PREDICTING ICU REQUIREMENTS OF COVID-19 PATIENTS USING ARTIFICIAL NEURAL NETWORK

(YAPAY SİNİR AĞI KULLANARAK COVID-19 HASTALARININ YOĞUN BAKIM ÜNİTESİ GEREKLİLİKLERİNİN TAHMİNİ)

by

Yeliz ÇOTOY, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

June 2022

ACKNOWLEDGEMENTS

I want to express my gratitude to my thesis advisor [redacted], for his mentoring and support of my master thesis. Besides my supervisor, I would also like to thank [redacted] and [redacted] for their insightful comments and detailed information about the study. Their guidance helped me throughout the research and writing of this thesis.

My sincere thanks also go out to İstanbul Doğuş University Industrial Engineering Department. I would like to thank all faculty members in the Department of Industrial Engineering for their academic and individual contributions to me.

I would like to thank [redacted] for giving me a great vision to become an academician and for her valuable contributions to me.

I would like to thank my beloved family who always believed in me throughout my life and supported me in every decision I made.

My special thanks to my darling and friends who stood by me in good and bad times, motivated me, increased my enthusiasm, and made me stronger.

June 2022
Yeliz ÇOTOY

TABLE OF CONTENTS

LIST OF SYMBOLS

ADHD	**:** Attention-deficit/hyperactivity disorder
AI	**:** Artificial Intelligence
AIDS	**:** Acquired Immune Deficiency Syndrome
ANN	**:** Artificial Neural Network
BI	**:** Business Intelligence
BMI	**:** Body Mass Index
CNN	**:** Convolutional Neural Network
COPD	**:** Chronic Obstructive Pulmonary Disease
COVID-19	**:** Coronavirus Disease 2019
GDP	**:** Gross Domestic Product
GDPR	**:** General Data Protection Regulation
HIPAA	**:** Health Insurance Portability and Accountability Act
HIV	**:** Human Immunodeficiency Virus
ICU	**:** Intensive Care Unit
ML	**:** Machine Learning
MLP	**:** Multi-layer Perceptron
NLP	**:** Natural Language Processing
PHI	**:** Protected Health Information
ReLU	**:** Rectified Linear Unit
RNN	**:** Recurrent Neural Network
WHO	**:** World Health Organization

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Since the first pneumonia case caused by the new 2019 coronavirus (COVID -19) was found in Wuhan, the impact and consequences of this pandemic, which has affected the whole world, are still ongoing. The healthcare sector is one of the most affected sectors by this pandemic. During the most difficult times of the COVID 19 pandemic, the intensive care unit (ICU) was usually the bottleneck of a medical facility. As ICU admission rates increased, overworked healthcare providers faced critical decisions about the lives of their patients. ICU admission has evolved from a medical decision to a resource allocation issue.

One of the great difficulties experienced by health care was the increase in hospital admissions due to the spread of the disease. The need for decision support arises at all levels of health care. The ability to predict what type of individual needs will occur with a positive test, or even earlier, is beneficial for authorities to plan resources effectively.

In this study, a decision support framework is proposed for predicting the demand of COVID-19 patients in ICU within the framework of prescriptive analysis and developing a classification model using artificial neural networks. Risk factors are derived from the literature review. The model has been trained, validated, and tested using a total dataset of 120026 COVID 19 cases. During the test phase, performance is observed using confusion matrices, training and validation loss curves, and other performance metrics such as accuracy, precision, recall, and F1 score. The Synthetic Minority Oversampling Technique (SMOTE) is applied to unbalanced datasets to improve performance. It was predicted with 79% accuracy whether a person requires the Intensive Care Unit.

ÖZET

COVID -19'un neden olduğu ilk pnömoni vakasının Wuhan'da bulunmasından bu yana tüm dünyayı etkisi altına alan bu salgının etkisi ve sonuçları halen devam ediyor. Sağlık, bu salgından en çok etkilenen sektörlerden biridir. COVID-19 pandemisinin en zor zamanlarında, yoğun bakım ünitesi (YBÜ) genellikle bir tıbbi tesisin darboğazıydı. Aşırı çalışan sağlık hizmeti görevlileri, yoğun bakım ünitesine kabul oranları arttıkça hastalarının yaşamları hakkında kritik kararlarla karşı karşıya kaldı. Yoğun bakım ünitesine kabul, zamanla tıbbi bir karardan kaynak tahsisi sorununa dönüştü.

Sağlıkta yaşanan en büyük zorluklardan biri de hastalığın yayılması nedeniyle hastane başvurularının artmasıydı. Karar desteğine duyulan ihtiyaç, sağlık hizmetlerinin tüm seviyelerinde ortaya çıkmaktaydı. Hastanın testinin pozitif çıkmasından bile daha erken bir zamanda ne tür bireysel ihtiyaçların ortaya çıkacağını tahmin etme yeteneği, yetkililerin kaynakları etkili bir şekilde planlaması için faydalı olacaktı.

Bu çalışmada, Meksika Hükümeti'nin yayınladığı açık veri kaynaklarından yararlanılmış ve oradaki hastanelere başvuran hastaların COVID-19 sebebiyle yoğun bakıma ihtiyacının olup olmayacağı kuralcı analitik çerçevesinde incelenmiş ve yapay sinir ağları kullanılarak bir sınıflandırma modeli geliştirilmiştir. Aynı zamanda literatürdeki COVID-19 risk faktörleri araştırılmıştır. Model, toplam 120026 COVID-19 vakanın veri seti kullanılarak eğitilmiş, doğrulanmış ve test edilmiştir. Test aşamasında performansı ölçmek için, karışıklık matrisleri, eğitim ve doğrulama kaybı eğrileri ve doğruluk, kesinlik, geri çağırma ve F1 puanı gibi diğer performans ölçütleri kullanılmıştır. Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE), performansı artırmak için dengesiz veri kümelerine uygulanır ve bu tezde SMOTE kullanılarak veri seti dengelenmiştir. Sonuç olarak bir kişinin yoğun bakım ünitesine ihtiyacının olup olmadığı %79 doğruluk oranıyla tahmin edilmiştir.

1. INTRODUCTION

According to the World Health Organization as of March 14, 2022, the COVID 19 pandemic resulted in more than 450 million confirmed cases and more than 6 million deaths. In other words, the COVID 19 pandemic is one of the greatest public health challenges in human history. Since the detection of a new virus in Wuhan China in December 2019, COVID 19 has spread with a high infection rate (R0 value greater than 2) (Xu et al., 2021). In terms of scale, spread, violence, and death, this situation represents an unprecedented pandemic.

A catastrophic health crisis like the COVID-19 pandemic can be used to better our healthcare systems in the future. During an outbreak, one of the most crucial elements that we can enhance is the effective delivery of medical supplies. One of the major issues affecting healthcare professionals throughout the epidemic phase is a shortage of medical resources and an effective plan to distribute them properly.

Data and information have long played an important role in decision-making and healthcare delivery. Even if nations' strategies for combatting coronavirus illness differ, many common issues must be studied, researched, and assessed globally when battling this pandemic.

With each new case and discovery, information and data on coronavirus disease grow. The healthcare sector has become extremely data driven as the number of data provided by numerous sources has increased. The complexity and variety of data characterize this business. The necessary instruments must be in place for the health industry to efficiently control costs and enhance service quality. Health analytics must be used effectively to deliver meaningful insights and aid in medical decision making.

Analytics is the process of discovering new insights from large amounts of data. It uses various techniques such as machine learning and statistical algorithms to extract meaningful information from data. Due to the increasing number of chronic diseases and demand for specialized treatments, many countries have changed their approach from the old paradigm of medical decision making. The failure of the old paradigm of medical decision making has encouraged the healthcare industry to take advantage of Big Data and analytics to customize treatments according to the lifestyle and environment of the individual patient. This approach considers the different characteristics of the patient and the differences between the patient and the condition.

