of appetite.
3. Severe symptoms can complain of shortness of breath, fever above (39.4°C), and persistent pain or pressure in the chest.
4. More serious symptoms are classified as critical cases, that patients can complain in addition to severe symptoms, they can develop life-threatening symptoms like a drop in heart rate, cold limbs that may lead to respiratory failure, sepsis, and multi-organ failure

## **2.6. Clinical finding**

In addition to the aforementioned symptoms of COVID-19 that manifested of lower respiratory tract infections, the upper respiratory tract symptoms were less common, suggest the most affected and targeted cells to the COVID-19 virus might be located in the lower airway. Which explains that around 2-9% of asymptomatic patients were discovered and diagnosed by abnormal lung appearance on CT scan. Therefore lung CT scan examination with IGG and IGM laboratory or serum antibody tests should be done for persons with a history of exposure to infected patients with COVID-19 (Shi et al., 2020). Relevant studies indicate that the right lower lobes of the right lung were the most commonly affected compared to the left lung(Tahamtan & Ardebili, 2020). This might be due to anatomical characteristics of the trachea and bronchi; that the right bronchus is shorter and more straight in comparison to the left one, the causative COVID-19 virus might tend to attack this site (Shi et al., 2020).

## 2.7. Laboratory finding

The importance of laboratory tests is not limited to diagnosing of COVID-19, as it has a major role in monitoring the patients and the body systems function, where in this section we highlight the most important laboratory tests that reveal the development of the patient's condition.

The hematopoietic system (*the system elaborate in the formation of the cells of blood*) was also impacted by COVID-19. Therefore, Lymphopenia (*reduced number of lymphocytes in the blood*) is one of the important laboratory findings. The platelet/lymphocyte ratio and Neutrophil/lymphocyte ratio are a prognostic indicator to determining the severity level of COVID-19. Lymphocyte count and inflammatory indices, including, Interleukin 6, CRP, and LDH, are cardinal during disease course for cases evaluation of those with a dismal prognosis and urgent intervention. Ferritin and procalcitonin are also linked with poor disease outcome. Thrombophilia or hypercoagulability (*an imbalance in naturally occurring blood-clotting proteins, or clotting factors*(Ferrari et al., 2020)) which is manifested in elevated D-Dimer levels is another essential test to monitor disease course as gradual elevated D-Dimer level usually associated with case worsening. Furthermore, partial Thromboplastin time (PTT) or activated partial Thromboplastin time (aPTT or APTT) is another test to monitor blood coagulation abnormalities, which when increase with severe thrombocytopenia lead to life-threatening as it indicates that there are blood clots throughout the bloodstream, may blocking the small blood vessels or Disseminated intravascular coagulation (DIC) (Terpos et al., 2020)(Ferrari et al., 2020).

## **2.8. Radiological finding**

Chest computed tomography (CT scan) of patients with COVID-19 usually show unilateral or bilateral ground-glass opacification, possibly progressed to consolidation within the first to the third week from symptom onset. However, most infected patients showed bilateral lung embroilment, with opacification or consolidation usually located peripherally (sub-pleural). The dominant pattern on CT scan finding of infected patients was ground-glass opacity, with air bronchograms, ill-defined margins, septal, and adjacent pleura thickening. Findings at chest radiography in infected patients reflect those at CT scan of infiltration and opacification with less sensitivity (Wong et al., 2020) (F. Pan et al., 2020).

## **2.9. Co-morbidities**

As mentioned earlier, the existing of comorbidities is one of the most important factors that affect the severity symptoms of COVID-19. Some comorbidities have more influence than others on the disease outcomes. In the following sections we will discuss some of the co-morbidities that affect and affected by COVID-19:

### **2.9.1. Hypertension**

According to the American Heart Association (AHA) and American College of Cardiology (ACC), hypertension defined as elevated diastolic blood pressure (BP)  $\geq 80$  mm<sup>3</sup> or systolic blood pressure (BP)  $\geq 130$ mm<sup>3</sup>, or blood pressure in arteries is persistently elevated (Kanwal et al., 2020).

Relevant reports from multiple COVID-19 infected hot spot areas, including New York City and Wuhan, indicates a high correlation and proportion relation between severity level of the disease and having hypertension, as a higher rate of patients with hypertension were identified among severity ill of hospitalized COVID-19 infected patients. Despite these reports, the mechanisms of affection and link between COVID-19 and hypertension are unclear. As the severity of COVID-19 symptoms is skewed toward elderly patients with a higher prevalence of hypertension, this over the presentation of patients with hypertension comorbidities could confound by age or other comorbidities like diabetes mellitus, cardiac disease, chronic kidney disease, and obesity (Kanwal et al., 2020)

### **2.9.2. Diabetes mellitus**

Diabetes mellitus (DM) defined as a disorder that impairs the ability of the body to process blood glucose, where the body does not produce sufficient insulin or does not use it efficiently (Anderson et al., 2020). Many studies linked between having diabetes and increase the chance of developing severe symptoms or worse outcome of patients with COVID-19, while experiencing severe symptoms and serious complication is probably to be lower with well-managed diabetes patients. However, in the case of fluctuating blood sugar, patients with COVID-19 fall at risk of many of diabetes related complication behind COVID-19 symptoms. Forth more having other comorbidities like hypertension or heart disease in addition to diabetes could compromise the body ability to resist the virus and worsen the chance of developing serious complications from COVID-19 (Anderson et al., 2020).

### **2.9.3. Malignancy**

Which is the most popular characteristic of cancer cells, or the term refers to the existence of that cancerous cells; an abnormal and uncontrolled division of cells that can invade nearby body tissue. The effect of malignancy as comorbidities of COVID-19 patients varies according to the site of disease, different biological behaviors of cancer cells, stage, patient's health status, and medication use. But many studies agreed that COVID-19 patients with malignancy had a higher risk of getting severer symptoms and worse outcomes compared with patients without malignancy. As cancer patients are advised to avoid immunosuppressant drugs in case of confirmed to have COVID-19 infection (Radin & Tsirka, 2020).

### **2.9.4. Cerebrovascular diseases**

Conditions and disorders of blood vessels that supplied the brain includes blockage (carotid stenosis, stroke, vertebral stenosis, and intracranial stenosis), aneurysms, and malformation that restrict blood flow to the brain, which prevent its cells from getting enough oxygen (Matt Craven , Linda Liu is a, n.d.) cerebrovascular disorders were found to be associated with an increase chance of developing severe COVID-19 symptoms and critical illness (Matt Craven , Linda Liu is a, n.d.).

### **2.9.5. Chronic kidney disease**

Chronic kidney failure is a term that describes gradual damage of kidney function, as kidneys can't do their function to filter blood from excess fluids and wastes, and this dysfunction can get worse over periods (Ahani & Nilashi, 2020). CKD also seems to be strongly associated with an increase

in the risk of severe COVID-19 symptoms, as many reports advise CKD patients to take extra caution to reduce the risk of infected with COVID-19 (Ahani & Nilashi, 2020).

### **2.9.6. Chronic liver disease**

Is gradual destruction of the tissue of the liver over time, which can be categorized into liver fibrosis or liver cirrhosis (Evans, 2020) it remains ambiguous that chronic liver disease considers as a risk factor and correlated with the risk of a severe course of COVID-19, despite elevated liver enzymes; AST and ALT and slightly elevated bilirubin for COVID-19 patients at the time of admission and association of elevated those enzymes with higher mortality (Evans, 2020).

### **2.9.7. Neurological diseases**

Disorders that affect the brain and nerves system throughout the human body can be electrical, structural, or biochemical abnormalities (Wu et al., 2020). Viral infection in general can impact nerves system function, even can produce severe neurological damage. It was reported that the COVID-19 virus can be found in cerebrospinal fluid or even in the brain, however, it's still unknown how COVID-19 impact effect the nerves system, despite many reported confirmed the relation of having a previous neurological disorder and developing severe symptoms (Wu et al., 2020).

### **2.9.8. Cardiovascular disease (CVD)**

Is a class of conditions or disease that affecting heart or blood vessels, commonly associated with fatty deposits in arteries, that increase the risk of a build-up of blood-clots. That can cause many types of CVD as Coronary heart disease, Strokes and TIAs, Peripheral arterial disease, and Aortic disease (Apicella et al., 2020a). CVD comorbidities are highly correlated at a higher risk of mortality and morbidity, as most COVID-19 patients with CVD experience severe or critical symptoms (Carvalho et al., 2020).

### **2.9.9. Respiratory disease**

It's a wide range of lung diseases that affect the respiratory system and limit its function such as Chronic obstructive pulmonary disease (COPD), asthma, and cystic fibrosis (Pranata et al., 2020). Since COVID-19 is a respiratory syndrome, it's been logical to increase the chance of developing severe health complication when having a previous respiratory disorder, that can put COVID-19 patients at risk of poor prognosis and worse disease outcome (Pranata et al., 2020).

### **2.10. Treatment of COVID-19**

Scientists and researchers around the world are making tremendous efforts to develop an effective treatment of COVID-19. Till the time of writing this thesis, there is no specific treatment for the disease, with many drugs are under investigation, physicians on their side follow international protocols provided by WHO, and giving some medication to reduce the exacerbation of the disease and relief its complications.

Chloroquine phosphate is one of the most drugs used in China and Italy. However, the drug is recommended by National Health Commission to be included for prevention and treatment of COVID-19 next version Guidelines as it showed acceptable safety and effectiveness toward COVID-19 associated pneumonia (Carvalho et al., 2020). Moreover, Hydroxychloroquine shows efficacy against COVID-19, as when reinforced by azithromycin, be associated with viral load reduction within COVID-19 infected patients (Heymann & Shindo, 2020). Glucocorticoids suggest to likely decrease mortality and mechanical ventilation through severe COVID-19 patients. And there is uncertainty if Remdesivir drug has an impact on an important outcome like mortality, despite it reduce the time of severe symptoms. Table (1) below shows a summary of the effects of the drug interventions on the outcomes. The following table is a current systematic review and network meta-analysis for 25 electronic databases from the US Centers for Disease Control and Prevention and six additional Chinese databases (Carvalho et al., 2020).

Table 1: summary of the effects of the drug interventions on the outcomes (Carvalho et al., 2020).

|                                   | Mortality           | Mechanical ventilation | Adverse events    | Viral clearance    | Admission to hospital | Duration of hospital stay | ICU length of stay | Duration of mechanical ventilation | Time to symptom resolution | Time to viral clearance |
|-----------------------------------|---------------------|------------------------|-------------------|--------------------|-----------------------|---------------------------|--------------------|------------------------------------|----------------------------|-------------------------|
| Standard care*                    | 330 per 1000        | 116 per 1000           | 15 per 1000       | 500 per 1000       | 41 per 1000           | 7 days                    | 10 days            | 10 days                            | 19 days                    | 7 days                  |
| Glucocorticoids                   | -31 (-55 to -5)**   | -28 (-45 to -9)***     |                   |                    |                       | -1.0 (-1.4 to -0.6)****   |                    |                                    |                            |                         |
| Favipiravir                       | -330 (-330 to 670)  |                        |                   |                    |                       |                           |                    |                                    |                            |                         |
| Hydroxychloroquine                | 13 (-15 to 43)**    | 19 (-4 to 45)***       | 16 (-11 to 192)** | 82 (-343 to 414)   | -19 (-43 to 26)       | -0.4 (-3.8 to 2.4)        |                    |                                    | -4.7 (-6.0 to -3.0)        | -0.7 (-3.9 to 5.5)      |
| Hydroxychloroquine + azithromycin | -105 (-246 to 102)  | 57 (-15 to 162)        |                   |                    |                       | 0.6 (-0.8 to 2.0)****     |                    |                                    |                            |                         |
| Lopinavir-ritonavir               | -71 (-181 to 77)    |                        |                   | -243 (-479 to 237) |                       | -1.3 (-2.4 to -0.3)****   |                    |                                    | -1.1 (-2.1 to -0.4)        |                         |
| Remdesivir                        | -91 (-154 to -14)** | -23 (-47 to 8)***      | 3 (-7 to 43)      | 11 (-470 to 473)   |                       | 0.3 (-3.8 to 4.5)         |                    |                                    | -2.6 (-4.3 to -0.6)        |                         |
| Umifenovir                        | -330 (-330 to 670)  |                        |                   |                    |                       |                           |                    |                                    |                            |                         |

|                         |                 |                      |                       |         |
|-------------------------|-----------------|----------------------|-----------------------|---------|
|                         | Most beneficial | Intermediate benefit | Not different from SC | Harmful |
| High/moderate certainty |                 |                      |                       |         |
| Low/very low certainty  |                 |                      |                       |         |

\*Numbers presented are absolute risk differences (95% credible interval) per 1000 patients or mean difference (95% credible interval) when compared to standard care  
 \*\* Random effects NMA estimates (versus standard care): Glucocorticoids, -25 (-89 to 77); Hydroxychloroquine, 16 (-56 to 110); Remdesivir, -85 (-161 to 20)  
 \*\*\* Random effects NMA estimates (versus standard care): Glucocorticoids, -23 (-56 to 53); Hydroxychloroquine, 22 (-35 to 106); Remdesivir, -24 (-63 to 35)  
 \*\*\*\*The best estimate of effect is from direct (pairwise) meta-analyses  
 Empty cells: there was insufficient or no evidence for this drug/outcome

## 2.11. Data mining and machine learning in COVID-19

While researchers around the world are implementing a different research approach to face this epidemic, data mining techniques had a major role in explaining and analyzing some hidden matters regarding this disease. In general Artificial intelligence can be used to discover and explore novel, knowledge and hidden patterns from the dataset (J. C. Saire & Panford-Quainoo, 2020).

Researchers around the world use data mining techniques and artificial intelligence in different aspects to understanding the disease behaviors.

### **2.11.1. Data mining techniques**

Multiple Data mining techniques were applied to classify and predict the severity level of COVID-19, therefore we applied the most popular Machine learning algorithm of Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Networks (ANNs).

#### **2.11.1.1 Logistic regression (LR)**

It is a statistical and data mining model that can be used in a situation relating to variable outcome with 3 or more categories that have many potential predictor variables (multinomial, dependent variables) with sigmoid function, it's only deal with output value from zero to one representing the probabilities of output in between (Menard, 2002).

#### **2.11.1.2 Decision tree (DT)**

Decision tree (DT) is one of the popular data mining models to remotely classify data with a systematic flowchart-like sequential structure, containing root nodes, internal nodes, branches, and leaf. Which each internal nodes represent a "test" on the feature, branches represent test outcome, leaf node represents the class label, DT consider to have an advantage over other models of

Machine Learning (ML), while DT refers to fast logarithm computation that didn't make a statistical assumption and interpret the results easily (Pal & Mather, 2003).

### **2.11.1.3 Support Vector Machine (SVM)**

Support vector machine is a supervised model in machine learning, analyzing the data for regression analysis and classification with associated learning algorithms. SVM has been widely practiced in sciences, hypertext categorization, image, and hand-written recognition (Barzilay & Brailovsky, 1999).

### **2.11.1.4 Artificial neural networks (ANN's)**

ANN is a section of a computing system intended to simulate the human brain technique to analyze and process information. Formed by input, hidden and output layer(s), the networks between the nodes of adjacent layers have weights allied for each level. The goal of learning is to give correct weights for these edges, as ANN have self-learning abilities that allow them to produce better results as more data becomes offered (Barzilay & Brailovsky, 1999; Zurada, 1992).

## **2.11.2. Model evaluation**

Regarding our intention to develop a model to classify and predict the development of the patient's condition (Severe/ Non-Severe) during their stay in the hospital, it is important to find a practical and scientific way to test the accuracy of the multiple methods used in this regard, for a given severity level, the best likely test can be used based on the given attributes are sensitivity,

specificity, and accuracy, which are commonly used statistics used to measure how good and reliable the diagnostic tests are. While sensitivity assesses the ability of the model in detecting a positive result (Severe symptoms). Specificity assesses how likely patients with Non-severe symptoms can be correctly predicted by the used model. Area Under the Curve (AUC) estimates the ability of a classifier to differentiate between classes, as the higher the AUC, the better the performance of the model at distinctive between negative and positive classes. However, the accuracy of the used models is measured how correct is it to identifies and exclude a given outcome, which can be calculated and determined from the sensitivity and specificity of that model (Zhu et al., 2010).

Numerous terms are usually used along with the description of sensitivity, specificity accuracy, true positive (TP), true negative (TN), false negative (FN), and false-positive (FP). If the severity level of symptoms is proven to be severe in a patient, and the given results of the model indicate to be severe, the results of the used model are considered a true positive. Similarly, if the severity level of symptoms is proven to be non-severe in a patient, and the given results of the model indicate to be non-severe, the results of the used model are considered a true negative. Nevertheless, no medical test is perfect. If the severity level of symptoms is proven to be non-severe in a patient, and the given results of the model indicate to be severe, the results of the used model are considered a false positive. Similarly, if the severity level of symptoms is proven to be severe in a patient, and the given results of the model indicate to be non-severe, the results of the used model are considered a false negative. While both false positive and false negative indicate that the used model is opposite to the actual status (Zhu et al., 2010).

We can calculate the accuracy, sensitivity, and specificity by using these formulas:

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative}}$$

$$\text{Sensitivity} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

$$\text{Specificity} = \frac{\text{TrueNegative}}{\text{TrueNegative} + \text{FalsePositive}}$$

### 2.11.3. Relevant Studies

The using of artificial intelligence is not limited to machine learning, as some researchers have used data mining techniques to present an exploratory analysis of COVID-19 datasets in Sao Paulo State / Brazil, to produce a cleaned dataset for future researches by reducing data inconsistencies like outliers and null reference values for further analyses, (J. E. C. Saire, 2020).

Carlos Medel-Ramírez, et al, conducted an algorithm of data mining on open-source data for daily COVID-19 patients registry according to their sex, age, past medical history for the tree classification of cases (positive, negative, suspected) regarding their status as an outpatient or hospitalized attending medical development (Deaths, Recovered) to provide fast and timely

information needed to estimate the medical care scenarios of COVID-19 patients (Medel-ramírez et al., n.d.).

Fatima Ahouz, et al, used techniques of learners regression to predict the prevalence of COVID-19 for 17,136 records in different geographical regions include the variables of longitude, latitude, date, and records, the model technique presenter for 3 groups (prevalence <200 cases, error rates 9.42), (prevalence 200-1000 cases, error rates 17.8), (prevalence > 1000 cases, error rates 12.26), she uses models predict that the prevalence in the next two weeks from (March 30- 2020) expected to be over 300,000 in America, over 600,000 in Europe, 14,000 in Australia, 100,000 in Asia, and over 19,000 new cases in Africa (Ahouz & golabpour, 2020).

Regarding uses of data mining models in prediction (L. J. Muhammad et al) develop a model for the prediction of COVID-19 patient's recovery using the epidemiological dataset of COVID-19 in South Korea, by python program language applying the algorithms of K-nearest neighbour algorithms, support vector machine SVM, decision tree, and naive Bayes to predict the lowest and highest duration for COVID-19 patients to recover. The model developed by decision tree algorithm is the most efficient for recovery possibility with an accuracy of 99.85% which consider being the best model developed among the use's models.