Prescriptive analytics is a method that uses machine learning to learn how to anticipate better results. It operates by asking, What adjustments can we make in accordance with our adverse event prediction? Prescriptive analytics is developing as a powerful technique in medical decision-making. Although machine learning algorithms are commonly employed in epidemiological investigations, their potential for COVID-19 has not been well investigated.(Vepa et al., 2021).

In these bad times, being able to forecast what type of resource that individual would need when confirmed positively, or before, will be of enormous assistance to authorities, allowing them to give and alter the required resources, and this prediction may save that patient's life. Improved disease management and hospital administration can benefit from more precise estimates of ICU demands. It allows for more precise use of limited resources (country health budget and hospital capacity).

High admission rates in hospitals and intensive care units have put pressure on healthcare systems all over the world, with critical patients suffering from multi-organ failure due to hyper inflammation, thrombosis, and other reasons. Healthcare resources may be insufficient to fulfill demand due to an inflow of patients suffering from serious or life-threatening disorders. Patients with acute illnesses may be admitted to the emergency room right away, deteriorate after being admitted to the hospital, or be referred to emergency physicians by critical care units. Overworked medical professionals may be forced to make difficult judgments concerning the patient's ICU care in such an unusual situation. There is only local guidance (Azoulay et al., 2020) to aid professionals in

organizing patients' journeys and standardizing processes across facilities to avoid acting on a first-come, first-served basis.

Tyrrell et al. consolidated current recommendations and highlighted the many methods utilized across the world to deal with these tough situations, examining how the rules were similar and different. During the epidemic, experts compiled a list of factors to consider when developing or revising recommendations for controlling ICU admissions (Tyrrell et al., 2021).

This thesis presents a decision support framework for reliably predicting the requirement for ICU of a patient with COVID-19 in the course of the disease, based on the patient's features at admission to the hospital, in the context of prescriptive analytics. As a result, this research makes two contributions: The first one constructs a model for accurately predicting a COVID19 patient's ICU requirement, and then it examines the use of machine learning approaches in making rapid medical decisions.

This study continues as follows: the importance of data and healthcare analytics are discussed in the next section. Prescriptive analytics in healthcare is also discussed in terms of its usefulness, benefits, and approaches. The COVID -19 risk factors were evaluated and categorized from literature. The third section explains the data and methodology used in the study, which includes prescriptive analytics and an artificial neural network. The artificial neural network model is explained in the fourth section. It is explained there how the model is constructed. A detailed numerical application explaining the use of artificial neural network methods is presented, and the results obtained using the proposed method are presented and discussed in Section 5. Finally, the sixth section concludes the study.

2. LITERATURE REVIEW

2.1 Healthcare Analytics

The literature on healthcare analytics is examined and given here. Analytics forms are studied, and the benefits of prescriptive analytics are presented. Prescriptive analytics methodologies are also described here.

The healthcare industry's fast development and evolution have generated the significant potential for healthcare professionals and organizations to strengthen their operations and improve patient health outcomes. The development and evolution of healthcare analytics have an impact on how the healthcare sector is now conducted. (Blessy Trencia Lincy & Suresh Kum, 2018).

The collecting, organizing, and processing of medical and other health information is referred to as health analytics. It is driven by 5 main segments. To evaluate healthcare data, five types of healthcare analytics are employed which include descriptive healthcare analytics for identifying relevant trends and health information modeling for designing and implementing evidence-based strategies.

A healthcare practitioner using a smartphone to detect a person's health state in real-time is an example of such an application. The result is insufficient to educate the user of the most effective and actionable technique. Because of the complexity and fragmentation of the data, it is difficult for healthcare practitioners to make educated judgments about the best appropriate therapy. This procedure is also known as medical decision making.

Because of the huge trend of predictive healthcare analytics, has transformed the daily healthcare sector. It can forecast the onset and severity of sickness, for instance, by identifying persons at risk for specific diseases. According to on the insights supplied by the other three analytics, prescriptive analytics offer answers to often asked questions such as what should we do and what should we avoid doing..

2.1.1 Diagnostic Analytics

The term diagnostic analytics refers to the process of examining data to answer various questions. It is done through the utilization of various techniques such as discovery, mining, correlation and associations (Mosavi & Santos, 2020).

2.1.2 Descriptive Analytics

Descriptive analytics are the most essential part of the analytics and are used by most of the industries and organizations. They help in uncovering the motives behind the events and results that happened in the past (Poornima & Pushpalatha, 2020).

2.1.3 Predictive Analytics

The reduction of data is the first step in the process of data reduction. Analyzing past trends and drifts in the data can help in formulating strategies that are feasible in the present and future.

Predictive analytics is mainly used to learn the various trends and observe patterns to formulate predictions for the future. It uses various techniques such as web server profiling, statistical modelling, and predictive text mining(Lopes et al., 2020).

Predictive analytics is an improved approach to solve problems that cannot be answered by conventional BI methods. It offers an improved method to make future predictions that are not possible with conventional methods.

2.1.4 Prescriptive Analytics

Big data can help improve the efficiency and effectiveness of a business by uncovering the various factors that led to the events or issues that occurred. It can also help create suggestions for the future needs of the organization. Prescriptive analytics are done for various reasons such as to identify the best possible outcome from the available data, to classify the uncertainties in the data, and to produce better decisions (Poornima & Pushpalatha, 2020).

Prescriptive analytics are done based on various data elements such as organization's internal and external data. It can be defined as an analysis of various business regulations and their various interpretations. These are also reasonably multifaceted and are not until now being used in a day-to-day basis.

2.1.5 Discovery Analytics

Through discovery analytics, healthcare providers and researchers can identify unknown diseases and medical conditions, as well as explore new treatment and medicine options(Mosavi & Santos, 2020).

2.2 Prescriptive Analytics in Healthcare

Prescriptive analytics are by far the most advanced sort of analytics, evaluating different approaches to goal achievement. They use various techniques such as optimization, simulation, heuristics, and multi-criteria decision making, and also enablers (e.g., deep learning, cognitive computing, and big data) to evaluate the results. (Mosavi & Santos, 2020)

2.2.1 Why Prescriptive Analytics?

Prescriptive analytics is based on the capability of artificial intelligence tools, which enable computers to analyze massive volumes of data without the need for human intervention. Computer programs automatically change to take advantage of new data as it becomes available. This technique is significantly more extensive and quicker than what

people are capable of. Another sort of data analytics is predictive analytics, which employs statistical modeling and predictive analysis to forecast future performance. It works by predicting the future based on current and past data sets and what is expected to happen and then suggest on how to proceed. Prescriptive analytics can assist in determining the most likely plan of action to pursue in the case of a change in circumstances.

As we collect data, we will demonstrate the usability of this method, and we will also measure the performance of people's risk estimates of getting coronavirus diseases (COVID-19) based on risk factors. By modelling the probability of various scenarios, they can help healthcare organizations plan strategies that are more likely to work. The guiding references in the study are given in the Table 2.1.

Table 2.1: Literature review on prescriptive analytics in healthcare

Author	Title	Objective
(Bohr & Memarzadeh, 2020)	Current healthcare, big data, and machine learning	Investigation
(Mosavi & Santos, 2020)	How prescriptive analytics influences decision making in precision medicine	Highlight
(Lepenioti et al., 2020)	International Journal of Information Management Prescriptive analytics: Literature review and research challenges	Investigation
(Poornima & Pushpalatha, 2020)	A survey on various applications of prescriptive analytics	Literature Survey
(Mehta et al., 2019)	Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study	Case Study
(Schwartz et al., 2017)	Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience	Case Study
(Srinivas & Ravindran, 2018)	Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework	Framework

2.2.2 Prescriptive Analytics Methods

The existing literature on prescriptive analytics and prominent methods for its implementation has been reviewed (Lepenioti et al., 2020). The literature review article has been reached, and the following Figure 2.1 is taken from this review article. Machine learning algorithms are generally used in health applications of prescriptive analytics (Mehta et al., 2019).

2.3 Artificial Intelligence in Healthcare

With the increasing amount of data that patients have accumulated over the years, it is now possible to consolidate these records and use them to improve the treatment of patients. Big data has the potential to provide valuable information to healthcare organizations and to transform the way healthcare is delivered. Its processing and interpretation can provide valuable insights and recommendations to improve patient care.

AI techniques can help gather this data and use it to improve the efficiency of their operations. These techniques draw its expertise from various fields of science such as statistics, data science, and optimization. It uses learning algorithms to analyze and predict the future based on available data.

2.3.1 Machine Learning

Machine learning (ML) is a sub-discipline of AI systems that contains a range of techniques for analyzing data and identifying connections within a dataset. These models can be used to analyze difficult health problems (e.g., disease risk factors) or make predictions health status (e.g., predicting disease prognosis). In addition to standard algorithms, ML employs learning techniques and numerical simulations based on facts and probabilities, as well as other scientific domains such as computer science, statistics, and optimization. It differs from typical programming in that it uses prediction to forecast unavailable data using data that is already accessible. Data may be used to train learning algorithms, which can then be fed back with the desired outputs.

The three forms of machine learning are: (1) supervised learning, (2) unsupervised learning, and (3) semi-supervised learning. Supervised learning train algorithms to handle regression and classification problems using datasets with pre-labeled outcomes. It is highly suited for regression issues since it employs input factors to predict a specific output, but it is time demanding owing to manual data labeling (supervision) and requires vast volumes of data.

By creating correlations between patient characteristics as inputs and the result of interest as outputs, supervised learning is ideal for predictive modeling.

Unsupervised learning makes use of datasets with unlabeled inputs, allowing algorithms to extract features and structures from the data autonomously. It recognizes the data it is exposed to using pre-learned characteristics. It is frequently used to discover hidden patterns in data, with the objective of locating these patterns without the usage of human input (Holzinger, 2016).

Semi-supervised learning combines unsupervised and supervised learning methods. Because fully labeled data involves extensive investigation, it is a model in which just a fraction of the output data is labeled. When there is existing unlabeled data and labeling is time-consuming, this technique might be beneficial. Self-training is a model in which classifiers are trained with labeled data first, then fed unlabeled data, with the predictions included in the training dataset (Jiang et al., 2017).

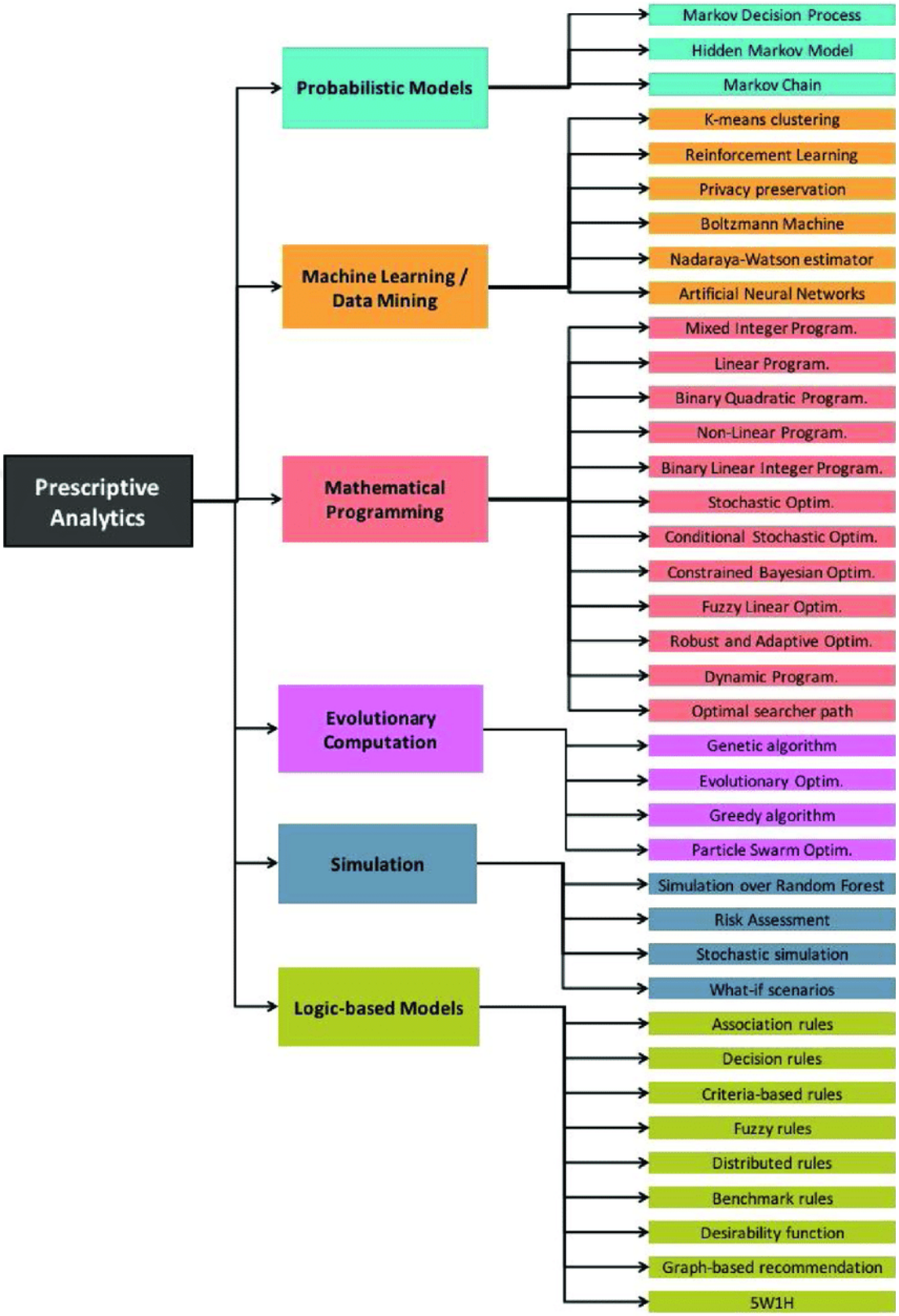

Figure 2.1: Prescriptive analytics methods (Lepenioti et al., 2020)

2.4 COVID-19 Risk Factors Assessment

The literature on COVID-19 has been evaluated, and 40 risk factors for the disease have been found. Some of the underlying disorders that cause COVID-19 that were discovered in the literature study are in the form of compounds. They symbolize the possibility of contracting COVID-19.

- Chronic kidney disease -Having dialysis or has severe (stage 5) long-term kidney disease (R. Wadhera, P. Wadhera, Gaba, Figueroa, Yeh, Shen, 2020)
- Cancer (Alger et al., 2020)
- Chronic lung diseases, including COPD (chronic obstructive pulmonary disease), asthma (moderate-to-severe), interstitial lung disease, cystic fibrosis, and pulmonary hypertension (Alger et al., 2020)
- Dementia or other neurological conditions (Medicine & Wars, 2021)
- Stroke or cerebrovascular disease, which affects blood flow to the brain (Terpos et al., 2020)
- Mental disorders like schizophrenia, ADHD, autism, and cerebral palsy (Wang et al., 2021)
- Down syndrome (Wang et al., 2021)
- Diabetes (type 1 or type 2) (R. Wadhera, P. Wadhera, Gaba, Figueroa, Yeh, Shen, 2020)
- Heart conditions (such as heart failure, coronary artery disease, cardiomyopathies or hypertension) (Roedl et al., 2020)
- Immunocompromised state (weakened immune system) (Allam et al., 2020)
- Have been treated in the past 5 years for a cancer of the blood or bone marrow (such as leukemia, lymphoma or myeloma) (Terpos et al., 2020)
- Have been treated in the past 1 year for a cancer that did not start in the blood or bone marrow (Terpos et al., 2020)
- Solid organ or blood stem cell transplant (Terpos et al., 2020)
- HIV/AIDS (Hao et al., 2020)
- Long-term use of prednisone or similar drugs that weaken your immune system such as steroids (Wooding & Bach, 2020)
- Sickle cell disease or thalassemia (Oakes et al., 2021)