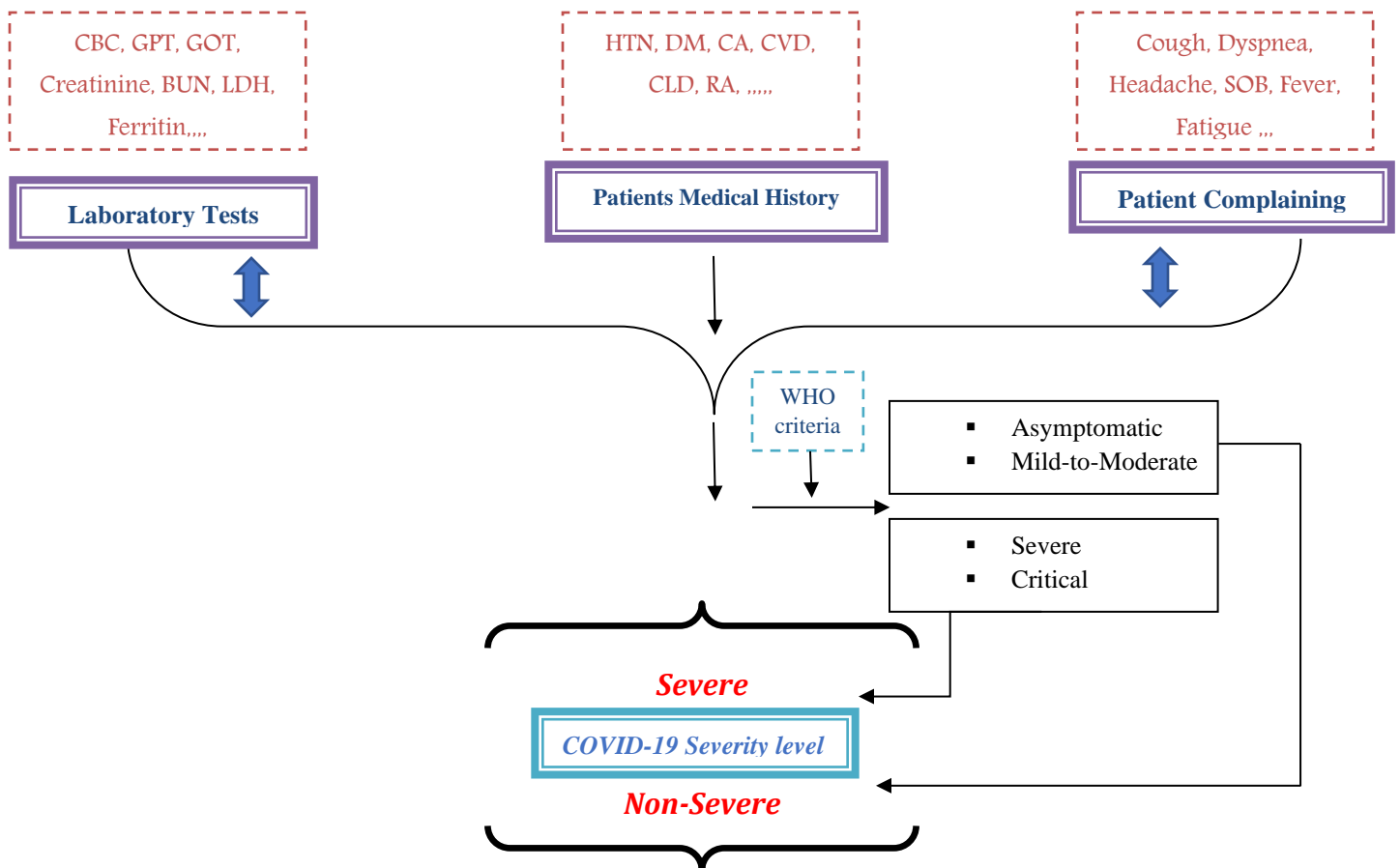
To our knowledge, no research utilizes artificial intelligence in Palestine to determine the health status of the COVID-19 patients in the future. As in this study, we considered analyzing four different popular data mining tools, to evaluate which tool scored the best accuracy in classifying the severity level of COVID-19. Starting with evaluating and describe the situation of COVID-19, in term of discovering its characteristics and link it with severity level of hospitalized patients.

### 3. Chapter 3: Conceptual Framework for COVID-19

---

#### 3.1. Introduction

In this chapter we will illustrate our conceptual framework and defined the variables with explaining how it related to each other, as we considered the severity level (Severe / Non-Severe) as a dependent variable that is directly affected by independent variable of patient's medical history, patient demographic factors, laboratory tests, and patient's complaining according to the following chart:



## **3.2. Variables**

### **3.2.1. Demographic variables**

Demographic variables are an independent variable used to describe the nature and distribution of the study sample, related to the patient sex (Male, Female), age (no determined age group), and its association with study outcome of severity level.

### **3.2.2. Medical history**

The patient medical history may contain relevant information about the patient health status in the past that being an explanation of all medical events and harms a patient has experienced, that considered a key tool in the medical management of the patient.

The past medical history or underlying comorbidities were categorized into 11 major groups (Hypertension (HTN), Diabetes (DM), Malignancy (CA), Cerebrovascular Disease, Chronic Kidney Disease (CKD), Chronic Liver Disease, Rheumatoid (RA), Cardiac Disease, Neurological Disorder, Chronic obstructive pulmonary disease (COPD) and, Respiratory Disorder)., where each underlying comorbidity were defined in previous chapter (chapter two).

### **3.2.3. Laboratory tests**

We chose the laboratory tests applicable to COVID-19 patients by first, collecting all the laboratory tests in the first two days from the patient's admission to the hospital in the beginning, and then we assign the tests for all patients were approved as a protocol of COVID-19 patients and excluded the tests specifically applied to some patients and not others, also we excluded the tests

that have many missing values. However, the final laboratory tests we work with were: white blood cells (WBCs), Neutrophils granulocyte percentage, lymphocytes percentage, hemoglobin levels (HGB), platelets count, coagulation testing, liver and renal function (BUN, Creatinine, GPT, GOT), measures of electrolytes, C-reactive protein (CRP), lactate dehydrogenase (LDH), D-dimer, and Ferritin level.

### **3.2.4. Complaining**

Patient's Chief complaint which represents the reason for the patient visit to the hospital as itemized in the patient's own words. There was a variation from one patient to another, However, the complaints were limited to 19 complaints as follows (Cough, Dyspnea, Headache, Diarrhea, shortness of breath SOB, Fever, Nausea, Chest Pain, Sore throat, Chills, Tachypnea, Vomiting, Fatigue, Runny nose, Dizziness, Hemoptysis, Weakness, Arthralgia, Loss of appetite). These complains were used as independent variable to explore its relationship with the study outcome of (severe, non-severe).

### **3.3. WHO criteria for severity level staging**

Disease severity *is a term used to characterize the impact that a disease process has on the utilization of resources, comorbidities, and mortality* (Gambert, 2013). The World Health Organization had classified the symptoms of COVID-19 disease into four levels, as shown in Table (2).

Table 2: COVID-19 symptoms classification (Gao et al., 2020).0.01

|                  |             |   |   |
|------------------|-------------|---|---|
| Mild-to-moderate | main        | 1 | Fever   |
|                  |             | 2 | Cough   |
|                  |             | 3 | Fatigue   |
|                  | less common | 1 | sore throat   |
|                  |             | 2 | Headache  |
|                  |             | 3 | nasal congestion  |
|                  |             | 4 | Malaise   |
|                  |             | 5 | muscle aches  |
|                  |             | 6 | gastrointestinal issues, such as diarrhea, nausea, or vomiting    |
|                  |             | 7 | lack of appetite  |
| Severe cases     |             | 1 | high fever above (39.4°C)   |
|                  |             | 2 | difficulty breathing  |
|                  |             | 3 | blue lips or face   |
|                  |             | 4 | persistent pain or pressure in the chest                          |
|                  |             | 5 | new-onset confusion (“altered mental status”)                     |
|                  |             | 6 | trouble waking up, even when another person tries to wake them    |
| Critical cases   |             | 1 | weak pulse  |
|                  |             | 2 | cold hands or feet  |
|                  |             | 3 | requiring medical treatment to maintain organ function and life   |
| Asymptomatic     |             |   | do not have any known symptoms of COVID-19 or associated symptoms |

For the purposes of this study, we divided the severity of symptoms into two groups, first: classified the patient's symptoms as Not Severe when they were asymptomatic or have a mild to

moderate symptoms. Second: classified the patient's symptoms as Severe when they were having a severe or critical symptom.

## **4. Chapter 4: Study methodology**

---

### **4.1. Introduction**

In this chapter, we described the study design used in our study. A quantitative research design using retrospective review of patient's medical records was carried out in two hospitals designated to provide services for COVID-19 in Palestine. The process of data collection, cleaning, and analysis were clarified in this chapter.

### **4.2. Study settings and population characteristics**

#### **4.2.1. Study population**

The study was conducted in West Bank (WB). Data were obtained from two hospitals designated for providing services for COVID-19 infected patients in Palestine, namely, Hugo Chavez Hospital and Hebron Governmental Hospital. These hospitals were the only two hospitals for COVID-19 patients. It was firstly selected and fully equipped by the Ministry of Health to be highly qualified centers to deal with COVID-19 cases in the south and north Palestine and under the supervision of the general hospital directorate and to be able to receive the largest numbers of COVID-19 patients.

### **4.2.2. Study setting**

1. Hugo Chavez hospital: is a governmental ophthalmic hospital located in Ramallah city. That hospital was converted to be COVID-19 hospital to provide services for infected cases from the north of the West Bank at the beginning of disease spread since March 2020 with a capacity of 60 beds and 10 ICU beds.

2. Hebron governmental hospital is a central governmental hospital in Southern West Bank in Hebron city. That hospital designated to provide services to COVID-19 patients since June 2020 with a capacity of 120 beds and 20 ICU beds.

So, the study population included all admitted patients confirmed with COVID-19, with a total population of 199 patients at the time of the study

### **4.2.3. Study setting and participants**

The study setting consisted of the two COVID 19 designated hospitals (Hebron governmental hospital and Hugo Chavez hospital). All patients admitted to these hospitals during the period from 19 March 2020 to 20 July 2020 were included.

## **4.3. Study design**

A Cross-Sectional study was conducted using a quantitative retrospective review of patient's medical records of COVID -19 patients. More details on the study sampling and data collection will be explained in the following sections.

## **4.4. Study sampling and sampling techniques**

A web-based program (RAOSOFT Sample size calculator) was used to calculate the sample size according to (confidence interval CI 95% and 5% margin of error). A total of 132 patient records

were randomly selected by simple randomization. Medical records were reviewed from the two main hospitals mentioned between 19 March 2020 and 20 July 2020.

## **4.5. Data collection**

All patients medical record was imported on excel sheet from the health information system of the Palestinian Ministry of Health, then all patient file reviewed by data collection team. The data collection team received intensive training on data collection methods with a real training on a ten-patient file to ensure reliability and validity of the data collected. Data collection team were one nurse and one medical technology with extensive experience in hospital health information system. The researcher made sure that the data collection team is completely committed to the data collection procedure. First, we did a statistical analysis on the data extracted from 132 patient's medical record. The first phase of data analysis was to determine the most relevant variables that needed to be used for data mining. Another 67 patient records of the total population were added from the same population for data mining analysis. Increasing the sample or input records will increase the results accuracy of the used model. Inclusion and exclusion criteria for extracted medical records are described below.

### **4.5.1. Inclusion criteria**

1. All hospitalized COVID-19 patient that had a laboratory-confirmed SARS-CoV-2 infection with RT-PCR assay (Nasal and/or pharyngeal swabs) setting.
2. Hospitalized patients between 19 March 2020 and 20 July 2020.

### 4.5.2. Exclusion criteria

1. The patients who died in the first 12 hours and had no chance to ensure required laboratory tests.
2. Features inside Patients record were empty or invalid and did not contain the required data.
- 3.

### 4.6. Study tools

The data collection sheet (as shown in Appendix 1) was divided into six categories table (3): demographic data, past medical history, vital sign observation, laboratory findings, treatment plan, and clinical event. Details of each category can be followed below:

*Table 3: Data collection categories*

| No | Category             | Differentials   |
|----|----------------------|---|
| 1  | demographic data     | Age, sex.   |
| 2  | Past medical history | Past medical history or underlying comorbidities were categorized into 11 major groups (Hypertension (HTN), Diabetes (DM), Malignancy (CA), Cerebrovascular Disease, Chronic Kidney Disease (CKD), Chronic Liver Disease, Rheumatoid (RA), Cardiac Disease, Neurological Disorder, Chronic obstructive pulmonary disease (COPD) and, Respiratory Disorder). |
| 3  | Vital sign           | Which consists of Blood Pressure (BP). heart rate (P). Temperature (T), and respiratory rate (RR)   |
| 4  | laboratory findings  | Laboratory findings include white blood cells (WBCs), Neutrophils granulocyte percentage, lymphocytes percentage, hemoglobin levels (HGB), platelets count, coagulation   |

---

|   |                 |  |
|---|-----------------|--|
|   |                 | testing, liver and renal function, measures of electrolytes, C-reactive protein, lactate dehydrogenase (LDH), and Ferritin level were assessed).   |
| 5 | Treatment plan  | were categorized into 15 major groups (Antibiotic, Antipyretics, Supplementary, Oxygen's therapy, Glucocorticosteroids, antihypertensive, anticoagulant, Anti-asthmatic, Anti-cough, Antihistamines, Proton pump inhibitors (PPIs), Antiviral and Hydroxychloroquine anti-rheumatic) |
| 6 | Clinical events | The primary composite endpoint was defined by following up the patient electronic medical records for evaluating case severity level (severe or non-severe).   |

---

#### 4.7. Study variables

For purpose of data mining a review of many previous up to date studies and a comprehensive discussion with two intensive care specialist physicians working in COVID-19 treatment hospitals, the following variables were identified to be studied:

- Demographic data: age, sex
- Comorbidities: (hypertension, diabetes, malignancy, cerebrovascular disease, chronic kidney disease, chronic liver disease, rheumatoid arthritis, cardiac disease, neurological disorder, and respiratory disorder).
- Laboratory findings include white blood cells, neutrophils granulocyte percentage, lymphocytes percentage, hemoglobin levels, platelets count, coagulation testing, liver and renal

function, measures of electrolytes, C-reactive protein, lactate dehydrogenase, and Ferritin level were assessed).

- Severity level: severe and non-severe.

#### **4.8. Data analysis**

**Descriptive statistical analysis** was performed using the SPSS software V 26, the raw data was exported to SPSS, Preprocessing and data cleaning was done. Continuous variables were reported as median and interquartile ranges; However, categorical variables were summarized as counts and percentages. Disease severity levels were performed and reported and compared among age, sex, and co-morbid groups. Two-tailed t-test used to examine the differences between subjects for continuous data while the chi-square test for a categorical variable, linear, and multiple regression analysis performed to test the effects of variables on the severity level. Significant level was considered at  $P\text{-value} < 0.05$ .

Visualization and graphical representation of data were done by using visual elements such as graphs, charts, and box plots, to provide an attainable way to see and understand trends, outliers, and data patterns.

MATLAB was used to run different data mining methodologies, we considered analyzing four different popular data mining tools, to evaluate which tool scored the best accuracy which is defined as "*the percentage of correct predictions for the test data*"(JORDAN, 2017). So it is a measurement used to decide which model is superior at recognizing relationships and patterns

between variables based on the input, or training data. The better a model can generalize to hidden data, the better predictions and insights it can produce.

Our study analyzed data using Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Artificial Neural Network (ANN). The analysis adopted a 70% training and 30% testing basis, then the accuracy for the four models is calculated as the final accuracy score for each tool.

To achieve the study objectives, data analysis went through the following process:

- **Step 1:** Data collection and dataset preparation: The data will be collected from Palestinian MOH, preprocessing tools like cleaning, outlier, normalization will be applied.
- **Step 2:** Use SPSS to produce a descriptive analysis to discover the correlation between variables.
- **Step 3:** Visualize the results of data prepared in step 2 to better understand our datasets and assign which statistically significant variables can be used in the next step.
- **Step 4:** Data mining tools, apply different computing and machine learning data analysis tools to develop a prediction model.

#### **4.9. Work pathway**

After statistically analyzed the captured datasets through SPSS, we identified the most influencing variables on the disease severity level following phases:

- An extensive literature review was carried out to take advantage of relevant work.
- Analyzed the datasets using statistical methods and results visualization to identify the most relevant variables.

- Expert consultation, two physicians (intensivist and general practitioner) who had experience in COVID-19 treatment centers.

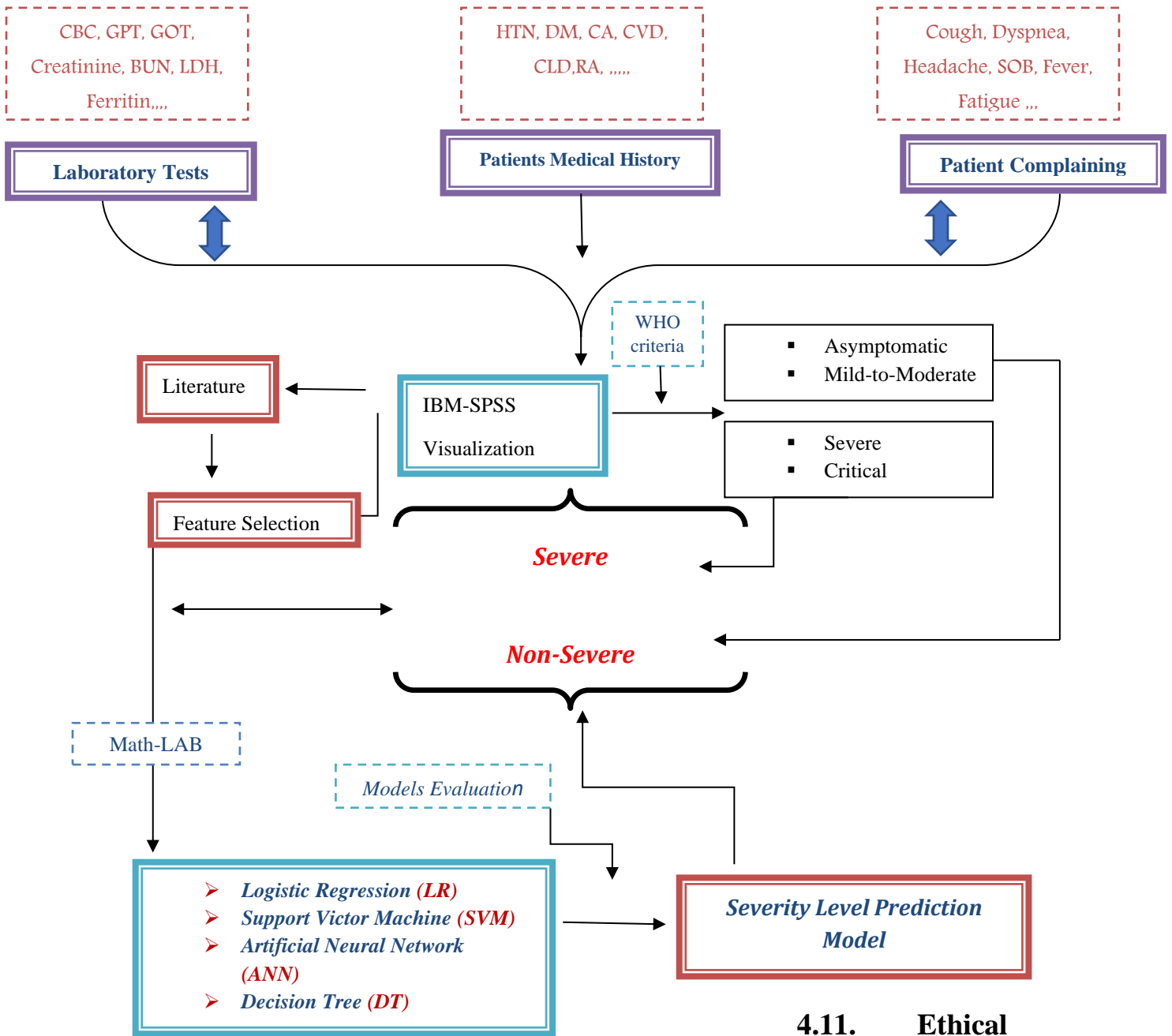
Where we selected the most influential variables on disease severity level that can be utilized to develop a severity level prediction model using an artificial intelligence (AI) models of Logistic Regression (LR) , Support Vector Machine (SVM) , Artificial Neural Network (ANN), and Decision Tree (DT) with performing an evaluation method for each produced model to define its accuracy and validity.