- Having hemoglobin blood disorders like sickle cell disease (SCD) or thalassemia (Levy et al., 2020)
- Substance use disorders (such as alcohol, opioid, or cocaine use disorder) (Wang et al., 2021)
- Liver disease (R. Wadhera, P. Wadhera, Gaba, Figueroa, Yeh, Shen, 2020)
- People with disability (Oakes et al., 2021)
- Overweight and obesity (Yadaw et al., 2020)
- Pregnancy and breastfeeding (Oakes et al., 2021)
- Smoking, current or former (Zheng et al., 2020)
- Older Age (Brandt et al., 2020)
- Gender (R. Wadhera, P. Wadhera, Gaba, Figueroa, Yeh, Shen, 2020)
- Race and Ethnicity (Oakes et al., 2021)
- Essential workers- Doctors, nurses, Grocery store employees, mail carriers, bus drivers, and others also have important jobs that can't be done at home. Another population besides healthcare professionals at risk of contracting COVID-19 and spreading the virus is essential workers (Allam et al., 2020).
- Poor ventilation (Albayati et al., 2021)
- Malnutrition (indirect in low-income nations) and living in poverty- such as homeless people (Albayati et al., 2021)
- Living area- metropolises (Cao et al., 2021)
- people in aged care facilities (Summers et al., 2020)
- Latitude (Cao et al., 2021)
- air pollution (Cao et al., 2021)
- wind speed (Cao et al., 2021)
- the total number of participants in major sports events (Cao et al., 2021)
- GDP per capita (Cao et al., 2021)
- Weather temperature (Cao et al., 2021)
- population density (Cao et al., 2021)
- industrial city (Aabed & Lashin, 2021)
- Sea level (Aabed & Lashin, 2021)

2.4.1 Risk Factor Groups

To begin, all characteristics influencing COVID -19 patients are derived from the literature and classified into four groups.

-Demographic Factors

Age, gender, race/ethnicity

-Comorbidities and health situation

Chronic kidney diseases, Cancer, lung diseases, hypertension, diabetes, Dementia or other neurological conditions, Stroke or cerebrovascular disease, Mental disorders, Down syndrome, Heart conditions, Immunocompromised state, leukemia, lymphoma or myeloma, Cancer treatment, Solid organ or blood stem cell transplant, HIV/AIDS, Sickle cell disease or thalassemia, Liver disease, Overweight and obesity, BMI, People with disability, Pregnancy

-Environmental and Geographic Factors

Climate, Air Pollution, Population Density, Relative humidity, Weather temperature, Latitude, Wind speed, Sea level, Living area

-Lifestyle and Socioeconomic Factors

Smoking, essential workers, poor ventilation, Substance use disorders, GDP per capita

3. DATA AND METHODOLOGY

Most of the elements that may have an impact on the COVID-19 are based on research. An extensive literature study was undertaken to identify COVID-19 risk variables. For this purpose, 40 factors were found in the literature review and then these factors were divided into groups such as socioeconomic parameters, demographic factors, comorbidities. As an example, also a database containing only publications about COVID-19 was used to scan the article. COVID Scholar is a knowledge portal designed with the unique needs of the COVID-19 research community in mind, utilizing natural language processing (NLP) to aid researchers in synthesizing the information spread across thousands of emergent research articles, patents, and clinical trials into actionable insights and new knowledge (Trewartha et al., 2020).

It was intended to make use of articles in the literature as well as open data published by the World Health Organization (WHO) on a country-by-country basis. There was no data set in the literature that included these risk variables. Some articles study them, but since their data is not open source, another dataset was sought, and the open-source dataset provided by the Mexican government was reached. The patient data includes information from the Mexican open-source database and the patient's demographic and additional diseases, the patient's COVID -19 test result, whether they smoke, whether they are inpatients or outpatients, and whether they are intubated. The most common risk factor is that the patient was diagnosed with pneumonia, hypertension, diabetes, and obesity, respectively.

Collecting patient data, which includes patient-specific information about medical history and habits of COVID -19 patients, is completely different and protected by law. This is largely due to regulatory security laws such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), which

make it nearly impossible for anyone to access compelled protected health information (PHI).

3.1 Data Collection

The dataset in this thesis incorporates information from the Epidemiological Surveillance System for Viral Respiratory Diseases, which is one of COVID-19 open databases in Mexico. Only data from an epidemiological examination of a suspected case of viral respiratory illness upon diagnosis in the Health Sector's medical units are included. The patient is classified as an outpatient or inpatient based on the clinical diagnosis at the time of admission. The database does not contain activities that happen during a patient's stay at a medical facility.

3.2 Data Pre-processing

All variables from patient data are converted to numeric data. Gender (male-female) is defined as a nominal binary variable (1–0). Binary variables are all Yes–No type variables (1–0). (Intubated, pneumonia, pregnancy, diabetes, COPD, asthma, immune-suppression, hypertension, other diseases, cardiovascular, obesity, renal chronic, smoking, COVID-19 test result, ICU admission). Age and the number of symptom days are scale variables. The dataset included 566602 laboratory-confirmed COVID-19 cases. There are several missing values in the original dataset. We excluded 446576 patients whose features had not been recorded, as well as those whose data was missing or unknown. Our study included a total of 120026 patients. Because of the large proportion of missing information, five features are removed from the data.

After the data cleaning phase, 18 characteristics from 120 026 patients were incorporated into the model to produce more accurate findings. The remaining 18 risk variables are shown and defined below after eliminating these missing 5 characteristics. These are the characteristics of the patient as described in the dataset.

Final Risk Factors

1. Gender: Identifies the patient's gender.
2. Age: Identifies the age of the patient.
3. Num_days_sympthom: Identifies the date on which the patient's symptoms began.
4. Intubated: Identifies if the patient required intubation.
5. Pneumonia: Identifies if the patient was diagnosed with pneumonia.
6. Pregnancy: Identifies if the patient is pregnant.
7. Diabetes: Identifies if the patient has a diagnosis of diabetes.
8. Copd: Identifies if the patient has a diagnosis of COPD.
9. Asthma: Identifies if the patient has a diagnosis of asthma.
10. Inmsupr: Identifies if the patient has immunosuppression.
11. Hypertansion: Identifies if the patient has a diagnosis of hypertension.
12. Other_disease: Identifies if the patient has a diagnosis of other diseases.
13. Cardiovascular: Identifies if the patient has a diagnosis of cardiovascular disease.
14. Obesity: Identifies if the patient is diagnosed with obesity.
15. Renal_chronic: Identifies if the patient has a diagnosis of chronic kidney failure.
16. Tobacco: Identify if the patient has a smoking habit.
17. Covid_result: Identifies the result of the COVID-19 test.
18. ICU (intensive care unit): Identifies if the patient required to enter an Intensive Care Unit.

The most common characteristics of the patients included in the study are shown in the Figure 3.1 below from most common to the least. The most common feature is that the patient was diagnosed with pneumonia, hypertension, diabetes, and obesity, respectively.