#### **4.10. Pathway model**

In this section we present our created research pathway model to achieve the study goals by collecting data from four main categories “demographic, laboratory test, patient complaining, and patient medical history”.

Which were used to define the severity level of COVID-19 patients according to WHO criteria that ended with descriptive analysis and visualization of main study findings that has been done by SPSS V26, then used in line with literature review and expert discussion to select feature to be used for four machine learning algorithms “LR, DT, SVM, and ANN” that has been accomplished by MATLAB program as shown in Figure (4)

Figure 4: Represent the research pathway model



**4.11. Ethical**

**consideration**

This study was approved by the Ethical Committee of the Palestinian Ministry of Health with consideration of all requirements to protect patient's privacy and ensuring an adequate level of

confidentiality (see appendix 2). IRB approval was obtained from Arab American University-Palestine (AAUP) to conduct this study. (See appendix 3)

Ethical consideration is one of the most important parts of this research, while after obtaining the necessary permissions from MOH; we ensure the following confidentiality steps:

- Research participant's privacy has been protected by ensuring an adequate level of confidentiality.
- Participants had not had subjected to harm in any way, Anonymity of individuals participating in the research was ensured.
- Any communication related to the research was done with honesty and transparency.
- Any representation of patient's data in a biased way, as well as misleading information, was avoided.

#### **4.12. Summary**

In this chapter, the methodology of our study was comprehensively presented as study setting, study population, and sample, in addition to determining those who were included/excluded in the study. Moreover, ethical considerations of our research were discussed to preserve patient privacy. The study findings and results will be reported in the following chapter.

## 5. Chapter 5: Data analysis and research findings

---

### 5.1. Introduction

This chapter presents the study results done to assess the current situation of COVID-19 in Palestine, that describes its characteristics and associate it with the severity level of hospitalized patients depends on statistical analysis and machine learning techniques. the chapter started with a brief description of how data was prepared, then demographic and clinical characteristics description of the sample, followed by descriptive analysis with visualization of the study sample, and end with testing and applying different prediction models.

As clearly mentioned previously patients with COVID-19 are considered to have severe illness if they meet the aforementioned WHO criteria for severe or critical level of illness. These patients may experience rapid clinical deterioration. The following variables are the main variables that are studied:

- Demographic data: age, sex.
- Comorbidities.
- Laboratory findings.
- Severity level: severe and non-severe.

## **5.2. Data preparation and cleaning**

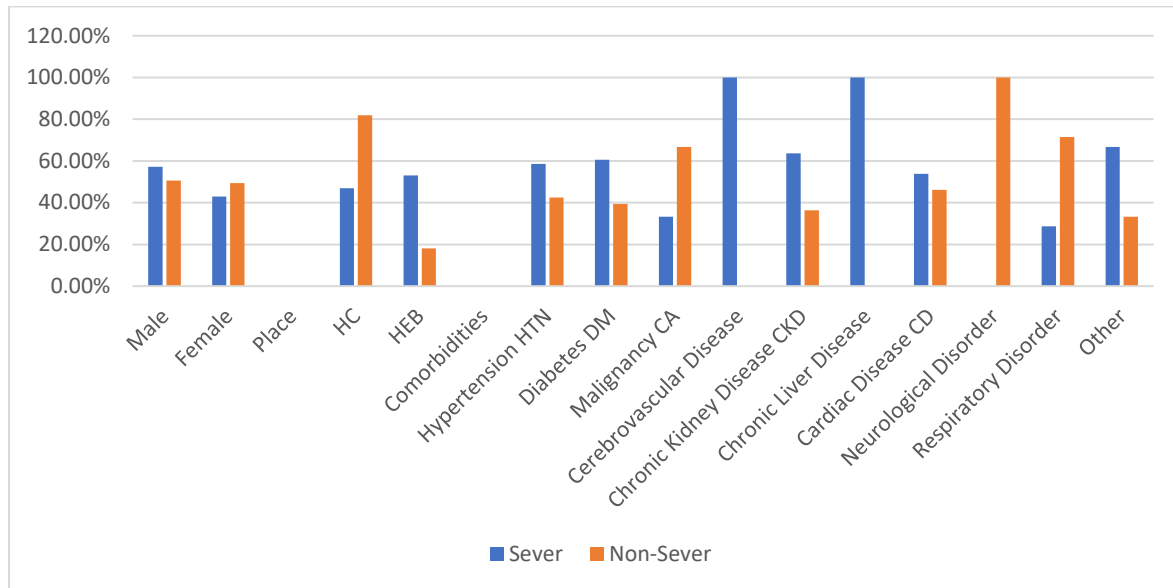
Before start analyzing the data, data was imported to SPSS version 26. Data cleaning was performed for outliers and missing values. Demographic data had no missing data. However, some clinical data were noted as missing in patients' files. So, missing values were replaced by mean that is only about 2 % of the sample.

From 199 patient files, a 132 Patient file was reviewed for descriptive analysis and visualization then 67 files were added to the sample for data mining that has been done using MATLAB.

### **5.3. Socio-demographic characteristics of the sample**

The study sample that was used for descriptive analysis and visualization included 132 COVID-19 hospitalized patients. Around (47%) of them were females. The median age was 45.5 years old (IQR, 31.0– 62.5). A total of (23.48%) of them were admitted and transferred to the ICU, and (37.12%) developed severe symptoms, where our study reported a high percentage of patients as severe cases (Figure 5). This could be related to the nature of patients who were asymptotic and mildly ill did not admit to hospitals and were not included in our study. However, lower patients were admitted to ICU (23.48%) than the percent that was recently reported globally(Wang et al., 2020).

Figure 5: Socio-demographic and comorbidities Characteristics of the Sample



Not all severe cases were admitted to the ICU because of the ICU capacity of the Palestinian healthcare system. The median length of stay in the hospital was about 8.5 days (IQR, 4.3 – 17.0). The length of stay of COVID-19 patients is varied from less than 14 days to nearly two months as published previously (Rees et al. 2020;)(Liu et al., 2020). Patients with COVID-19 in Palestine appeared to remain in the hospital for a shorter than elsewhere. This may be explained by differences in criteria for admission and discharge between countries, this result may be due to many matters specific to Palestine, such as the general healthcare situation in Palestine at the beginning of the pandemic and the lack of facilities that allow the presence of large numbers of patients in hospitals, and thus allowing patients to follow the last stages of their treatment in their homes. On the other hand, social stigma for COVID-19 patients especially at the beginning of pandemic may push many patients to follow their treatments at hospital in a late stage. Of the 132 patients, (12.2%) died during their hospitalization. Our case fatality rate findings in the real-world scenario might be lower when compared to the total infected COVID-19 patients not only admitted to the hospitals. Based on Palestinian Ministry of Health statistics the total number of confirmed

cases until July 20-2020, was 10520 cases where 65 cases were death, with (0.61%) case fatality. Most of our study sample (68.9%) were in north and middle Palestine while the rest were in the south (Table 4).

*Table 4: Study Population Characteristics*

|   |           | Median (IQR)       |
|---|-----------|--------------------|
| Age   |           | 45.5 (31.0 - 61.5) |
| Length of Stay  |           | 8.5 (4.3 - 17.0)   |
|   |           | <b>No. %</b>       |
| Sex   | Male      | 70 (53.0%)         |
|   | Female    | 62 (47.0%)         |
| Alive / Dead  | Dead      | 15 (12.2%)         |
|   | Alive     | 108 (87.8%)        |
| Location  | HC        | 91 (68.9%)         |
|   | HEB       | 41 (31.1%)         |
| Severity level  | Sever     | 49 (37.12%)        |
|   | Not Sever | 83 (62.88%)        |
| ICU / Non-ICU   | ICU       | 31 (23.48%)        |
|   | Non-ICU   | 101 (76.52%)       |
| <i>IQR: Interquartile Range / ICU: Intensive Care Unit/ HC: Hugo Chavez Hospital/ HEB: Hebron Governmental Hospital</i> |           |                    |

Patients who were at a severe level were significantly older compared with patients who were not (median age, 60 years (IQR, 45.0–66.0) vs. 41 years (IQR, 47.0 – 54.0);  $P < .001$ ). In general, the

risk of getting severe disease illness increases as getting older, as mentioned in several previous studies, the greatest risk of severe COVID-19 symptoms is among cases who are aged 80 years or older, taking into account the existence of other factors such as chronic diseases or comorbidities(Palmieri et al. 2020)(Gostic et al., 2020).

The median hospital length of stay for severe cases was 7.0 day (IQR,3.0 - 12.0) and significantly ( $P < .05$ ) lower than the median length of stay for not severe cases which is 10.0 days (IQR, 5.0 – 19.0), This may be because 15 cases out of 49 severe symptoms died within 5 days of hospitalization that ends the stay of severe cases in hospital (Table 5)

*Table 5: demographic Characteristics correlation with a severity level*

|                        | <b>Total (N =<br/>132)</b> | <b>No. (%)</b>             |                                | <b>P-Value<sup>1</sup></b> |
|------------------------|----------------------------|----------------------------|--------------------------------|----------------------------|
|                        |                            | <b>Severe<br/>(n = 49)</b> | <b>Non-Severe<br/>(n = 83)</b> |                            |
| Age, median (IQR), Y   | 45.5 (31.0 - 61.50)        | 60 (45 – 66)               | 41 (27 – 54)                   | 0.000*                     |
| L.o.S, median (IQR), D | 8.0 (4.0 - 17.0)           | 7 (3 – 12)                 | 10 (5 – 19)                    | 0.016*                     |
| <b>Sex</b>             |                            |                            |                                |                            |
| Male                   | 70 (53.0%)                 | 28 (57.14%)                | 42 (50.60%)                    | 0.467                      |
| Female                 | 62 (47.0%)                 | 21 (42.86%)                | 41 (49.40%)                    |                            |
| <b>Place</b>           |                            |                            |                                |                            |
| HC                     | 91 (68.9%)                 | 23 (46.9%)                 | 68 (81.9%)                     | 0.000                      |
| HEB                    | 41 (31.1%)                 | 26 (53.1%)                 | 15 (18.1%)                     |                            |

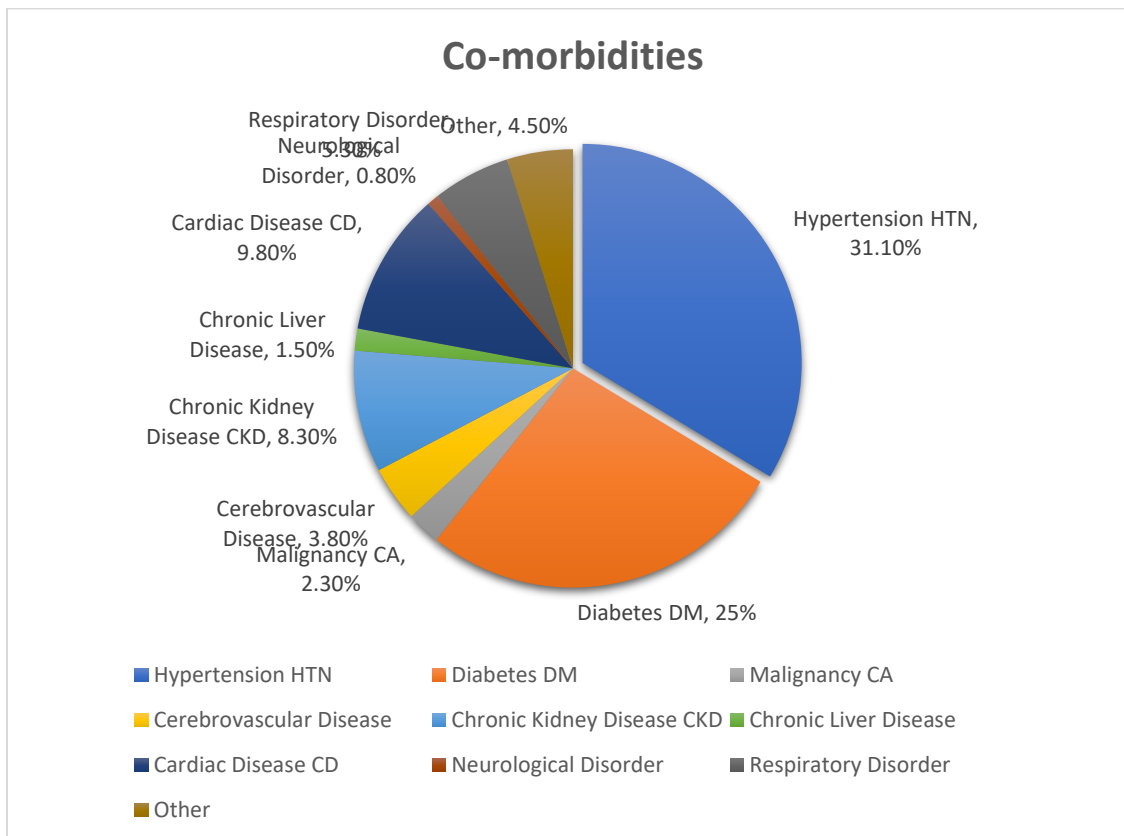
*IQR: Interquartile Range / ICU: Intensive Care Unit/ HC: Hugo Chavez Hospital/ HEB: Hebron*

*Governmental Hospital/ LoS: length of stay/ \*: significant relationship <0.05*

#### 5.4. Co-morbidities in the study sample

Of the 132 patients, (48.5%) had 1 or more preexisting medical conditions. Hypertension (31.1%), Diabetes (25%), Cardiac Disease (9.8%), and Chronic Kidney Disease (8.3%) were the most common preexisting comorbidities (Figure 6).

Figure 6: Sample Co-morbidities



When comparing patients who have severe symptom (n=49) with those of non-sever symptoms, (n=83), patient sever symptom were more likely to have preexisting medical condition, including hypertension (24 [58.50%] vs. 17 [41.50%]). Diabetes (20 [60.60%] vs 13 [39.40%]), and Cerebrovascular Diseases (5 [100.00%] vs. 0 [0.00%]) were statistically significant ( $P < .001$ ,  $P < .001$ ,  $P < .05$  respectively) in their association with severity level is shown in Table 6.

*Table 6: Co-morbidities correlation with a severity level*

| <b>COMORBIDITIES</b>                  | <b>Total</b> | <b>Severe</b> | <b>Non-Severe</b> | <b>P-value</b> |
|---------------------------------------|--------------|---------------|-------------------|----------------|
| <b>Hypertension HTN</b>               | 41 (31.1%)   | 24 (58.50%)   | 17 (41.50%)       | 0.001          |
| <b>Diabetes DM</b>                    | 33 (25.0%)   | 20 (60.60%)   | 13 (39.40%)       | 0.001          |
| <b>Malignancy CA</b>                  | 3 (2.3%)     | 1 (33.30%)    | 2 (66.70%)        | 0.891          |
| <b>Cerebrovascular Disease</b>        | 5 (3.8%)     | 5 (100.00%)   | 0 (0.00%)         | 0.003          |
| <b>Chronic Kidney Disease<br/>CKD</b> | 11 (8.3%)    | 7 (63.60%)    | 4 (36.40%)        | 0.057          |
| <b>Chronic Liver Disease</b>          | 2 (1.5%)     | 2 (100.00%)   | 0 (0.00%)         | 0.064          |
| <b>Cardiac Disease CD</b>             | 13(9.8%)     | 7 (53.80%)    | 6 (46.20%)        | 0.189          |
| <b>Neurological Disorder</b>          | 1 (0.8%)     | 0 (0.00%)     | 1 (100.00%)       | 0.441          |
| <b>Respiratory Disorder</b>           | 7 (5.3%)     | 2 (28.60%)    | 5 (71.40%)        | 0.630          |
| <b>Other</b>                          | 6 (4.5%)     | 4 (66.70%)    | 2 (33.30%)        | 0.125          |

Considering the presence of coexisting illness, (58.5%) of the hypertension patients developed a severe symptom with a statistically significant relationship between having hypertension as comorbidities and symptoms severity. Similar to other studies hypertension is one of the comorbidities that increase the risk of developing severe symptoms and worse outcomes (Schiffirin et al., 2020)(Fang et al., 2020). Diabetes is another important coexisting illness that also leads to a worse prognosis of admitted patients. (25%) of our population study have type2 diabetes, with a statistically significant relationship of developing severe symptoms, where (60.6%) of diabetic patients developed a severe illness. Our findings were more similar to Italian statistics(Apicella et al., 2020b). All of the cerebrovascular disease cases was developed a severe illness and admitted

to the ICU. Such an outcome could be attributed to cerebrovascular diseases that may lead to a problem in the coagulation system (Altable & de la Serna, n.d.). Our study also shows that half of the cardiac or cardiovascular disease patients develop some severe symptoms. WHO reported the underlying cardiovascular comorbidities is strongly associated with high mortality, conversely covid-19, in turn, lead to some cardiovascular problems like acute coronary syndrome, arrhythmias, and venous thromboembolism (Nishiga et al., 2020). By looking at the chronic kidney disease CKD, we found that (8.3%) of admitted patients were with coexisting CKD; (63.6%) of them developed a severe illness. Previous studies show that having CKD of any stage can increase the potential of developing severe symptoms, especially those with kidney transplants, and have an immunosuppressant treatment and hemodialysis(Ajaimy & Melamed, 2020)(Henry & Lippi, 2020). However, the most affected patients are those who had multiple coexisting illnesses, including hypertension, diabetes, and cardiovascular disease(Ajaimy & Melamed, 2020). Figure (7) represent a comparison between different comorbidities classified as severe (series 1) and non-severe (series 2)

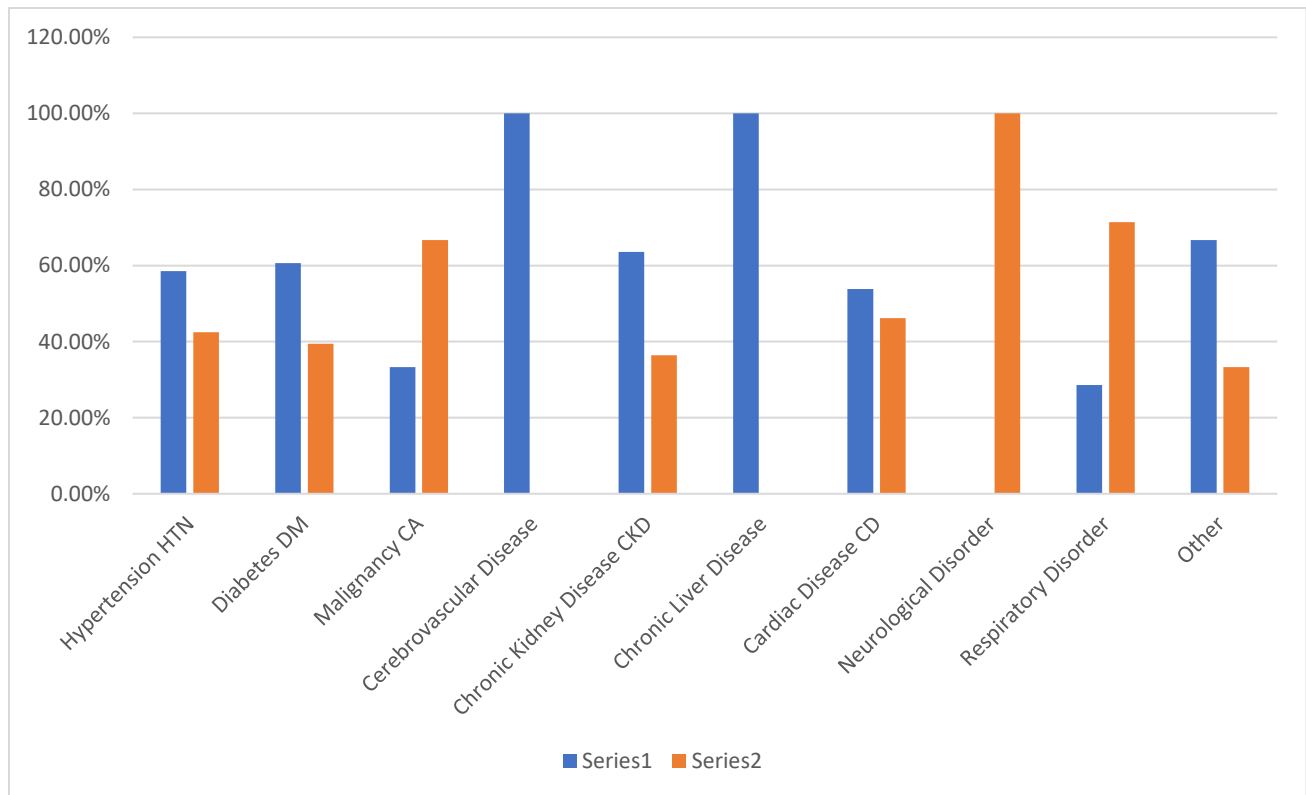


Figure 7: severity level comparison according Co-morbidities

## 5.5. Signs and complying

The most common patient's complying during the first week of admission were cough (51.5%), fever (41.7%), shortness of breath (25.8%), weakness (15.9%). Comparing with previous studies, similar symptoms were noticed in other countries (Carf`i et al., 2020) (Gostic et al., 2020). The European Surveillance System (TESSy) reports that the most common symptoms of COVID-19 in an observational study of 1420 covid19 cases were headache (70%), loss of smell (70.2%), cough (63.2%), myalgia (62%), sore throat (52%) and (45.4%) (European Surveillance System, n.d.). Less common symptoms on our sample were headache, diarrhea, hemoptysis, fatigue,

Dyspnea, nausea, chest pain, sore throat, chills, tachypnea, vomiting, and loss of appetite. Table (7) shows that cough, dyspnea, shortness of breath, fever, and loss of appetite were statistically significant in their association with severity level, as the percentage of COVID-19 patients who were at a severe level and suffered from these symptoms is significantly larger than the percentage of them whose status was not severe and suffered from these symptoms, while Figure (8), show a comparison between different complaining according to severity level.

*Table 7: Patients complaining relationship with a severity level*

|                                 | <b>Total (N = 132)</b> | <b>Relationship with a severity level</b> |                  |                            |
|---------------------------------|------------------------|---|------------------|----------------------------|
|                                 |                        | <b>Sever</b>                              | <b>Not Sever</b> | <b>P-Value<sup>1</sup></b> |
|                                 |                        | <b>No. (%)</b>                            |                  |                            |
| Cough                           | 68 (51.5%)             | 32 (65.31%)                               | 36 (43.37%)      | 0.015                      |
| Dyspnea                         | 10 (7.6%)              | 8 (16.33%)                                | 2 (2.41%)        | 0.004                      |
| Shortness of Breath<br>(SOB)    | 34 (25.8%)             | 29 (59.18%)                               | 5 (6.02%)        | 0.000                      |
| Fever                           | 55 (41.7%)             | 29 (59.18%)                               | 26 (31.33%)      | 0.002                      |
| Tachypnea                       | 2 (1.5%)               | 2 (4.08%)                                 | 0 (0.00%)        | 0.064                      |
| Loss of Appetite                | 3 (2.3%)               | 3 (6.12%)                                 | 0 (0.00%)        | 0.023                      |
| <i>SOB: Shortness of Breath</i> |                        |   |                  |                            |

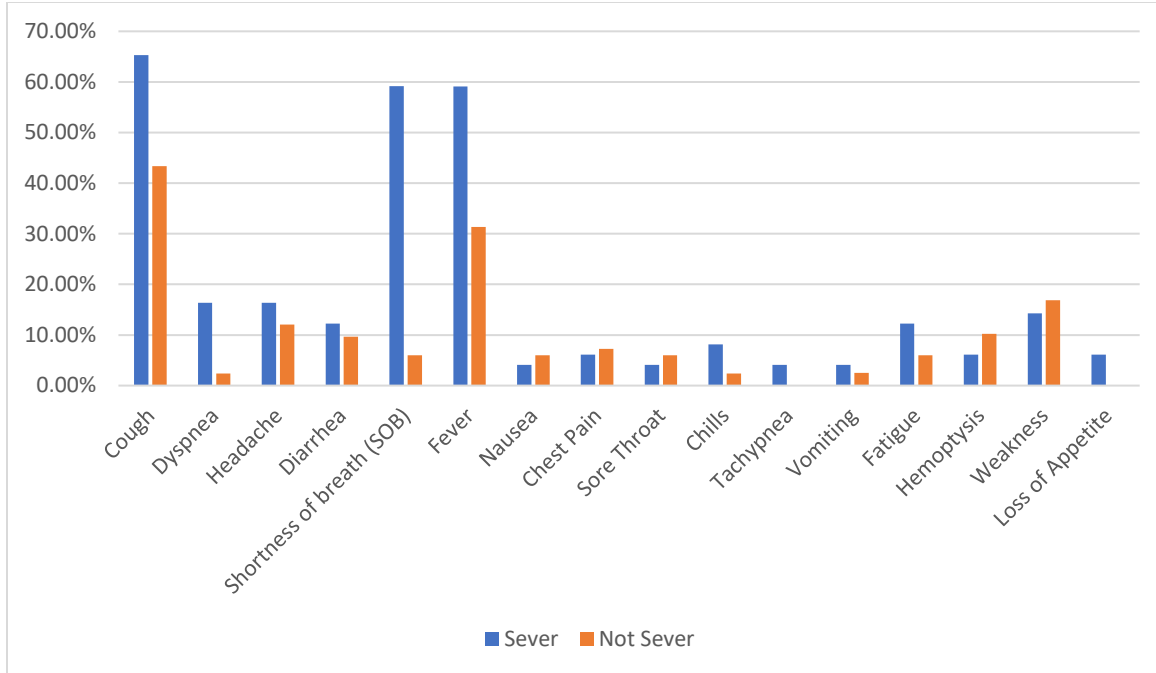


Figure 8: complaining according to severity level

## 5.6. Laboratory findings

According to the laboratory findings; numerous differences were reported between severe and non-severe cases as shown in table (8), including higher white blood cell and neutrophil counts, moreover higher level of Lactate dehydrogenase LDH, Ferritin, BUN, and Creatinine ( $P < .05$ ), and lower lymphocytes percentage, as well as lower saturation of arterial blood with oxygen ( $P < .001$ ). C - Reactive Protein – CRP tests was statistically significant in its association with severity level (P-value  $< 0.05$ ), as (26.67%) of patients who were at a severe level were Negative while (71.05%) of patients who were not at a severe level were Negative. Monocytes were statistically significant in their association with severity level (P-value  $< 0.05$ ), while PT and APTT were not.

Table 8: Lab Tests Findings of Patients Infected with COVID-19 on Admission to Hospital

|  |  | Relationship with a severity level |
|--|--|------------------------------------|
|  |  |                                    |

|  |          | Total                | Severe                | Non-Severe         | P-Value <sup>1</sup> |
|--|----------|----------------------|-----------------------|--------------------|----------------------|
|  |          | Median (IQR)         |                       |                    |                      |
| WBC  |          | 7.8 (6.2 - 10.3)     | 9.25 (5.65 - 13.75)   | 7.4 (6.2 - 9)      | 0.007                |
| Neutrophil granulocyte percentage  |          | 72.0 (62.5 - 83.1)   | 82.05 (70.4 - 88.05)  | 68.2 (56.8 - 77)   | 0.000                |
| Lymphocytes percentage   |          | 22.8 (14.2 - 34.0)   | 16.4 (9.2 - 23.4)     | 26.9 (18.8 - 37.5) | 0.000                |
| HGB  |          | 13.2 (11.9 - 14.6)   | 12.25 (10.8 - 14.35)  | 13.4 (12.2 - 14.9) | 0.038                |
| Platelets count  |          | 223 (177 - 279)      | 212.5 (140.5 - 276.5) | 224 (182 - 279)    | 0.283                |
| BUN  |          | 14.2 (10.0 - 23.1)   | 18.3 (14.2 - 36.7)    | 12.75 (9.9 - 17)   | 0.001                |
| Creatinine   |          | 0.8 (0.6 - 1.1)      | 0.97 (0.7 - 1.78)     | 0.79 (0.59 - 0.92) | 0.034                |
| GPT  |          | 22.6 (15.0 - 33.9)   | 21.8 (15.6 - 33.8)    | 24 (14.4 - 33.9)   | 0.760                |
| GOT  |          | 23.6 (18.0 - 35.1)   | 26.7 (20 - 42.6)      | 23 (16.1 - 34.1)   | 0.267                |
| D - Dimer Gen.2 (D-DI)   |          | 258 (0.0 - 536.0)    | 370.5 (44 - 789)      | 199 (0 - 450)      | 0.070                |
| Lactate dehydrogenase LDH*   |          | 211 (173 - 305)      | 276.5 (186 - 423.5)   | 203 (165 - 284)    | 0.012                |
| Ferritin*  |          | 188.3 (64.8 - 426.0) | 292.5 (121 - 777)     | 131 (54.5 - 267)   | 0.000                |
| C - Reactive Protein<br>- CRP  | Negative | 66 (54.5%)           | 12 (26.67%)           | 54 (71.05%)        | 0.000                |
|  | Positive | 55 (45.5%)           | 33 (73.33%)           | 22 (28.95%)        |                      |
| Monocytes  |          |                      | 0.65 (0 - 3.7)        | 0 (0 - 0)          | 0.001                |
| PT*  |          |                      | 13.2 (12.25 - 14.1)   | 12.9 (12 - 15)     | 0.995                |
| APTT*  |          |                      | 31.5 (30 - 37.9)      | 32 (29.4 - 37.3)   | 0.946                |
| Troponin   | Negative | 9 (7.3%)             | 3 (6.25%)             | 6 (7.89%)          | 0.731                |
|  | Positive | 115 (92.7%)          | 45 (93.75%)           | 70 (92.11%)        |                      |
| SPO2, a saturation of arterial blood with oxygen. IQR, interquartile range. BUN, blood urea nitrogen. *Mann-Whitney U was used instead of an independent t-test due to the smaller number of cases used. |          |                      |                       |                    |                      |

When measuring the median of Ferritin level of all COVID-19 patients, it was 188.3 (IQR, 64.8 - 426.0). However, for those who had diabetes the median Ferritin level was 248.50 (IQR, 129.0-731.68) and significantly larger than the median of Ferritin level 151.85 (IQR, 57.40 - 291.24) for those who didn't have diabetes (P-value < 0.05). Having diabetes is significantly associated with severity level, as (60.6%) of COVID-19 patients who had diabetes were at a severe level while (29.3%) of COVID-19 patients who didn't have diabetes were at a severe level (Table 9). That ferritin is a key mediator of immune dysregulation, especially in diabetic patients under extreme hyperferritinemia, via direct immune-suppressive and pro-inflammatory effects, contributing to the cytokine storm (Rismanbaf & Zarei, 2020).

*Table 9: Diabetes relationships with Ferritin and Severity*

| Relationship with Diabetes-DM |           |                      |                       |                         |                            |
|-------------------------------|-----------|----------------------|-----------------------|-------------------------|----------------------------|
|                               |           | <b>Total</b>         | <b>diabetes</b>       | <b>No diabetes</b>      | <b>P-Value<sup>1</sup></b> |
|                               |           | <b>Median (IQR)</b>  |                       |                         |                            |
| Ferritin                      |           | 188.3 (64.8 - 426.0) | 248.50 (129.0-731.68) | 151.85 (57.40 - 291.24) | 0.022                      |
|                               |           | <b>No. (%)</b>       |                       |                         |                            |
| Severity                      | Sever     | 49 (37.12%)          | 20 (60.6%)            | 29 (29.3%)              | 0.001                      |
|                               | Not Sever | 83 (62.88%)          | 13 (39.4%)            | 70 (70.7%)              |                            |

By comparing the COVID-19 impact on kidney function, through testing the biochemistry of kidneys (Blood urea nitrogen (BUN) and Creatinine), our targeted study population showed that there is a significant increase in both tests in favor of patients with severe illness. That was consistent with a previous retrospective analysis of 100 patients with covid-19 which showed that (35%) of patients within a duration from 2 to 4 days develop an increased level of both BUN and Creatinine (Mahmoudi et al., 2020), another study expose that SARS-CoV-2 effects on the kidney and in patient with underlying kidney disease is not well-characterized. Preliminary data has indicated that previous kidney disease could represent a risk factor, especially in elderly patients, for a more severe disease course. SARS-CoV-2 infects the kidneys and may induce acute kidney injury. While there is no current specific therapy, many drugs both antiviral and/or anti-inflammatory are being actively tested in randomized trials. Further studies are necessary to better understand disease pathology, acute kidney injury associated with infection, long-term renal consequences, and potential therapies (Rismanbaf & Zarei, 2020). However, for a liver test, there was no significant change in liver function tests between patients with severe and non-severe illness, the GPT or GOT (also known as AST and ALT) are the liver blood tests, relevant studies of these tests confirmed covid-19 patients expose the relation between abnormal liver function tests and risk of developing the severe disease (Cai et al., 2020). Possible explanations for that difference; our data were collected for patients during the first two days from admission. Relevant studies suggested that the lever injury related to some medication used in the disease treatments, that could be very early for appeared in our sample (Cai et al., 2020)(Fan et al., 2020).

Our study showed that there is no significant relationship between D-dimmer and severity level, however, a relevant study of 248 patients with covid-19 in Wuhan University, China, confirmed that it correlates with disease severity and can be used as a reliable prognostic marker for COVID-

19 patients(Yao et al., 2020), that D-dimer elevation has been observed in describing the clinical features of COVID-19, whether the level of D-dimer is a marker of severity has not been examined. C-reactive protein level was positively correlated with disease severity, where (73.33%) found to be positive in patients of a severe group in comparison to (28.95%) positive CRP in a non-severe group. An observation study was done into 22 confirmed COVID-19 patients aims to assess the usefulness of CRP levels for evaluating COVID-19 severity, found that there is a direct proportion relation between CRP serum level and degree of severity, that a high C-reactive protein (CRP) values are frequently found in patients with bacterial respiratory infection, and CRP testing has been shown to be useful in differentiating pneumonia from other respiratory infections. Raised CRP values may also be found in viral respiratory infection that may increase the severity of the disease (Ling, 2020). Some early observations studies consider Monocytes as an important player in the progression of severe symptoms of COVID-19 patients(Pence, 2020). Our study also came to confirm the same result, that of the severe group patients, Monocyte's level was found to be higher than patients in a non-severe group. Moreover, a strong significant relationship was found between ferritin and severity level of disease, being serum Ferritin median level (292.5) in patients with severe illness in comparison with (131) in non-severe cases. In agreement with this suggestion, another study confirmed the same relation(Terpos et al., 2020)(Shoenfeld, 2020), concluded that Ferritin levels were strongly related to the severity of COVID-19. Despite a recent study in 69 patients confirms that 20% of patients develop thrombocytopenia (low levels of platelets) and severe symptoms, our sample showed no significant difference between patients with severe symptoms and the non-severe group (Terpos et al., 2020).

### 5.7. Treatments provided to COVID-19 patients

The most common treatments given for COVID-19 patients were Antipyretic 99 (75.0%), Supplementary drugs 73 (55.3%), Anticoagulant 62 (47.0%), Antibiotic 61 (46.2%), IV fluid 45 (34.1%), Glucocorticosteroids 42 (31.8%). The Less common treatments were Anti asthmatic drugs 33 (25.0%), O2 Therapy 35 (26.5%), Antihypertensive drugs 28 (21.2%), Anti-cough drugs 22(16.7%), Hydroxychloroquine anti rheumatic drug 22 (16.7%), Proton-pump inhibitors (PPI) drugs 14 (10.6%), Antihistamine drugs 6 (4.5%), Antiviral 4 (3.0%) as shown in figure (9).

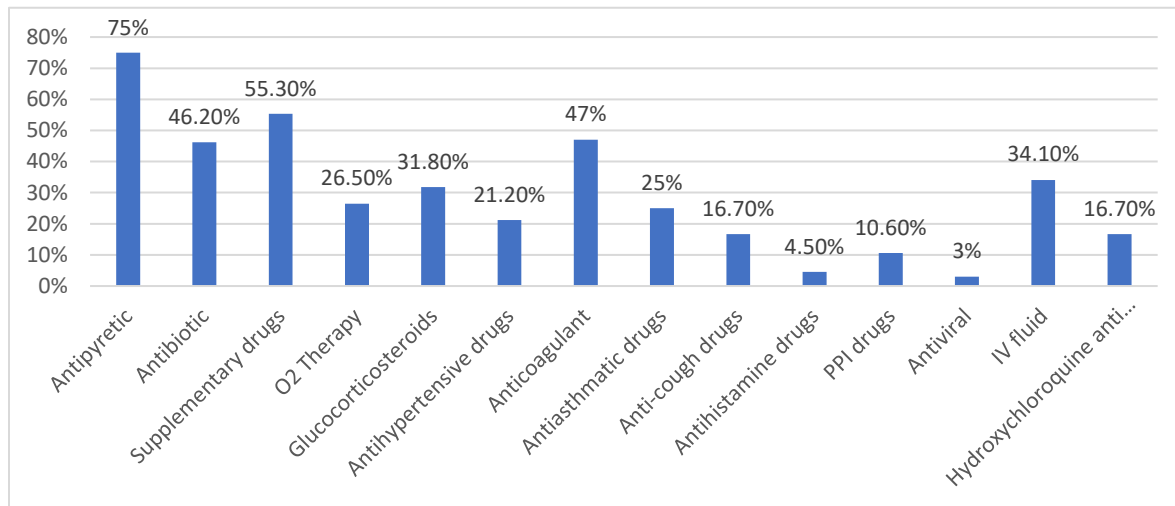


Figure 9: Treatments provided to COVID-19 patients

Scientists and researchers around the world are making tremendous efforts to develop an effective treatment of COVID-19. Till now there is no specific treatment for the disease, with many drugs are under investigation, physicians on their part follow international protocols provided by WHO, and giving some medication to reduce the exacerbation of the disease and relief its complications. Where (75.0%) of our sample receive an antipyretic drug, that might be related to the fever which is one of the most common symptoms of our sample, as antipyretics like acetaminophen and

ibuprofen are often taken to reduce the discomfort of fever (Jamerson & Haryadi, 2020). Although many reports indicate the positive effects of some antibiotics (like azithromycin) in relief complications of covid-19 patients, there have also been severe side effects, as our study showed that nearly (46.2%) of the patients took antibiotics during their hospitalization stay, despite Antibiotics is specified for treating bacterial infections with no effect in virus's infection cases (Beović et al., 2020). (55.3%) of patients included in our study received supplementary drugs, that according to recent reports, dietary supplements and vitamins D and C are very effective in helping the immune system resist the covid-19 infection(Grant et al., 2020). Previous literature analysis study proposed that oxygen therapy can inhibit virus replication and play an important role in autoimmunity regulation, physicians in these study used O2 therapy for (26.5%) of patients (Kayem et al., 2020)(Shen et al., 2020) (31.8%) of patients were given exogenous glucocorticoids (GCs). However, their use has been discouraged because linked with immune suppression (Isidori et al., 2020). Covid-19 is considered as pro-coagulability for unclear mechanisms until now, which could be a life threatening and increase the possibility of bleeding, regarding D-dimmer levels, physicians used anticoagulant for (47.0%) of patients. (Rico-Mesa et al., 2020) Some studies reviewed the relationships between anti-cough drugs and covid-19. Researchers found there is a direct relationship between the growth of the virus and the use of this drug, and also found that coughing may help relieve congestion in the lungs, and this explains the use of this drugs are relatively little, as our study indicates that about (16.7%) of the patients received this drug, especially at the beginning of its spread, as information about the disease was still vague.