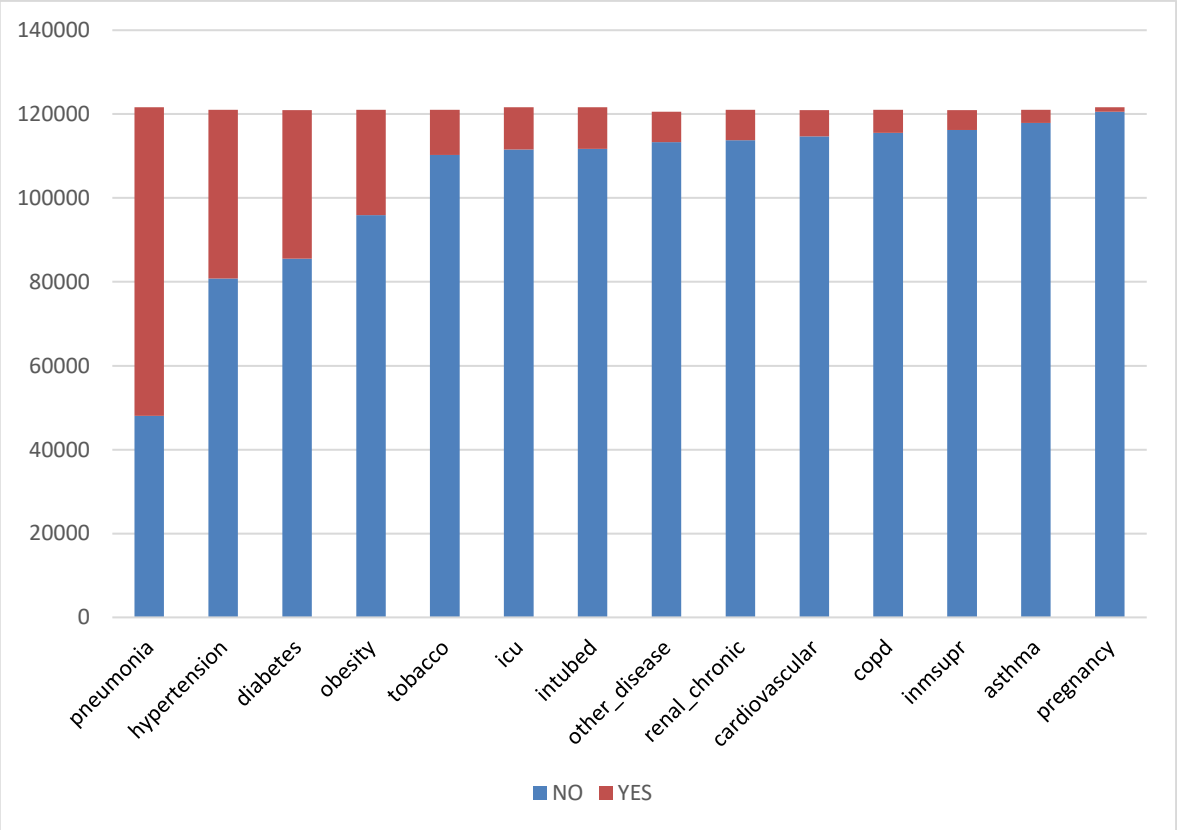

Figure 3.1: Common characteristics of the patient medical records

3.3 Methodology

This study's approach is divided into three major parts, as shown in Figure 3.2. The first step is to conduct a literature review to identify the important factors influencing the necessity for an ICU. We conducted a risk factor analysis. It was necessary to collect patient data as open access data for the study. The Mexican Ministry of Health authorized open access data was gathered and preprocessed to fit the model and establish the final components. The second step focuses with mathematical models and applications. In this study, machine learning techniques are studied, and artificial neural networks are used in practice. The third step is to conduct additional analysis of the results.

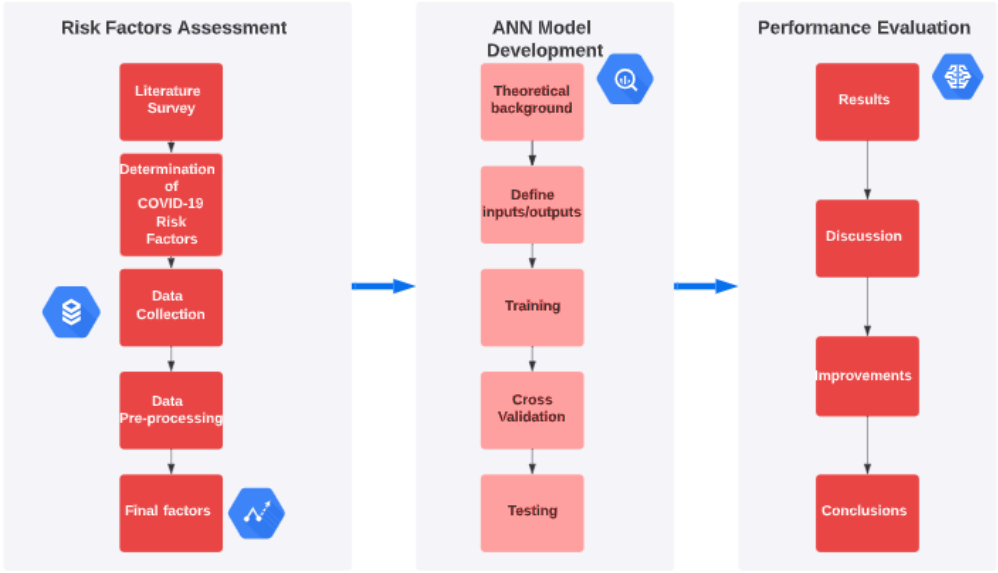

Figure 3.2: Methodology representation

3.3.1 Artificial neural networks

ANN is a three-layered intelligent mathematical algorithm. There is an input layer, an intermediate or concealed layer(s), and an output layer. The system's inputs comprise the initial structure, the input layer. The hidden layer, the second component, is the core of the artificial neural network and is composed of multiple substructures known as neurons. The major mathematical computations occur within these neurons to process the inputs and provide suitable outputs. It functions just like a real neuron in the human brain. The neurons in this layer collect data from the previous layer and pass them to the next layer. The neuron's received and transmitted values may change based on the weight value of the wire going to the neuron and conveying the neuron's value. Before passing the result to the next neuron, the weight value of the channel is multiplied by the value communicated, i.e., the weight value is derived by multiplying the preceding neuron's value. As the specified task varies, so does the weight value; the model decides this value by learning and remembering how to complete the task.

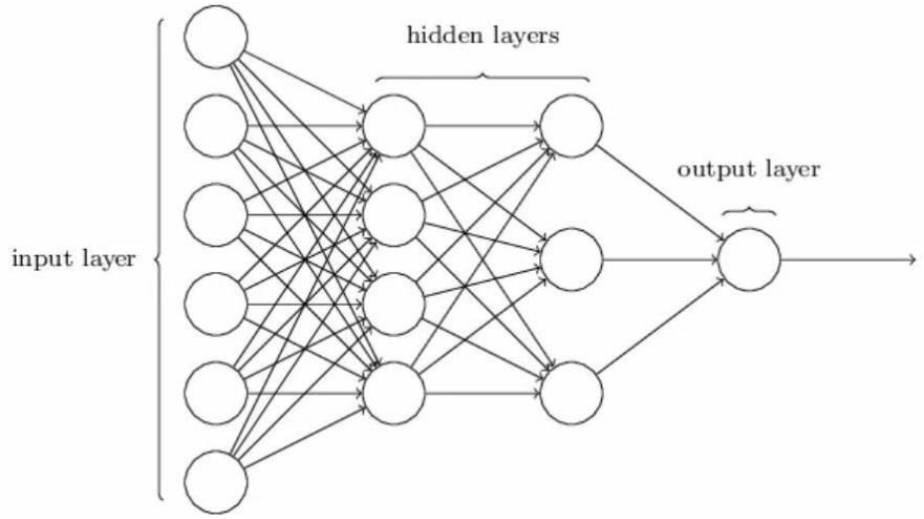

Figure 3.3: General structure of Neural Network

Each neuron takes input from several other neurons, multiplies it by its assigned weights, adds it all up, and delivers the result to one or more neurons. We create an artificial neural network that can do even highly difficult tasks when hundreds, thousands, or even millions of neurons are placed in numerous layers and layered on top of each other. The weights of the connections between neurons in artificial neural networks are first given random values. The ANN can accomplish its duty successfully if these weights are set to the correct numbers. When we're working with numerous layers and thousands of neurons, finding the proper weights isn't straightforward. Artificial neurons are the fundamental component of ANNs. The essential parts of an Artificial Neuron are shown in Figure 3.4.