Also, by browsing the patients' files, in addition to the aforementioned drugs, some additional treatment plans were recorded which were used at lower rates. It was found that about (25.0%) of

patients received an Antiasthmatic drug, (16.7%) Hydroxychloroquine anti-rheumatic drug, (10.6%) PPI drugs, and (3.0%) Antiviral drugs.

## **5.8. Prediction models**

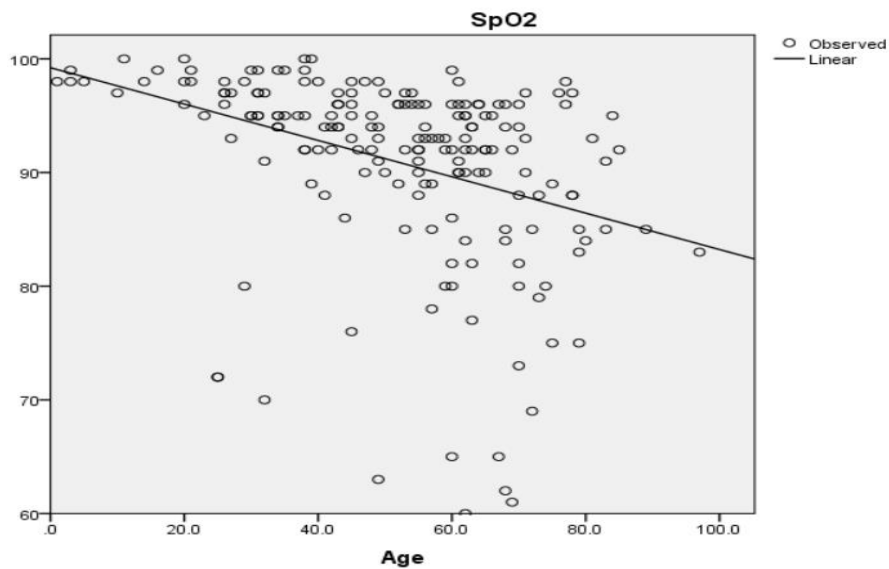
The global threat to living society has been brought by a catastrophic outbreak of Coronavirus infection known as COVID 19. Therefore, globally, incredible efforts are needed to fight the rapid spread of COVID19 in investment, infrastructure, data source, protective tools, treatment of life-risk, and many other resources. This made a challenge to the artificial intelligence researchers to concentrate their special experience in developing mathematical models to analyze the nature of the COVID-19 epidemic using a different dataset. So that to have a role towards the health of living society, our goal was to utilize prediction model that aims to understand the future of COVID-19 across the world by using real data from COVID-19 Palestinian hospitals.

## **5.9. Linear regression**

Regression analysis was done to assess and predict factors that may affect the severity of COVID-19 symptoms. Because patient-reported oxygen saturation (SPO2) is considered the first indicator of symptom severity, many studies evaluate the ability to use (SpO2) to determine the need for a COVID-19 patient hospitalization and used it as an effective indicator of the severity level and disease progression. (Teo et. al 2020). Decreased SPO2 level was independently associated with hospitalization mortality (Teo, 2020)(Xie et al., 2020). It appears through the previous analysis of the datasets that the level of SPO2 is one of the most important factors to determine the disease behavior and its severity level.

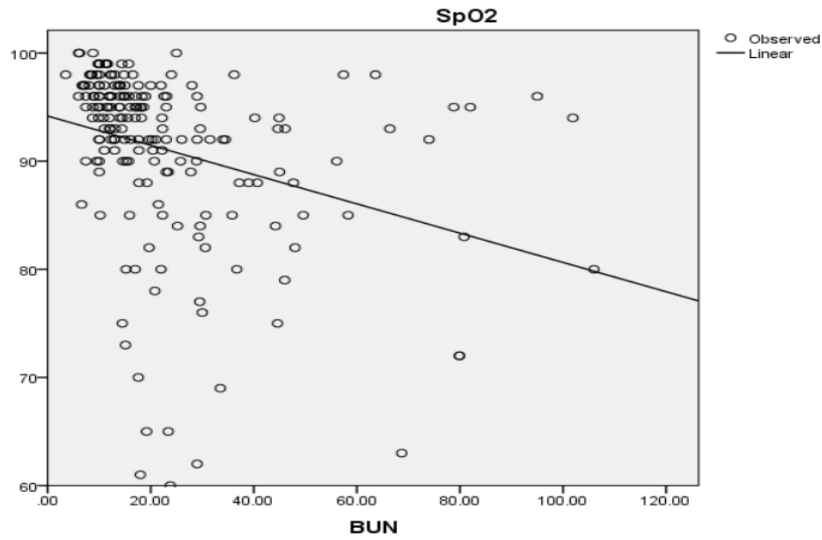
We used the generalized linear model (GLMs) to determine the quantitative features correlation with SPO2 level, considered COVID-19 infection as a hidden variable that can affect the relation, in order to clarify and confirm on indirect way the relationship of these variables with the disease symptoms severity, as all the results of the previous statistical analysis has been confirmed with the GLMs results, in this section we are going to present the most important results of the model.

In the first (GLMs), age (P-Value 0.001) was significantly correlated with symptom severity and a negative correlation with SPO2 level (figure 10), as this result came with compatibility with our previous results of patients who were at a severe level were significantly older compared with patients who were not.



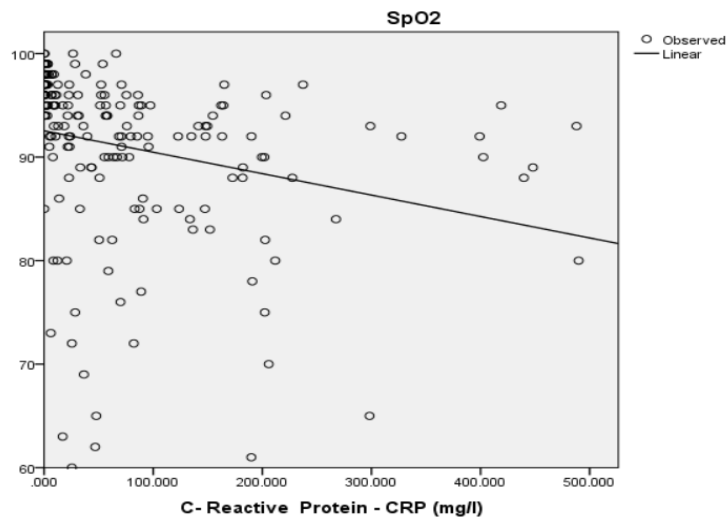
*Figure 10: linear regression correlating Age with SpO2*

Second (GLMs), the blood urea nitrogen (BUN) shows a negative correlation with SPO2 level (P-Value 0.001), and as missioned earlier our targeted study population showed that there is a significant increase in BUN level in favor of patients with severe illness Figure (11). That SPO2 is inversely proportional to disease severity level, so that when the patient-reported oxygen saturation is increase, the severity of the disease and symptoms decrease.



*Figure 11:: Linear regression correlating BUN with SPO2*

C-reactive protein (CRP) level was positively correlated with disease severity, then negatively correlated with SPO2 level P-value 0.001 as shown in (Figure 12), as raised CRP values may also be found in viral respiratory infection that may increase the severity of the disease (Ling, 2020).



*Figure 12: Linear regression correlating CRP with SPO2*

GLMs for Lactate Dehydrogenase (LDH) level show a negative correlation with SPO2 (P-Value 0.001), as previous analysis confirms the relation between higher level of LDH and disease severity level of worse outcome (figure 13)

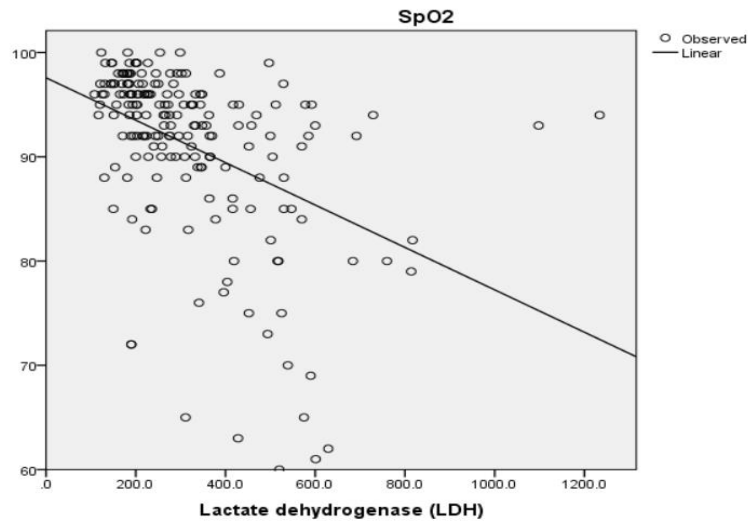


Figure 13: Linear regression correlating LDH with SPO2

On other hand, a strong significant relationship was found between ferritin and severity level of disease, GLMs for serum ferritin level show a negative correlation with SPO2 (P-value 0.001), (figure 14)

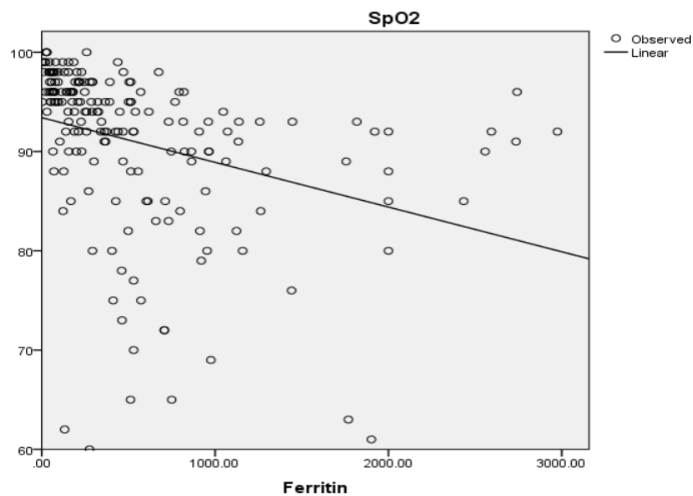


Figure 14: Linear regression correlating Ferritin with SPO2

Patient's age, BUN, CRP, LDH, and Ferritin level were significantly correlated with SPO2 level and by combining all these variables in a simple linear regression model proved a strong relationship with the SPO2 ( $R = 0.581$ , adjusted  $R$  square = 0.337), (P-value 0.001).

This result can be used by medical providers in COVID-19 centers to improve patient health status for the patients by maintaining their SPO2 level in the normal range. These consequences may guide the clinical management of patients with severe COVID-19 symptoms, mainly in settings that require s strategic allocation of restricted critical care resources.

### 5.10. Data mining techniques

In this section, we applied different machine learning methods; logistic regression, support vector machine, decision tree, and multilayer perceptron (neural network) to classify cases with severe and non-severe symptoms of COVID-19 infected patients. The dataset used in this part of the experimental consists of the demographic characteristics of the infected persons, past medical history, laboratory findings, and patients complaining, with a total sample of 199 records that consist of 20 features input and one output "severe and non -severe". After reviewing many previous studies and a comprehensive discussion with COVID-19 treatment physician experts we were advised to use these 20 features in utilizing prediction model table (10).

*Table 10: features in utilizing prediction model*

|   | <b>Variables</b>          | <b>Discretion</b>   |
|---|---------------------------|---|
| L | WBC                       | <i>White blood cells also called leukocytes or leucocytes</i>   |
|   | Neutrophil's granulocyte% | <i>neutrocytes or heterophils are the most abundant type of granulocytes and makeup 40% to 70% of all white blood cells inhumans.</i> |
|   | Lymphocytes %             | <i>white blood cells are also one of the body's main types of immunecells</i>   |

|                  |                            |  |
|------------------|----------------------------|--|
| Laboratory Tests | Platelet's count           | <i>A platelet count is a lab test to measure how many platelets you have in your blood. Platelets are parts of the blood that helps the blood clot. They are smaller than red or white blood cells.</i>  |
|                  | BUN                        | <i>Blood urea nitrogen is a medical test that measures the amount of urea nitrogen found in the blood.</i>   |
|                  | Creatinine                 | <i>is a waste product produced by muscles from the breakdown of a compound called creatine.</i>  |
|                  | GPT                        | <i>(also known as AST and ALT) liver blood tests are often slightly elevated, usually less than twice normal. Ultrasound of the abdomen usually shows the liver is "echogenic," which is more dense than usual toward sound waves. The liver can also be enlarged due to fat.</i>  |
|                  | GOT                        |  |
|                  | D - Dimer Gen.2 (D-DI)     | <i>is a fibrin degradation product (or FDP), a small protein fragment present in the blood after a blood clot is degraded by fibrinolysis.</i>   |
|                  | LDH                        | <i>The lactate dehydrogenase (LDH) test looks for signs of damage to the body's tissues.</i>   |
|                  | Ferritin                   | <i>a blood protein that contains iron. A ferritin test helps your doctor understand how much iron your body stores.</i>  |
|                  | PT                         | <i>The PTT assesses the amount and the function of certain proteins in the blood called coagulation or clotting factors... The partial thromboplastin time (PTT; also known as activated partial thromboplastin time (aPTT)) is a screening test that helps evaluate a person's ability to appropriately form blood clots.</i> |
|                  | APTT                       |  |
|                  | Troponin                   | <i>Troponin, or the troponin complex, is a complex of three regulatory proteins that are integral to muscle contraction in skeletal muscle and cardiac muscle, but not smooth muscle.</i>  |
|                  | C - Reactive Protein - CRP | <i>measures the level of c-reactive protein (CRP) in your blood. CRP is a protein made by your liver. It's sent into your bloodstream in response to inflammation.</i>   |
| Observation      | SpO2                       | <i>Oxygen saturation (SpO2) is a measurement of how much oxygen your blood is carrying as a percentage of the maximum it could carry.</i>  |
| Complaining      | SOB                        | <i>Shortness of breath — known medically as dyspnea — is often described as an intense tightening in the chest, air hunger, difficulty breathing, breathlessness, or a feeling of suffocation.</i>   |
| Demographic      | Age                        |  |
|                  | Sex                        |  |
| Co-morbidities   | Past Medical History       | <i>including major illnesses, any previous surgery/operations (sometimes distinguished as past surgical history or PSH), any current ongoing illness</i>   |

### 5.10.1. Logistic regression (LR)

The Area Under the Curve or (AUC) is commonly used to estimate the ability of the function to classify the class of an object. Figure (15) shows that Logistic Regression (LR) can classify up to 74% of the objects' classes. On the other hand, Figure (16) shows the produced confusion matrix showed that 84% of the objects were classified as true positive and 16% were false positive, while 58% of the negative objects were classified truly and 42% of them were classified falsely. So, sensitivity and specificity of LR were calculated using previously mentioned equation and found to be 66% LR sensitivity while specificity was 78% which finally means that LR is better in predicting negative cases (non-severe). The prediction accuracy of Logistic Regression (LR) as calculated in the aforementioned formula is 71%. Comparing study result with previous studies was impossible as COVID-19 pandemic was recent and none of up-to-date COVID-19 machine learning research were focused on prediction of COVID-19 symptom severity.

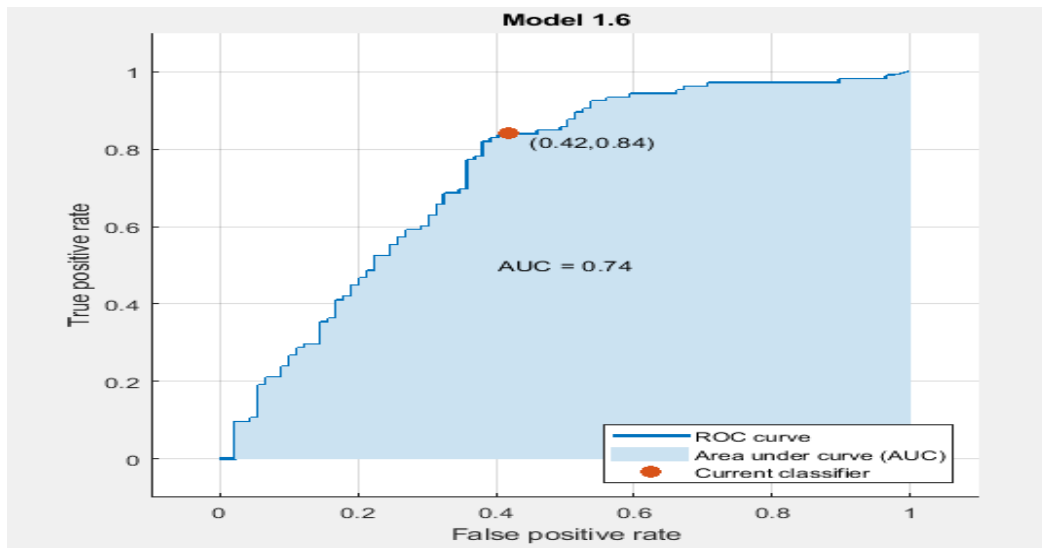


Figure 15: Logistic regression Area Under Curve

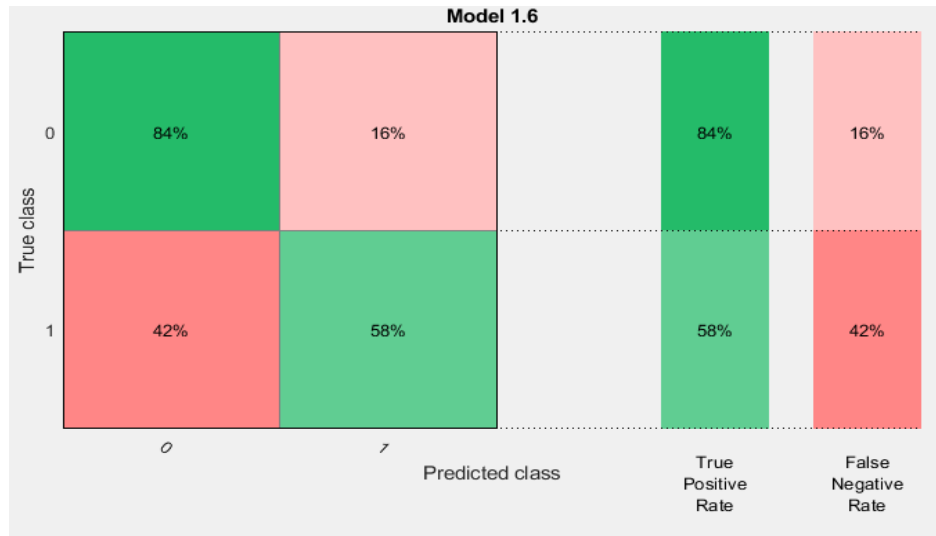


Figure 16: Logistic regression Confusion Matrix

### 5.10.2. Decision Tree (DT)

DT model can classify up to 68% of the objects' classes, according to the Area Under the Curve (figure 17), where (figure 18) the produced confusion matrix showed that 81% of the objects were classified as true positive and 19% were false positive, while 62% of the negative objects were classified truly and 19% of them were classified falsely. Therefore, the sensitivity and specificity of DT as calculated on the aforementioned equations were (68%), (76%) respectively. Concluded that DT's ability to predict negative cases is better than predicting positive ones, and the prediction accuracy of the Decision Tree (DT) is 71.5%.

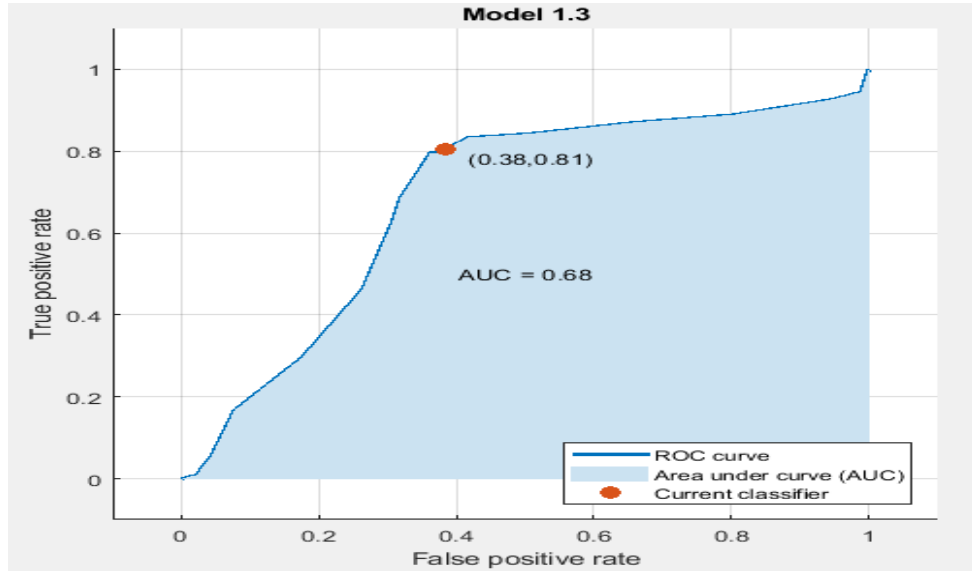


Figure 17: Decision Tree Area Under Curve



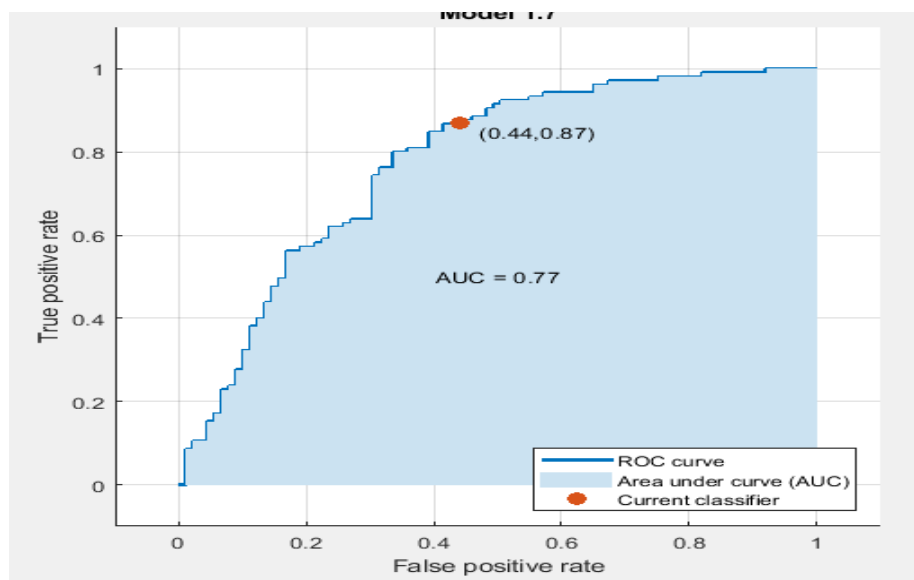
Figure 18: Decision Tree Confusion Matrix

### 5.10.3. Support vector machine (SVM)

By implementing the SVM model, we were able to classify up to 77% of the objects' classes according to the AUC (figure 19). Figure 20 shows the produced confusion matrix showed that

87% of the objects were classified as true positive and 13% were false positive, while 56% of the negative objects were classified truly and 44% of them were classified falsely.

The sensitivity and specificity of SVM were (66%), (81%) respectively. That findings concluded that SMV's ability to predict negative cases is better than predicting positive ones, and the prediction accuracy of Support Vector Machine (SVM) was 71.5%.



*Figure 19:Support victor machine Area Under Curve*

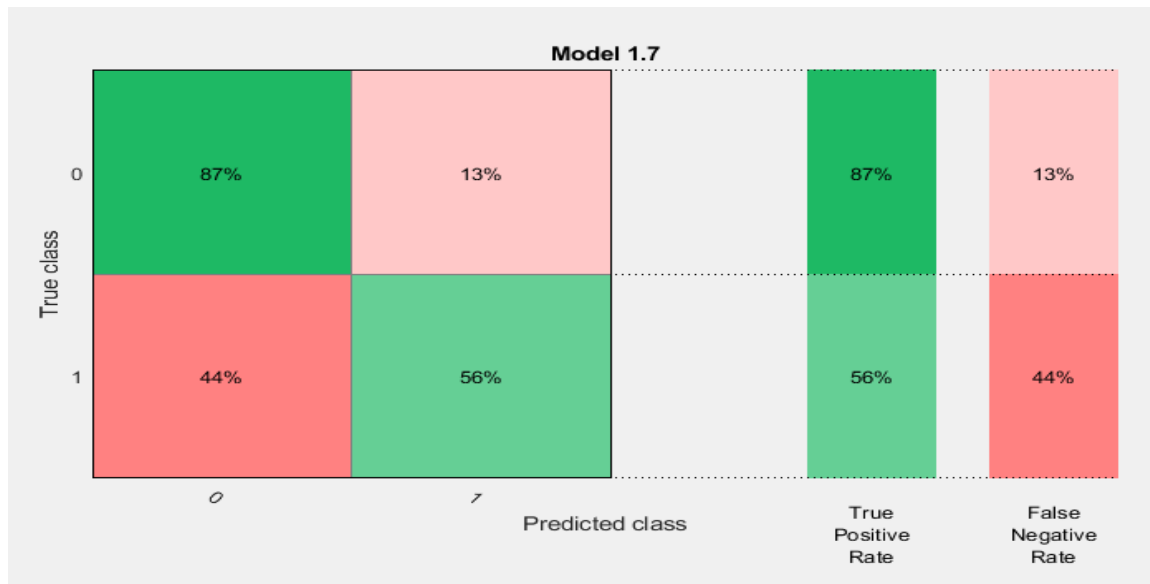


Figure 20: Support vector machine Confusion Matrix

#### 5.10.4. Artificial neural networks (ANNs)

Artificial neural network is mathematical method that is based on simple mathematical models. which allow complex nonlinear relationships between the variable and its predictors, ANNs model can classify up to 81% of the objects' classes, according to the Area Under the Curve figure (21), figure (22) shows the produced confusion matrix showed that 72% of the objects were classified as true positive and 28% were false positive, while 78% of the negative objects were classified truly and 22% of them were classified falsely. So, sensitivity and specificity of ANN's were calculated using previously mentioned equation and found to be 84.3% ANN's sensitivity while specificity was 61.5% which finally means that ANN's is better in predicting positive cases. The prediction accuracy of ANN's as calculated in the aforementioned formula is 74%.

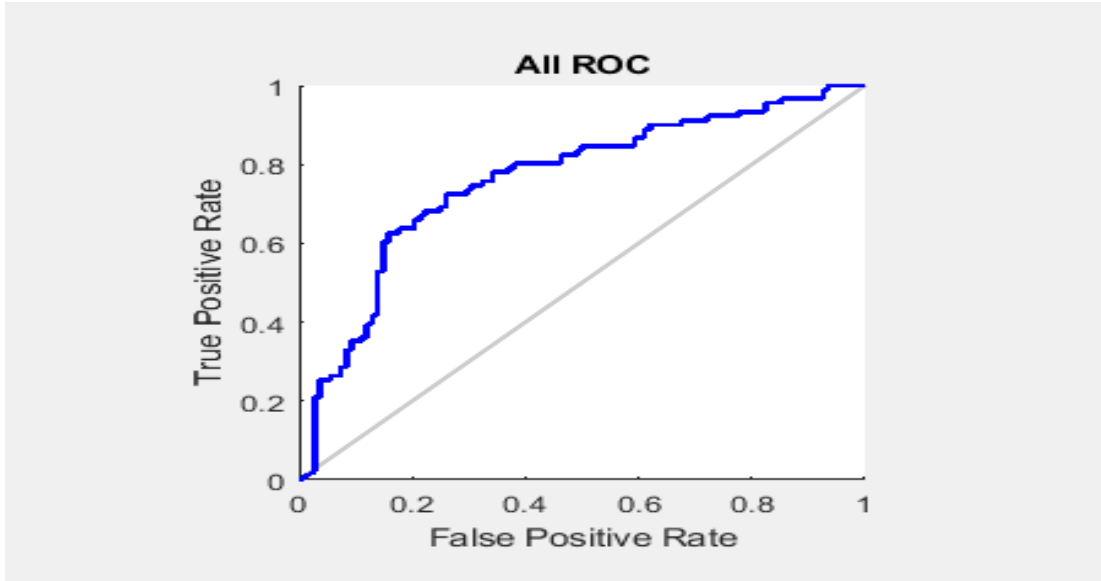


Figure 21: Artificial neural network Area Under Curve

**All Confusion Matrix**

|              |   |                       |                       |                       |
|--------------|---|-----------------------|-----------------------|-----------------------|
| Output Class | 0 | <b>91</b><br>45.7%    | <b>35</b><br>17.6%    | <b>72.2%</b><br>27.8% |
|              | 1 | <b>17</b><br>8.5%     | <b>56</b><br>28.1%    | <b>76.7%</b><br>23.3% |
|              |   | <b>84.3%</b><br>15.7% | <b>61.5%</b><br>38.5% | <b>73.9%</b><br>26.1% |
|              |   | 0                     | 1                     |                       |
|              |   | <b>Target Class</b>   |                       |                       |

Figure 22: Artificial neural network Confusion Matrix

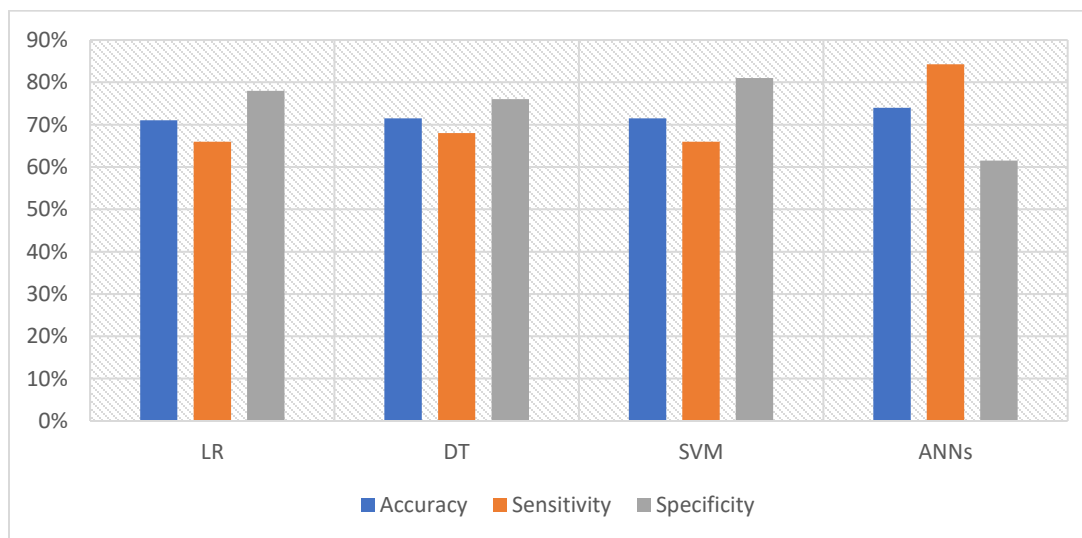
### 5.10.5. Model comparison

After applying four machine learning methods; logistic regression, support vector machine, decision tree, and multilayer perceptron (neural network) to classify cases with severe and non-sever symptoms of COVID-19 infected patients, we found that they are close to each other in terms of accuracy, sensitivity, specificity and the ability to identify cases with some slight discrepancies as shown in Table 11.

*Table 11: Data mining model evaluation*

| ML method | Accuracy | Sensitivity | Specificity |
|-----------|----------|-------------|-------------|
| LR        | 71%      | 66%         | 78%         |
| DT        | 71.50%   | 68%         | 76%         |
| SVM       | 71.50%   | 66%         | 81%         |
| ANNs      | 74%      | 84.30%      | 61.50%      |

ANNs are slightly better than other models in terms of their ability to identify positive cases that may develop a severe symptom in the future, with the best accuracy of 74% as shown in Figure (23). Comparison with previous studies was not possible as COVID-19 pandemic was recent and we were unable to find published studies focused on prediction of COVID-19 symptom severity.



*Figure 23: model evaluation comparison*

## 5.11. Summary

This chapter presents the results of the study done in two sections first by applying a conventional statistic of SPSS to expose the correlation of demographic characteristics of the sample, comorbidities, and laboratory tests with the severity level of the disease, second we use the data mining tools, applied a different machine learning data analysis tools to develop a prediction model.

#

## 6. Chapter 6: Conclusion and Recommendation

---

### 6.1. Conclusion

Palestine is one of the countries that has been affected by the COVID-19 pandemic. A retrospective review of 199 patient's medical records was conducted in this research study aims to understand and improve the clinical knowledge of COVID-19 patients in Palestine using conventional statistical analysis and data mining tools to develop model can classify and predict patient future health situation.

Around (47%) of our sample were females. The median age was 45.5 years old. A total of (23.48%) of them were admitted and transferred to the ICU, and (37.12%) developed severe symptoms, where (12.2%) died during their hospitalization. Having comorbidity of Hypertension HTN, Diabetes DM, Malignancy CA, and Cerebrovascular Disease will increase the chance of developing severe COVID-19 symptoms, as it considers as a risk factor of worse disease outcomes. The most common patients complying during the first two days of admission were cough (51.5%), fever (41.7%), shortness of breath (25.8%), weakness (15.9%). Laboratory tests of WBC, Neutrophil's granulocyte percentage, Lymphocytes percentage, Hgb, Platelets count, Monocytes, kidney function tests, LDH, Ferritin, and C - Reactive Protein – CRP, were significantly correlated with developing severe symptoms for COVID-19.

Although statistical models have been widely used for modeling the COVID-19-epidemiological characteristics, we also used the machine learning tools as a new trend in determining the severity level. Four popular machine learning tools were used to develop a prediction model for COVID-19 severity level; Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Networks (ANNs) with accuracy of each model that reflect how correct is it

to identifies the severity level are (71%, 71,5%, 71,5%, and 74%) respectively. However, these results laying down the practical foundations for developing model capable of classify the severity level of the disease with better accuracy would help the Palestinian health system to understand the nature and behavior of this disease, that will supports policymakers to effectively tackle the epidemic situation, better resource allocations and prepare control plans and needs of the healthcare system to save lives.

The medical staff must emphasize doing all laboratory tests, and carefully take notes on the patients' medical history, which are essential in understanding the patient's condition and determining the appropriate treatment. Also, special attention must be given to patients with comorbidities of hypertension, diabetes, malignancy, and Cerebrovascular Disease as they are the most fortunate in the deterioration of their health condition.

## **6.2. The implication to public health, research, and health informatics**

A key implication of this study is that developing mathematical models could assist in analyzing the COVID-19 symptom severity helps control measures including risk classification which needed to fight the rapid spread of COVID-19 in terms of infrastructure, protective tools, infection control policies, treatment of life-risk, and many other resources.

Such research could also help identify patient at risk of developing severe symptoms in those infected with COVID-19 and help in rapid and efficient triage of COVID-19 patient that provide clinical decisions, recommendations, and practice guidelines.

### **6.3. Recommendations and further research**

The study summarizes the importance of conducting a comprehensive evaluation of COVID-19 patients early at the infection onset of the appearance of symptoms, whether it is a clinical evaluation or through laboratory and radiology tests.

Due to its great importance in determining the patient's condition and behavior of the disease in future, which helps to identify the medical needs that are required to deal with the case, and also contributes to preparing the health system in Palestine in a way that allows to deal professionally with these cases and avoid the lack of hospitals beds and medical teams. We recommend health institutions in Palestine adopt artificial intelligence programs with their pioneering role in the speed and accuracy of analyzing medical data, that contribute and support decision-makers to do what is necessary to confront the epidemic. Where prediction models contribute greatly to developing plans for a healthcare organization.

There is still a lack of comprehensive research on COVID-19 patients in Palestine, as we recommend to support and encourage scientific research on this field, as establishing a classification system for COVID-19 severity level, will aid healthcare staffs prioritize medical care for patients with more severe symptoms, however, decrease the burden on health care system caused by the growing numbers of infected people.

Our findings revealed that patients had a past medical history of diabetes, hypertension, malignancy, and cardiovascular disease will significantly increase their chance to develop severe COVID-19 symptoms and suffer from worse disease outcome, however further research of COVID-19 clinical characteristics for these group of patients is highly recommended since these

chronic diseases are widespread in the Palestinian community and consider as a risk factor of developing a severe COVID-19 symptom.

On other hand, back to our results we got from the prediction model using machine learning methods, where the highest accuracy we got is 74% for ANN's model, further research for prediction of the severity level, other epidemiological characteristics features can be used as input to machine learning tools. Also, other data mining technique could be used as hybrid machine learning to produce better accuracy of severity level in COVID-19 infected patients.

#### **6.4. Study limitation**

This study was confronted with several limitations. First, lack of data standardization reported by MOH, as the outbreak occurs, there were no special pre-established procedures for reporting cases or examinations. As well as the logistical issues surrounding the involved medical centers, that it was not ideal for dealing with this epidemic, in terms of capacity and appropriate equipment. While some treatment and quarantine centers use electronic record forms to manage patient data and others still use paper forms.

Second, the quality of reported data depends on the personal factor, where data is as accurate as the reporter has skills to report reliable data. That data accuracy falls under human behavior and skills with the absence of one standardized formal way to document data.

Third, according to the statements issued by the Palestinian Ministry of Health, many asymptomatic or mildly ill patients were quarantined and treated at home, so our study represents so far the clinical characteristics of patients with symptoms.

Finally: the features used in the prediction model were determined by two physicians, which was presented under the subject of personal competence, since many aspects of the disease are still ambiguous, this can lead to limit the accuracy of the model.

### **6.5. Strength of the study**

This study is the first of its kind in Palestine, as no study has linked almost all patient's medical information of (demographic, laboratory tests, past medical history, and complaining) to the severity level of the disease, also we have for the first time- used artificial intelligence as an effective tool to predict the patient's health status, which would lead to better understand the medical needs necessary to deal with this epidemic.