To find the right weights, the network is trained with historical data. The neural network continually changes its weights as additional training examples are presented, matching each input with the corresponding output. During training, the network uses the data to adapt to recognize and gather certain patterns within the data. One of the most difficult parts of training is finding training samples of the right quantity and quality. Also, the larger the AI model is, the more computational resources are needed to train the model.

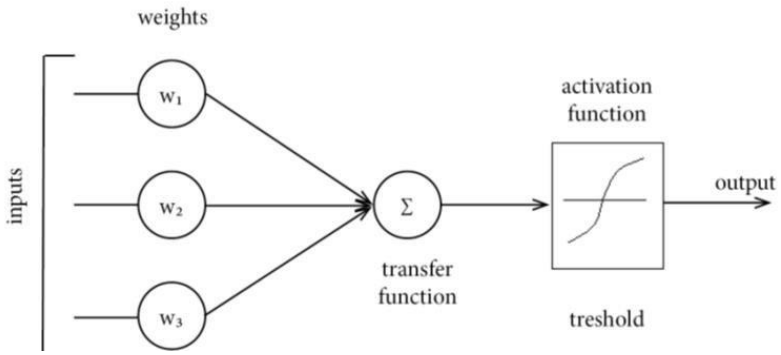

Figure 3.4: Basic elements of an Artificial Neuron

Every day, more health data is acquired, and it is required to analyze and understand this data to make sense of it and create value, which necessitates the use of advanced analytics. Big Data is now seen as being too vast and diverse to be stored and used effectively. While Big Data is useless as an asset, utilizing artificial intelligence (AI) to interpret Big Data to generate predictions or choices has the potential to transform existing health care procedures. Artificial intelligence algorithms can extract useful information from Big Data and apply it to healthcare advancements (Bohr & Memarzadeh, 2020).

An artificial neural network is a type of neural network that works by receiving inputs from other neurons and carrying out computational problems. An ANN is a neural network that consists of artificial neurons with numerous connections between them. It can be trained using both unsupervised and supervised learning. An ANN is a type of neural network that can predict an outcome based on a set of complex nonlinear relationships between two or more inputs.

Given the unpredictable course of the pandemic in which we live (COVID -19), the authorities have failed to understand how many resources they can muster even in the next week. In these difficult times, it is a great help to authorities if they can predict what kind of resources a person will need when they test positive or before, as they can obtain and organize the necessary resources. This can save the patient's life.

Health care systems around the world are facing unprecedented strain as the volume and acuity of patients hospitalized with COVID -19 has increased, while at the same time resources for patient care have been reduced due to interrupted supply chains (Vranas et al., 2021).

In Table 3.1 use of Artificial Neural Networks in COVID-19 literature is summarized with authors, years, types, and sources of publication.

Table 3.1: Some Artificial Neural Networks publications in COVID-19

Publication	**Type**	**Journal**
(Elhag et al., 2021)	Classification	Results in Physics
(Toğa et al., 2021)	Forecasting	Journal of Infection and Public Health
(Mohammadi et al., 2021)	Classification	Biomedical Journal
(Toraman et al., 2020)	Classification	Chaos, Solitons and Fractals
(Saba & Elsheikh, 2020)	Forecasting	Process Safety and Environmental Protection
(Aljaaf et al., 2021)	Forecasting	Journal of Biomedical Informatics

4. MODEL DEVELOPMENT

Artificial neural networks (ANN) are a mechanism that can be trained on a data set with features whose relationships can be relatively complex and make accurate predictions for a test set. ANNs, on the other hand, can be harmed by over-fitting and are sensitive to contradictory observations. As a result, the data set is preprocessed using data analysis. In theory, this technique allows for a reduction in the data set used to train the ANN without compromising performance (Misiunas et al., 2016).

4.1 The details of proposed ANN model

In Artificial Neural Network models, input node selection is an important step. It plays a major role in the success of the models. The first step in building an Artificial Neural Network model is to determine the output and input layers. In this step, the output layer consists of a single binary variable called ICU.

The goal of the ANN model is to categorize patients based on whether they would be admitted to intensive care, taking into account their initial features. The nodes in the input and output layers are exhibited first when developing an artificial neural network. The output layer consists of a single node since the model classifies patients based on their features to determine whether they will be admitted to intensive care. The intensive care value is the output layer as a binary variable. The input layer consists of the characteristics that remain after pre-processing.

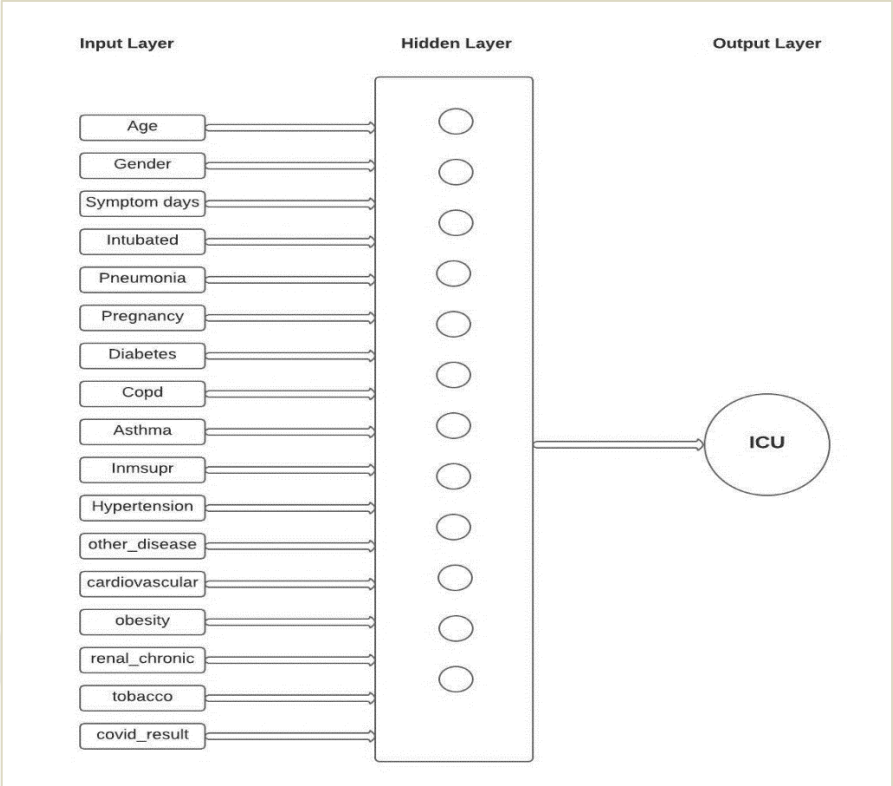

Figure 4.1: Model representation

The Figure 4.1 shows the artificial neural network used in the model. The model contains 17 input layers, 12 hidden nodes and 1 output layer. As a result of the output value, it is classified as 1 if the patient's ICU is required and 0 is if not. The example patient data used in the model are shown in Table 4.1.

Table 4.1: Sample Patient Data

id	gender	intubed	age	diabetes	asthma	covid_pos	covid_neg	icu
1beec8	1	0	47	1	0	1	0	1
077c96	1	1	66	0	0	1	0	1
0046aa	1	0	58	1	0	1	0	1
1d8440	0	0	31	0	0	1	0	1
03da5d	0	0	58	0	0	1	0	1

…			…			..	…	…
167386	0	0	54	0	0	1	0	0
0b5948	1	0	30	0	0	1	0	0
0d01b5	0	0	60	1	0	1	0	0
1beec8	1	0	47	1	0	1	0	1
1.75E+56	1	0	63	0	0	1	0	0

In the table, most of the values in the patient's medical records are binary values. A value of 1 is assigned if the patient has that disease and characteristic, and 0 otherwise. Gender is also represented numerically as 0 is female and 1 is male. Only two of the categories, age and how many days the patient had symptoms, remained unchanged.