Although the field of data mining is quite young, it has likely to play a major role in improving the healthcare system in general, as it can worthily push boundaries of healthcare provider's performance by refining their medical knowledge in a short time. Data mining and machine learning as an authentic part of modern health informatics sciences behind conventional descriptive analysis will definitely support front-line health care providers to better patient management, and policymaker to handling medical resources well.

Employing the medical data and information resulting from the health sector regarding COVID-19 through descriptive analysis or machine learning tools to predict the patient's condition in the future will improve healthcare service's outcome and save lives so that it contributes significantly to providing the patients' needs at the time they need in terms of hospital places, medical equipment, dedicated and adequate medical staff.

## Bibliography

1. Ahani, A., & Nilashi, M. (2020). Coronavirus Outbreak and its Impacts on Global Economy: The Role of Social Network Sites. *Journal of Soft Computing and Decision Support Systems*, 7(2), 19–22.
2. Ahouz, F., & golabpour, amin. (2020). *Predicting the COVID-19 Prevalence Rate Using Data Mining*. <https://doi.org/10.21203/rs.3.rs-21247/v1>
3. Ajaimy, M., & Melamed, M. L. (2020). *COVID-19 in patients with kidney disease*. *Am Soc Nephrol*.
4. Albulescu, C. (2020). Coronavirus and oil price crash. *Available at SSRN 3553452*.
5. Ali, M. Y., & Gatiti, P. (2020). The COVID-19 (Coronavirus) pandemic: reflections on the roles of librarians and information professionals. *Health Information & Libraries Journal*, 37(2), 158–162.
6. AlKhalidi, M., Kaloti, R., Shella, D., Al Basuoni, A., & Meghari, H. (2020). Health system's response to the COVID-19 pandemic in conflict settings: Policy reflections from Palestine. *Global Public Health*, 15(8), 1244–1256.
7. Altable, M., & de la Serna, J. M. (n.d.). *Cerebrovascular Disease in COVID-19: Is There a Higher Risk of Stroke?*
8. Anderson, L. J., Nuckols, T. K., Coles, C., Le, M. M., Schnipper, J. L., Shane, R., Jackevicius, C., Lee, J., Pevnick, J. M., & PHARM-DC Group Choudhry Niteesh K MD, P. D. O. D. M. D. S. C. M. D. (2020). A systematic overview of systematic reviews evaluating medication adherence interventions. *American Journal of Health-System Pharmacy*, 77(2), 138–

147.

9. Apicella, M., Campopiano, M. C., Mantuano, M., Mazoni, L., Coppelli, A., & Del Prato, S. (2020a). COVID-19 in people with diabetes: understanding the reasons for worse outcomes. *The Lancet Diabetes and Endocrinology*, 8(9), 782–792. [https://doi.org/10.1016/S2213-8587\(20\)30238-2](https://doi.org/10.1016/S2213-8587(20)30238-2)
10. Apicella, M., Campopiano, M. C., Mantuano, M., Mazoni, L., Coppelli, A., & Del Prato, S. (2020b). COVID-19 in people with diabetes: understanding the reasons for worse outcomes. *The Lancet Diabetes & Endocrinology*.
11. Ayittey, F. K., Ayittey, M. K., Chiwero, N. B., Kamasah, J. S., & Dzuvor, C. (2020). Economic impacts of Wuhan 2019-nCoV on China and the world. *Journal of Medical Virology*.
12. Barzilay, O., & Brailovsky, V. L. (1999). On domain knowledge and feature selection using a support vector machine. *Pattern Recognition Letters*, 20(5), 475–484.
13. Beović, B., Doušak, M., Ferreira-Coimbra, J., Nadrah, K., Rubulotta, F., Belliato, M., Berger-Estilita, J., Ayoade, F., Rello, J., & Erdem, H. (2020). Antibiotic use in patients with COVID-19: a snapshot. Infectious Diseases International Research Initiative (ID-IRI) survey. *Journal of Antimicrobial Chemotherapy*.
14. Cai, Q., Huang, D., Yu, H., Zhu, Z., Xia, Z., Su, Y., Li, Z., Zhou, G., Gou, J., Qu, J., & others. (2020). COVID-19: abnormal liver function tests. *Journal of Hepatology*.
15. Cao, M., Zhang, D., Wang, Y., Lu, Y., Zhu, X., Li, Y., Xue, H., Lin, Y., Zhang, M., Sun, Y., & others. (2020). Clinical features of patients infected with the 2019 novel coronavirus (COVID-19) in Shanghai, China. *MedRxiv*.

16. Carfì, A., Bernabei, R., Landi, F., & others. (2020). Persistent symptoms in patients after acute COVID-19. *Jama*, 324(6), 603–605.
17. Carvalho, P. M. M., Moreira, M. M., de Oliveira, M. N. A., Landim, J. M. M., & Neto, M. L. R. (2020). The psychiatric impact of the novel coronavirus outbreak. *Psychiatry Research*, 286, 112902.
18. Cleverley, J., Piper, J., & Jones, M. M. (2020). The role of chest radiography in confirming covid-19 pneumonia. *Bmj*, 370.
19. Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Bio-Medica: Atenei Parmensis*, 91(1), 157–160.
20. European Surveillance System. (n.d.). *Clinical characteristics of COVID-19*. <https://www.ecdc.europa.eu/en/covid-19/latest-evidence/clinical>
21. Evans, O. (2020). Socio-economic impacts of novel coronavirus: The policy solutions. *BizEcons Quarterly*, 7, 3–12.
22. Fan, Z., Chen, L., Li, J., Cheng, X., Yang, J., Tian, C., Zhang, Y., Huang, S., Liu, Z., & Cheng, J. (2020). Clinical features of COVID-19-related liver damage. *Clinical Gastroenterology and Hepatology*.
23. Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet. Respiratory Medicine*, 8(4), e21.
24. Ferrari, E., Sartre, B., Squara, F., Contenti, J., Ocellini, C., Lemoel, F., Levraut, J., Doyen,

- D., Dellamonica, J., Mondain, V., & others. (2020). High prevalence of acquired thrombophilia without prognosis value in patients with coronavirus disease 2019. *Journal of the American Heart Association*, 9(21), e017773.
25. Gambert, S. (2013). Disease Severity. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (p. 606). Springer New York. [https://doi.org/10.1007/978-1-4419-1005-9\\_1385](https://doi.org/10.1007/978-1-4419-1005-9_1385)
26. Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S., & Ma, K. (2020). A systematic review of asymptomatic infections with COVID-19. *Journal of Microbiology, Immunology and Infection*, January. <https://doi.org/10.1016/j.jmii.2020.05.001>
27. Gostic, K., Gomez, A. C. R., Mummah, R. O., Kucharski, A. J., & Lloyd-Smith, J. O. (2020). Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife*, 9, e55570.
28. Grant, W. B., Lahore, H., McDonnell, S. L., Baggerly, C. A., French, C. B., Aliano, J. L., & Bhattoa, H. P. (2020). Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. *Nutrients*, 12(4), 988.
29. Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., & Hui, D. S. C. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*.
30. Henry, B. M., & Lippi, G. (2020). Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection. *International Urology and Nephrology*, 1–2.
31. Heymann, D. L., & Shindo, N. (2020). COVID-19: what is next for public health? *The Lancet*, 395(10224), 542–545.

32. Isidori, A. M., Arnaldi, G., Boscaro, M., Falorni, A., Giordano, C., Giordano, R., Pivonello, R., Pofi, R., Hasenmajer, V., Venneri, M. A., & others. (2020). COVID-19 infection and glucocorticoids: update from the Italian Society of Endocrinology Expert Opinion on steroid replacement in adrenal insufficiency. *Journal of Endocrinological Investigation*, 1.
33. Jamerson, B., & Haryadi, T. H. (2020). The Use of Ibuprofen to Treat Fever in COVID-19: A Possible Indirect Association with Worse Outcome? *Medical Hypotheses*, 109880.
34. JORDAN, J. (2017). *Evaluating a machine learning model*.
35. Kanwal, A., Agarwala, A., & Warsinger Martin, L. (2020). COVID-19 and Hypertension: What We Know and Don't Know. *American College of Cardiology*, 6.
36. Kayem, G., Lecarpentier, E., Deruelle, P., Bretelle, F., Azria, E., Blanc, J., Bohec, C., Bornes, M., Ceccaldi, P.-F., Chalet, Y., & others. (2020). A snapshot of the Covid-19 pandemic among pregnant women in France. *Journal of Gynecology Obstetrics and Human Reproduction*, 49(7), 101826.
37. Ling, W. (2020). C-reactive protein levels in the early stage of COVID-19. *Medecine et Maladies Infectieuses*.
38. Liu, X., Zhou, H., Zhou, Y., Wu, X., Zhao, Y., Lu, Y., Tan, W., Yuan, M., Ding, X., & Zou, J. (2020). Risk factors associated with disease severity and length of hospital stay in COVID-19 patients. *Journal of Infection*, 81(1), e95–e97.
39. Mahmoudi, H., Alikhani, M. Y., Taheri, N. M., & Behzadi, A. (2020). *Assessment of changes in blood urea and creatinine levels in patients with coronavirus disease 2019 (COVID-19)*.

40. Matt Craven , Linda Liu is a, M. W. ; and M. M. (n.d.). No Title COVID-19: Implications for business, Our latest perspectives on the coronavirus pandemic. *McKinsey & Company, Risk Practice*.
41. Medel-ramírez, C., Medel-lópez, H., Veracruzana, U., Investigaciones, I. De, & Económicos, S. (n.d.). *Data Article . Data mining for the study of the Epidemic ( SARS- CoV-2 ) COVID-19 : Algorithm for the Authors Corresponding author A bstract Keywords*. 1–8.
42. Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.
43. Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B., & Wu, J. C. (2020). COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nature Reviews Cardiology*, 1–16.
44. Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554–565.
45. Palmieri, L., Vanacore, N., Donfrancesco, C., Lo Noce, C., Canevelli, M., Punzo, O., Raparelli, V., Pezzotti, P., Riccardo, F., Bella, A., & others. (2020). Clinical characteristics of hospitalized individuals dying with COVID-19 by age group in Italy. *The Journals of Gerontology: Series A*, 75(9), 1796–1800.
46. Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., Zheng, D., Wang, J., Hesketh, R. L., Yang, L., & others. (2020). Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*.
47. Pan, L., Mu, M., Yang, P., Sun, Y., Wang, R., Yan, J., Li, P., Hu, B., Wang, J., Hu, C.,

Jin, Y., Niu, X., Ping, R., Du, Y., Li, T., Xu, G., Hu, Q., & Tu, L. (2020). Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. *American Journal of Gastroenterology*, *115*(5), 766–773.

<https://doi.org/10.14309/ajg.0000000000000620>

48. Pence, B. D. (2020). Severe COVID-19 and aging: are monocytes the key? *GeroScience*, *42*(4), 1051–1061.

49. Pranata, R., Soeroto, A. Y., Huang, I., Lim, M. A., Santoso, P., Permana, H., & Lukito, A. A. (2020). Effect of chronic obstructive pulmonary disease and smoking on the outcome of COVID-19. *The International Journal of Tuberculosis and Lung Disease*, *24*(8), 838–843.

50. Radin, D. P., & Tsirka, S. E. (2020). Interactions between Tumor Cells, Neurons, and Microglia in the Glioma Microenvironment. *International Journal of Molecular Sciences*, *21*(22), 8476.

51. Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., Pearson, C. A. B., Jombart, T., Procter, S. R., Knight, G. M., Group, C. W., & others. (2020). *COVID-19 length of hospital stay: a systematic review and data synthesis*.

52. Response, I., & Lives, C. S. (2020). *Occupied Palestinian Territory : COVID-19 Emergency*. *1*(March), 1–10.

53. Rico-Mesa, J. S., Rosas, D., Ahmadian-Tehrani, A., White, A., Anderson, A. S., & Chilton, R. (2020). The role of anticoagulation in COVID-19-induced hypercoagulability. *Current Cardiology Reports*, *22*(7), 1–6.

54. Rismanbaf, A., & Zarei, S. (2020). Liver and kidney injuries in COVID-19 and their effects on drug therapy; a letter to editor. *Archives of Academic Emergency Medicine*, *8*(1), e17--

e17.

55. Saire, J. C., & Panford-Quainoo, K. (2020). Twitter Interaction to Analyze Covid-19 Impact in Ghana, Africa from March to July. *ArXiv Preprint ArXiv:2008.12277*.
56. Saire, J. E. C. (2020). Data mining approach to analyze Covid19 dataset of Brazilian patients. *ArXiv*.
57. Salameh, J. P., Leeflang, M. M. G., Hooft, L., Islam, N., McGrath, T. A., van der Pol, C. B., Frank, R. A., Prager, R., Hare, S. S., Dennie, C., Spijker, R., Deeks, J. J., Dinnes, J., Jenniskens, K., Korevaar, D. A., Cohen, J. F., Van den Bruel, A., Takwoingi, Y., van de Wijgert, J., ... McInnes, M. D. F. (2020). Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews*, 2020(9).  
<https://doi.org/10.1002/14651858.CD013639.pub2>
58. Schiffrin, E. L., Flack, J. M., Ito, S., Muntner, P., & Webb, R. C. (2020). *Hypertension and COVID-19*. Oxford University Press US.
59. Shen, C., Yue, X., Wang, J., Shi, C., & Li, W. (2020). Nocturnal oxygen therapy as an option for early COVID-19. *International Journal of Infectious Diseases*, 98, 176–179.
60. Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., Fan, Y., & Zheng, C. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet Infectious Diseases*, 20(4), 425–434.  
[https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4)
61. Shoenfeld, Y. (2020). Corona (COVID-19) time musings: our involvement in COVID-19 pathogenesis, diagnosis, treatment and vaccine planning. *Autoimmunity Reviews*.

62. Tahamtan, A., & Ardebili, A. (2020). *Real-time RT-PCR in COVID-19 detection: issues affecting the results*. Taylor & Francis.
63. Teo, J. (2020). Early detection of silent hypoxia in Covid-19 pneumonia using smartphone pulse oximetry. *Journal of Medical Systems*, 44(8), 1–2.
64. Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T. N., Politou, M., Psaltopoulou, T., Gerotziafas, G., & Dimopoulos, M. A. (2020). Hematological findings and complications of COVID-19. *American Journal of Hematology*.
65. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., & others. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus--infected pneumonia in Wuhan, China. *Jama*, 323(11), 1061–1069.
66. WHO. (2020). Coronavirus disease 2019 (COVID-19) Situation Report – 65. *World Health Organization, March 2020*.
67. Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., Lui, M. M. S., Lee, J. C. Y., Chiu, K. W. H., Chung, T. W. H., Lee, E. Y. P., Wan, E. Y. F., Hung, I. F. N., Lam, T. P. W., Kuo, M. D., & Ng, M. Y. (2020). Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology*, 296(2), E72–E78. <https://doi.org/10.1148/radiol.2020201160>
68. Wu, Y., Xu, X., Chen, Z., Duan, J., Hashimoto, K., Yang, L., Liu, C., & Yang, C. (2020). Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain, Behavior, and Immunity*, 87, 18–22. <https://doi.org/https://doi.org/10.1016/j.bbi.2020.03.031>

69. Xie, J., Covassin, N., Fan, Z., Singh, P., Gao, W., Li, G., Kara, T., & Somers, V. K. (2020). Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clinic Proceedings*.
70. Yao, Y., Cao, J., Wang, Q., Shi, Q., Liu, K., Luo, Z., Chen, X., Chen, S., Yu, K., Huang, Z., & others. (2020). D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study. *Journal of Intensive Care*, 8(1), 1–11.
71. Zachariah, P., Johnson, C. L., Halabi, K. C., Ahn, D., Sen, A. I., Fischer, A., Banker, S. L., Giordano, M., Manice, C. S., & Diamond, R. (2020). Epidemiology, clinical features, and disease severity in patients with coronavirus disease 2019 (COVID-19) in a children's hospital in New York City, New York. *JAMA Pediatrics*, 174(10), e202430–e202430.
72. Zhang, L., Zhu, F., Xie, L., Wang, C., Wang, J., Chen, R., Jia, P., Guan, H. Q., Peng, L., Chen, Y., & others. (2020). Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Annals of Oncology*.
73. Zhang, Y. P. (2020). Analysis of Epidemiological characteristics of new coronavirus pneumonia. *Chin. J. Epidemiol*, 41(2), 1–7.
74. Zhu, W., Zeng, N., Wang, N., & others. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland*, 19, 67.
75. Zurada, J. M. (1992). *Introduction to artificial neural systems* (Vol. 8). West St. Paul.

76. Ahani, A., & Nilashi, M. (2020). Coronavirus Outbreak and its Impacts on Global Economy: The Role of Social Network Sites. *Journal of Soft Computing and Decision Support Systems*, 7(2), 19–22.
77. Ahouz, F., & golabpour, amin. (2020). *Predicting the COVID-19 Prevalence Rate Using Data Mining*. <https://doi.org/10.21203/rs.3.rs-21247/v1>
78. Ajaimy, M., & Melamed, M. L. (2020). *COVID-19 in patients with kidney disease*. *Am Soc Nephrol*.
79. Albulescu, C. (2020). Coronavirus and oil price crash. *Available at SSRN 3553452*.
80. Ali, M. Y., & Gatiti, P. (2020). The COVID-19 (Coronavirus) pandemic: reflections on the roles of librarians and information professionals. *Health Information & Libraries Journal*, 37(2), 158–162.
81. AlKhaldi, M., Kaloti, R., Shella, D., Al Basuoni, A., & Meghari, H. (2020). Health system's response to the COVID-19 pandemic in conflict settings: Policy reflections from Palestine. *Global Public Health*, 15(8), 1244–1256.
82. Altable, M., & de la Serna, J. M. (n.d.). *Cerebrovascular Disease in COVID-19: Is There a Higher Risk of Stroke?*
83. Anderson, L. J., Nuckols, T. K., Coles, C., Le, M. M., Schnipper, J. L., Shane, R., Jackevicius, C., Lee, J., Pevnick, J. M., & PHARM-DC Group Choudhry Niteesh K MD, P. D. O. D. M. D. S. C. M. D. (2020). A systematic overview of systematic reviews evaluating medication adherence interventions. *American Journal of Health-System Pharmacy*, 77(2), 138–147.

84. Apicella, M., Campopiano, M. C., Mantuano, M., Mazoni, L., Coppelli, A., & Del Prato, S. (2020a). COVID-19 in people with diabetes: understanding the reasons for worse outcomes. *The Lancet Diabetes and Endocrinology*, 8(9), 782–792. [https://doi.org/10.1016/S2213-8587\(20\)30238-2](https://doi.org/10.1016/S2213-8587(20)30238-2)
85. Apicella, M., Campopiano, M. C., Mantuano, M., Mazoni, L., Coppelli, A., & Del Prato, S. (2020b). COVID-19 in people with diabetes: understanding the reasons for worse outcomes. *The Lancet Diabetes & Endocrinology*.
86. Ayittey, F. K., Ayittey, M. K., Chiwero, N. B., Kamasah, J. S., & Dzuovor, C. (2020). Economic impacts of Wuhan 2019-nCoV on China and the world. *Journal of Medical Virology*.
87. Barzilay, O., & Brailovsky, V. L. (1999). On domain knowledge and feature selection using a support vector machine. *Pattern Recognition Letters*, 20(5), 475–484.
88. Beović, B., Doušak, M., Ferreira-Coimbra, J., Nadrah, K., Rubulotta, F., Belliato, M., Berger-Estilita, J., Ayoade, F., Rello, J., & Erdem, H. (2020). Antibiotic use in patients with COVID-19: a snapshot Infectious Diseases International Research Initiative (ID-IRI) survey. *Journal of Antimicrobial Chemotherapy*.
89. Cai, Q., Huang, D., Yu, H., Zhu, Z., Xia, Z., Su, Y., Li, Z., Zhou, G., Gou, J., Qu, J., & others. (2020). COVID-19: abnormal liver function tests. *Journal of Hepatology*.
90. Cao, M., Zhang, D., Wang, Y., Lu, Y., Zhu, X., Li, Y., Xue, H., Lin, Y., Zhang, M., Sun, Y., & others. (2020). Clinical features of patients infected with the 2019 novel coronavirus (COVID-19) in Shanghai, China. *MedRxiv*.
91. Carfì, A., Bernabei, R., Landi, F., & others. (2020). Persistent symptoms in patients

after acute COVID-19. *Jama*, 324(6), 603–605.