4.2 Numerical application of ANN

The purpose of the mathematical application is to estimate if a patient requires an ICU based on the patient's features using artificial neural networks and the dataset given in previous sections. With this forecast, hospital resources, equipment, and medical staff may be allocated to more efficient locations, while timely diagnosis and treatment can save patients' lives.

Python 3.9 was used to create the model. As libraries, Keras and Tensorflow were employed. A total of 120,026 patient data points were assigned at random, with sections of 70%, 15%, and 15%. Data for training, validation, and testing include 84018, 18004, and 18004 cases, respectively.

The model is sequential. There are 17 inputs and just one output. Trial and error expanded the number of hidden nodes. There is no standard for achieving the optimum hidden layer.

N represents the number of input variables in the neural network, and the neural network of the neural network with at least N/2 hidden neurons is trained a certain number of times and tested with pre-selected input data for a certain period.

The optimal number is thus found by trial and error, and the number of hidden nodes determined for this model is 12.

After these processes are completed, the percentage of incorrect predictions is determined and recorded. The network with more hidden neurons is created by increasing the number of neural layers and is trained and tested over and over with the same pre-existing input data. For newly created neural networks, the learning process is repeated many times, each time starting with different random connection weights between neurons (Trifonov et al., 2018).

Input values are given to a neural network to be taught, and weight values are placed in the hidden layers during training. It generates a prediction of the new input value, which is fed into the model once training is completed. To strengthen the model's performance, the parameters in the hidden layers are adjusted using the training backscatter approach.

4.2.1 Reading Data

The first step for the predictive model is to read the data to be used as input. In this example, the data set coviddataset is used. To get started and read the data, the Pandas library was used.

4.2.2 Splitting the dataset into input and target

Then it is necessary to separate the dataset into input (train_X) and target (train_y). As input, there are 17 other columns except ICU , because after the model is trained, the value we try to estimate will be the required or not required value in the column ICU. Therefore, ICU is output.

4.2.3 Splitting the dataset into training dataset and testing dataset

In total, there are many data. 30% of this data has been split as test and validation data and the remaining 70% has been split as training data, for which the train_test_split function of Sklearn library has been used.

4.2.4 Building the Model

After the preliminary steps are completed, the next step is to create a model. The type of model to use is the sequential model. Sequential is the easiest way to create a model in Keras. Layer by layer, we can create a model. Each layer has weights that specify the following layer. The add () function is used to add layers to the model. There are 17 input layers, 12 hidden neurons and one output layer added.

Dense is the type of the layer. Dense is a default type of layer that works in most cases. In a dense layer, all nodes of the previous layer are connected to nodes of the current layer.

The number of neurons in a layer can be hundreds or even thousands. Increasing the number of nodes on each layer increases the model capacity as well as the training time and the storage space occupied by the model on disk, which leads to undesirable situations. For this reason, an attempt is made to find the optimal number of neurons and hidden layers. It is impossible to say exactly how many hidden layers there will be but adding tens or hundreds of layers in a row for a small dataset does more harm than good. It is impossible to predict how many neurons there will be in a hidden layer, but there are several approaches to this issue in the literature.

The activation function used for hidden layers is ReLU or Rectified Linear Activation. This activation function takes values less than zero as zero and values greater than zero as is. It is written in a formula as f (x) = max (0, x). For the output layer, the Sigmoid Function is chosen.

4.2.5 Compiling the model

The next process is compiling the model. Compiling the model is done using the model.compile () function. This function takes three parameters: Optimizer, Loss, and Metric. It takes many parameters, but three of them are used in this work.

The optimizer controls the learning rate. Adam is used as the optimizer. Adam is generally a good optimization algorithm used in many situations. The Adam algorithm adjusts the learning rate throughout the training.

The learning rate determines how fast the optimal weights for the model are calculated. A smaller learning rate can result in more accurate and good weights (up to a certain point), which means that the model learns better, but the training time becomes longer because the time needed to calculate the weights is longer.

The metric accuracy is used to see the amount of accuracy and loss achieved by the validation set at the end of each period to interpret how the model performed during training.

4.2.6 Training the model

To train the model, the function fit () is used with the following parameters: Training data (X_train), target data (y_train), validation set (validation_data), amount of data to be trained simultaneously (batch_size), and the number of periods that the data set should be used in the model. The number of periods that determine that they will be trained through. The batch size indicates how many data sets will be processed simultaneously when training the model. In our model, the batch size 256 is set, which means 256 data will go into training at the same time, then the error rate will be calculated using the optimization function and the next 256 data will be run through. This process continues until all the data is processed. After all the data is processed, it moves to the next epoch. If the batch size is low, the training time will be long because the data requires more optimization calculations. If this number is large, it may affect the performance of the model while the training time remains relatively short. This number is usually the exponential multiple of two, so numbers like 2,4,8,16,32...

The epoch number indicates how many times the data passes through the model while the model is being trained. If this value is a small number, the training time is short and the model's performance may not be fully developed, while if the number is large, the training time is too long and the model may have already completed its development, i.e., it may

have been trained unnecessarily much. The best method for this is to set a large number and stop training at the point where the model has completed its development, this is called early stopping.

As can be seen in the Figure 4.2, the epoch number was set at 400 because after 400 it was seen that the value remained constant for 50 epochs, and it was decided that it should be truncated at 400 for the model to work efficiently.

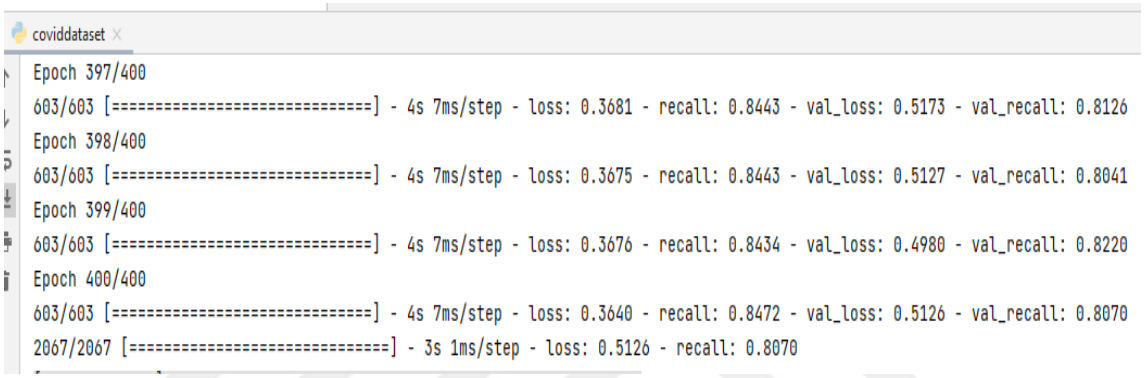

Figure 4.2: Training stages of the model

4.2.7 Making predictions with new data

The model. predict () function is used to create predictions when given new data as input. Any test data can be chosen and used as input. For this model, however, different pairs of tests were performed. These results are then graphically represented in Figure 5.2.

4.2.8 Parameter Settings

We conducted trial and error to find the optimal parameter values in the model. According to the Table 4.2, the best results were obtained when the epoch number was 300 and the neuron number in the hidden layer was 12.

Table 4.2: Parameter settings

Trial-Error No	Epoch number	Hidden Layers Neurons	Accuracy	Precision	Recall	F1 score
1	300	17	79,13%	79,09%	79,20%	79,14%
2	300	12	79,30%	78,20%	81,25%	79,70%
3	300	9	79,28%	79,02%	79,80%	79,41%
4	300	7	79,25%	78,82%	81,89%	80,33%

The confusion matrices of these trial and error are shown in the Figure 4.3.