92. Carvalho, P. M. M., Moreira, M. M., de Oliveira, M. N. A., Landim, J. M. M., & Neto, M. L. R. (2020). The psychiatric impact of the novel coronavirus outbreak. *Psychiatry Research*, 286, 112902.

93. Cleverley, J., Piper, J., & Jones, M. M. (2020). The role of chest radiography in confirming covid-19 pneumonia. *Bmj*, 370.

94. Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Bio-Medica: Atenei Parmensis*, 91(1), 157–160.

95. European Surveillance System. (n.d.). *Clinical characteristics of COVID-19*.  
<https://www.ecdc.europa.eu/en/covid-19/latest-evidence/clinical>

96. Evans, O. (2020). Socio-economic impacts of novel coronavirus: The policy solutions. *BizEcons Quarterly*, 7, 3–12.

97. Fan, Z., Chen, L., Li, J., Cheng, X., Yang, J., Tian, C., Zhang, Y., Huang, S., Liu, Z., & Cheng, J. (2020). Clinical features of COVID-19-related liver damage. *Clinical Gastroenterology and Hepatology*.

98. Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet. Respiratory Medicine*, 8(4), e21.

99. Ferrari, E., Sartre, B., Squara, F., Contenti, J., Ocelli, C., Lemoel, F., Levraut, J., Doyen, D., Dellamonica, J., Mondain, V., & others. (2020). High prevalence of acquired thrombophilia without prognosis value in patients with coronavirus disease 2019. *Journal of the American Heart Association*, 9(21), e017773.

100. Gambert, S. (2013). Disease Severity. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (p. 606). Springer New York. [https://doi.org/10.1007/978-1-4419-1005-9\\_1385](https://doi.org/10.1007/978-1-4419-1005-9_1385)
101. Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S., & Ma, K. (2020). A systematic review of asymptomatic infections with COVID-19. *Journal of Microbiology, Immunology and Infection, January*. <https://doi.org/10.1016/j.jmii.2020.05.001>
102. Gostic, K., Gomez, A. C. R., Mummah, R. O., Kucharski, A. J., & Lloyd-Smith, J. O. (2020). Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife, 9*, e55570.
103. Grant, W. B., Lahore, H., McDonnell, S. L., Baggerly, C. A., French, C. B., Aliano, J. L., & Bhattoa, H. P. (2020). Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. *Nutrients, 12*(4), 988.
104. Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., & Hui, D. S. C. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*.
105. Henry, B. M., & Lippi, G. (2020). Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection. *International Urology and Nephrology, 1–2*.
106. Heymann, D. L., & Shindo, N. (2020). COVID-19: what is next for public health? *The Lancet, 395*(10224), 542–545.
107. Isidori, A. M., Arnaldi, G., Boscaro, M., Falorni, A., Giordano, C., Giordano, R.,

Pivonello, R., Pofi, R., Hasenmajer, V., Venneri, M. A., & others. (2020). COVID-19 infection and glucocorticoids: update from the Italian Society of Endocrinology Expert Opinion on steroid replacement in adrenal insufficiency. *Journal of Endocrinological Investigation*, 1.

108. Jamerson, B., & Haryadi, T. H. (2020). The Use of Ibuprofen to Treat Fever in COVID-19: A Possible Indirect Association with Worse Outcome? *Medical Hypotheses*, 109880.

109. JORDAN, J. (2017). *Evaluating a machine learning model*.

110. Kanwal, A., Agarwala, A., & Warsinger Martin, L. (2020). COVID-19 and Hypertension: What We Know and Don't Know. *American College of Cardiology*, 6.

111. Kayem, G., Lecarpentier, E., Deruelle, P., Bretelle, F., Azria, E., Blanc, J., Bohec, C., Bornes, M., Ceccaldi, P.-F., Chalet, Y., & others. (2020). A snapshot of the Covid-19 pandemic among pregnant women in France. *Journal of Gynecology Obstetrics and Human Reproduction*, 49(7), 101826.

112. Ling, W. (2020). C-reactive protein levels in the early stage of COVID-19. *Medecine et Maladies Infectieuses*.

113. Liu, X., Zhou, H., Zhou, Y., Wu, X., Zhao, Y., Lu, Y., Tan, W., Yuan, M., Ding, X., & Zou, J. (2020). Risk factors associated with disease severity and length of hospital stay in COVID-19 patients. *Journal of Infection*, 81(1), e95–e97.

114. Mahmoudi, H., Alikhani, M. Y., Taheri, N. M., & Behzadi, A. (2020). *Assessment of changes in blood urea and creatinine levels in patients with coronavirus disease 2019 (COVID-19)*.

115. Matt Craven , Linda Liu is a, M. W. ; and M. M. (n.d.). No Title COVID-19: Implications for business, Our latest perspectives on the coronavirus pandemic. *McKinsey & Company, Risk Practice*.
116. Medel-ramírez, C., Medel-lópez, H., Veracruzana, U., Investigaciones, I. De, & Económicos, S. (n.d.). *Data Article . Data mining for the study of the Epidemic ( SARS- CoV-2 ) COVID-19 : Algorithm for the Authors Corresponding author A bstract Keywords*. 1–8.
117. Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.
118. Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554–565.
119. Palmieri, L., Vanacore, N., Donfrancesco, C., Lo Noce, C., Canevelli, M., Punzo, O., Raparelli, V., Pezzotti, P., Riccardo, F., Bella, A., & others. (2020). Clinical characteristics of hospitalized individuals dying with COVID-19 by age group in Italy. *The Journals of Gerontology: Series A*, 75(9), 1796–1800.
120. Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., Zheng, D., Wang, J., Hesketh, R. L., Yang, L., & others. (2020). Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*.
121. Pan, L., Mu, M., Yang, P., Sun, Y., Wang, R., Yan, J., Li, P., Hu, B., Wang, J., Hu, C., Jin, Y., Niu, X., Ping, R., Du, Y., Li, T., Xu, G., Hu, Q., & Tu, L. (2020). Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. *American Journal of Gastroenterology*, 115(5), 766–773.  
<https://doi.org/10.14309/ajg.0000000000000620>
122. Pence, B. D. (2020). Severe COVID-19 and aging: are monocytes the key? *GeroScience*,

42(4), 1051–1061.

123. Pranata, R., Soeroto, A. Y., Huang, I., Lim, M. A., Santoso, P., Permana, H., & Lukito, A. A. (2020). Effect of chronic obstructive pulmonary disease and smoking on the outcome of COVID-19. *The International Journal of Tuberculosis and Lung Disease*, 24(8), 838–843.

124. Radin, D. P., & Tsirka, S. E. (2020). Interactions between Tumor Cells, Neurons, and Microglia in the Glioma Microenvironment. *International Journal of Molecular Sciences*, 21(22), 8476.

125. Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., Pearson, C. A. B., Jombart, T., Procter, S. R., Knight, G. M., Group, C. W., & others. (2020). *COVID-19 length of hospital stay: a systematic review and data synthesis*.

126. Response, I., & Lives, C. S. (2020). *Occupied Palestinian Territory : COVID-19 Emergency*. 1(March), 1–10.

127. Rico-Mesa, J. S., Rosas, D., Ahmadian-Tehrani, A., White, A., Anderson, A. S., & Chilton, R. (2020). The role of anticoagulation in COVID-19-induced hypercoagulability. *Current Cardiology Reports*, 22(7), 1–6.

128. Rismanbaf, A., & Zarei, S. (2020). Liver and kidney injuries in COVID-19 and their effects on drug therapy; a letter to editor. *Archives of Academic Emergency Medicine*, 8(1), e17--e17.

129. Saire, J. C., & Panford-Quainoo, K. (2020). Twitter Interaction to Analyze Covid-19 Impact in Ghana, Africa from March to July. *ArXiv Preprint ArXiv:2008.12277*.

130. Saire, J. E. C. (2020). Data mining approach to analyze Covid19 dataset of Brazilian patients. *ArXiv*.
131. Salameh, J. P., Leeflang, M. M. G., Hooft, L., Islam, N., McGrath, T. A., van der Pol, C. B., Frank, R. A., Prager, R., Hare, S. S., Dennie, C., Spijker, R., Deeks, J. J., Dinnes, J., Jenniskens, K., Korevaar, D. A., Cohen, J. F., Van den Bruel, A., Takwoingi, Y., van de Wijgert, J., ... McInnes, M. D. F. (2020). Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database of Systematic Reviews*, 2020(9).  
<https://doi.org/10.1002/14651858.CD013639.pub2>
132. Schiffrin, E. L., Flack, J. M., Ito, S., Muntner, P., & Webb, R. C. (2020). *Hypertension and COVID-19*. Oxford University Press US.
133. Shen, C., Yue, X., Wang, J., Shi, C., & Li, W. (2020). Nocturnal oxygen therapy as an option for early COVID-19. *International Journal of Infectious Diseases*, 98, 176–179.
134. Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., Fan, Y., & Zheng, C. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet Infectious Diseases*, 20(4), 425–434.  
[https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4)
135. Shoenfeld, Y. (2020). Corona (COVID-19) time musings: our involvement in COVID-19 pathogenesis, diagnosis, treatment and vaccine planning. *Autoimmunity Reviews*.
136. Tahamtan, A., & Ardebili, A. (2020). *Real-time RT-PCR in COVID-19 detection: issues affecting the results*. Taylor & Francis.
137. Teo, J. (2020). Early detection of silent hypoxia in Covid-19 pneumonia using

smartphone pulse oximetry. *Journal of Medical Systems*, 44(8), 1–2.

138. Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T. N., Politou, M., Psaltopoulou, T., Gerotziafas, G., & Dimopoulos, M. A. (2020). Hematological findings and complications of COVID-19. *American Journal of Hematology*.
139. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., & others. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus--infected pneumonia in Wuhan, China. *Jama*, 323(11), 1061–1069.
140. WHO. (2020). Coronavirus disease 2019 (COVID-19) Situation Report – 65. *World Health Organization, March 2020*.
141. Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., Lui, M. M. S., Lee, J. C. Y., Chiu, K. W. H., Chung, T. W. H., Lee, E. Y. P., Wan, E. Y. F., Hung, I. F. N., Lam, T. P. W., Kuo, M. D., & Ng, M. Y. (2020). Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology*, 296(2), E72–E78. <https://doi.org/10.1148/radiol.2020201160>
142. Wu, Y., Xu, X., Chen, Z., Duan, J., Hashimoto, K., Yang, L., Liu, C., & Yang, C. (2020). Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain, Behavior, and Immunity*, 87, 18–22. <https://doi.org/10.1016/j.bbi.2020.03.031>
143. Xie, J., Covassin, N., Fan, Z., Singh, P., Gao, W., Li, G., Kara, T., & Somers, V. K. (2020). Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clinic Proceedings*.
144. Yao, Y., Cao, J., Wang, Q., Shi, Q., Liu, K., Luo, Z., Chen, X., Chen, S., Yu, K., Huang, Z., & others. (2020). D-dimer as a biomarker for disease severity and mortality in COVID-19

patients: a case control study. *Journal of Intensive Care*, 8(1), 1–11.

145. Zachariah, P., Johnson, C. L., Halabi, K. C., Ahn, D., Sen, A. I., Fischer, A., Banker, S. L., Giordano, M., Manice, C. S., & Diamond, R. (2020). Epidemiology, clinical features, and disease severity in patients with coronavirus disease 2019 (COVID-19) in a children's hospital in New York City, New York. *JAMA Pediatrics*, 174(10), e202430–e202430.

146. Zhang, L., Zhu, F., Xie, L., Wang, C., Wang, J., Chen, R., Jia, P., Guan, H. Q., Peng, L., Chen, Y., & others. (2020). Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Annals of Oncology*.

147. Zhang, Y. P. (2020). Analysis of Epidemiological characteristics of new coronavirus pneumonia. *Chin. J. Epidemiol*, 41(2), 1–7.

148. Zhu, W., Zeng, N., Wang, N., & others. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland*, 19, 67.

149. Zurada, J. M. (1992). *Introduction to artificial neural systems* (Vol. 8). West St. Paul.

# Appendices

## Appendices 1: Data collection form

### DATA COLLECTION FORM

|                      |       |                             |       |
|----------------------|-------|-----------------------------|-------|
| No.                  | ..... | Chest Pain                  | ..... |
| patint ID            | ..... | Sore throught               | ..... |
| Dep.                 | ..... | Chilss                      | ..... |
| discharge            | ..... | Tachypnia                   | ..... |
| Age                  | ..... | Vomiting                    | ..... |
| Date of admission    | ..... | Fatigue                     | ..... |
| Date of discharge    | ..... | Runny Nose                  | ..... |
| LoS                  | ..... | Dizziness                   | ..... |
| PLAN                 | ..... | Hemoptysis                  | ..... |
| Past medical history | ..... | Weakness                    | ..... |
| Date                 | ..... | Arthralgia                  | ..... |
| SpO2                 | ..... | Loss Of Apetite             | ..... |
| Complaining          | ..... | PMH                         | ..... |
| severity level       | ..... | Hypertention (HTN)          | ..... |
| Complaining          | ..... | Diabetes(DM)                | ..... |
| Cough                | ..... | Malignancy(CA)              | ..... |
| Dyspnea              | ..... | Cerbrovascular Cisease      | ..... |
| Headach              | ..... | COPD                        | ..... |
| Diarrhea             | ..... | Chronic Kidny Disease (CKD) | ..... |
| SOB                  | ..... | Chronic Liver Disease       | ..... |
| Fever                | ..... | Rheumatoid (RA)             | ..... |
| Nausea               | ..... | Cardiac Disease (CD)        | ..... |


Note.....  
 .....  
 .....

Reporter Name & Signature

## Appendices 2: Ethical Committee of the Palestinian Ministry of Health Approval



**State of Palestine**  
Ministry of Health  
Minister's Office



**دولة فلسطين**  
وزارة الصحة  
مكتب الوزير

---



دولة فلسطين  
وزارة الصحة  
مكتب الوزير  
رقم الملف: 11258/148  
التاريخ: 2020/7/12

حضرة الباحث الدكتور ابراهيم صباح

تمت الموافقة على البحث المقدم من طرفكم بعنوان:

**Correlations of laboratory tests with COVID-19  
Symptoms severity in Palestine.**

على أن يتم الالتزام بالشروط والمعايير التالية:

- 1- الالتزام باخلاقيات البحث العلمي و السريه التامه
- 2- تحديد البيانات المطلوبة ومصدرها من خلال ارسال كتاب للجنة على الايميل [.dr.mustafa.qawasma@gmail.com](mailto:dr.mustafa.qawasma@gmail.com)
- 3- تحديد الباحث الرئيسي والباحثين المشاركين على ان يتم الاخذ بعين الاعتبار كل من ساهم في هذا البحث من وزارة صحة او غيرها.
- 4- يجب الاشارة عند نشر البحث الى وزارة الصحة الفلسطينية كمصدر للمعلومات
- 5- يجب اخذ موافقة لجنة البحث العلمي على نتائج البحث قبل النشر .
- 6- يمنع اعطاء المعلومات المقدمة من وزارة الصحة بهدف هذا البحث مع اي جهة او اشخاص وذلك تحت المسؤولية.



الدكتور مي سالم الكيلة  
وزيرة الصحة



---

Ministry of Health - Nablus- Tel.: 09/2384771/6 - Fax : 09/2384777  
Ministry of Health -Ramallah- Behind Palestine Medical Complex  
Tel.: 02/2964183 - Fax : 02-2964182  
Ministry of Health - Gaza- Tel. : 08/2846949 - Fax : 08/2826295

وزارة الصحة - نابلس - تلفون : 09/2384771/6 - فاكس : 09/2384777  
وزارة الصحة - رام الله - خلف مجمع فلسطين الطبي  
تلفون : 02/2964183 فاكس : 02/2964182  
وزارة الصحة - غزة - تلفون : 08/2846949 فاكس : 08/2826295

## ملخص الدراسة

## الخصائص الطبية لمرضى كوفيد 19 المدخلين للمستشفيات في فلسطين ( منهجية تحليل وصفي و التنقيب في البيانات)

## مقدمة

فلسطين هي واحدة من الدول التي تأثرت بوباء فايروس الكورونا، حيث تهدف هذه الدراسة البحثية إلى فهم وتحسين المعرفة السريرية والتشخيصية لمرضى هذا الفايروس في فلسطين باستخدام التحليل الإحصائي التقليدي وأدوات التنقيب والدكاء الصناعي في البيانات لتطوير نموذج يمكن أن يصنف ويتنبأ بالحالة الصحية المستقبلية للمريض.

## منهجية الدراسة

تم إجراء تصميم البحث الكمي باستخدام التحليل الوصفي للحصول على نظرة وفكرة عامة لـ 132 مريضاً خلال فترة مكوثه بالمستشفى. تم إجراء مراجعة بأثر رجعي للسجلات الطبية للمرضى. تم جمع البيانات من مستشفين مخصصين في الضفة الغربية في فلسطين للفترة من 19 آذار إلى 20 تموز 2020. حيث تشمل البيانات السريرية متابعة الفحوصات المخبرية، والملاحظة السريرية، وخطة العلاج خلال أول يومين من تاريخ ادخالهم للمستشفى.

## النتائج

كانت الأعراض الأكثر شيوعاً عند الدخول هي السعال (51.5%) والحمى (41.7%) وضيق التنفس (25.8%). كما و كان هنالك تفاوت واختلافات عدة بين الفحوصات المخبرية للحالات شديدة الاعراض والحالات العادية التي تشكو من اعراض متوسطه بما في ذلك ارتفاع ملحوظ في مستويات الفحوصات التالية

(WBC, neutrophil, LDH, ferritin, BUN, CRP, Ferritin level, Monocytes and )

(creatine) حيث كانت القيمه الاحتمالية لنتائج هذه الفحوصات وعلاقتها مع شدة اعراض المرض ( $P < .05$ ) كما ولوحظ انخفاض في مستويات الفحوصات

التالية (lymphocytes percentage and SPO2) حيث كانت القيمه الاحتمالية لنتائج هذه الفحوصات وعلاقتها مع شدة اعراض المرض

( $P < .001$ ).

ترتبط الإصابة بمرض السكري وارتفاع ضغط الدم والأمراض الدماغية والشرابين بشكل كبير بمستوى شدة اعراض مرض فايروس الكرونا ( $P < .05$ ). كما وكانت

العلاجات الأكثر شيوعًا المقدمة لمرضى الفايروس في المستشفيات هي ( Antipyretic (75.0%), Supplementary drugs )

((55.3%), Anticoagulant (47.0%), and Antibiotic (46.2%).

تم استخدام أربع أدوات ذكاء اصطناعي شائعة وذلك لتطوير نموذج تنبؤ لمستوى شدة الاعراض لمرضى الفايروس في المستقبل:

Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), )

(and Artificial Neural Networks

بحيث تراوحت نسبة دقة هذه الادوات والتي تحدد قدرتها على التمييز الصحيح لمستوى شدة اعراض المرض المستقبلية من 71% الى 74% .