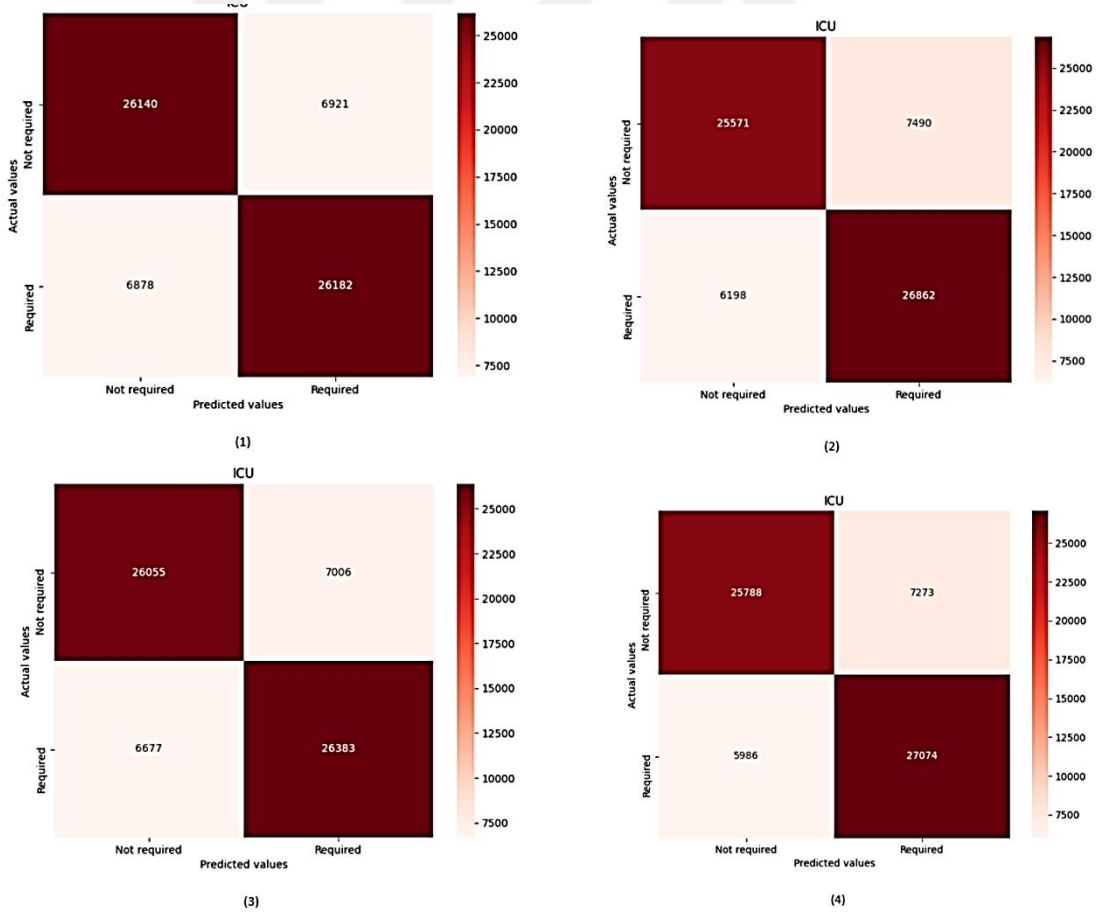

Figure 4.3: Confusion matrices of trial-errors

5. RESULTS AND DISCUSSION

This section summarizes the study's findings. Artificial neural networks are computational models that mimic human neurons to capture information and make decisions. As a supervised learning approach, the backpropagation method was applied. Python software was used to implement the predictive models. The number of neural nodes in the hidden layer of the model is determined experimentally and iteratively based on performance indicators. According to the applications, the number of neural nodes in the model architecture's hidden layer was 12. A 17x12x1 ANN model with 17 input layers, 12 hidden nodes, and 1 output layer was used. The original data set of 120,026 patients is randomly divided into three parts: 70%, 15%, and 15%. The model's prediction accuracy is 79 percent. When the ANN model is applied to the existing problem, the following results are achieved, as shown below. Table 5.1 displays the ANN model's performance metrics.

Table 5.1: ANN performance metrics

Class	Accuracy	Precision	Recall	F1 Score
0	79%	0.80	0.78	0.79
1	79%	0.78	0.81	0.80

Classification accuracy (ACC): Classification accuracy is one of the most extensively used metrics in evaluating the effectiveness of classification algorithms, and it is the measure of all properly described cases. When all classes are equally important, it is most usually utilized.

The following equation gives the ratio of actual positives and actual negatives obtained by the classifiers in the total number of samples:

$$ACC = \frac{TN+TP}{TP+FP+FN+TN} \tag{5.1}$$

where TP, FP, FN, TN denote the number of true positives, false positives, false negatives, and true negatives.

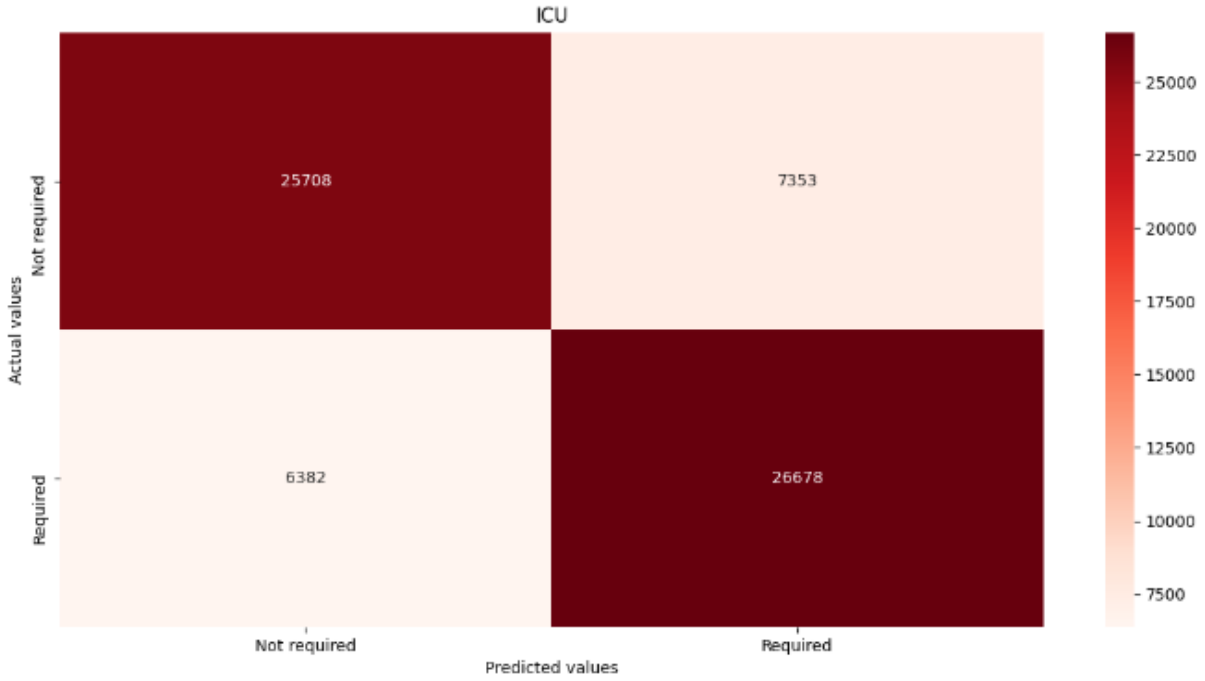

Figure 5.1: Graphical representation of the result

Regarding this illustration:

This ANN-based model predicts if the patient will require intensive care with excellent accuracy. The algorithm is more accurate in predicting patients who require ICU at real-world values than patients who do not require ICU.

Recall: It measures the accuracy of the model in predicting all actual objects.

$$Recall = \frac{TP}{TP+FN} = 78\% \quad (5.2)$$

Precision: It indicates the percentage of true positives out of the total number of positive predictions.

$$Precision = \frac{TP}{TP+FP} = 80\% \quad (5.3)$$

F1-score: This is the harmonic mean of Precision and Recall, and it provides a more accurate estimate of poorly classified cases than the accuracy metric.

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} = 79\% \quad (5.4)$$

Figure 5.2 depicts one of the performance graphs displaying the trained model's success rate during training. The following is a graph of training and validation losses:

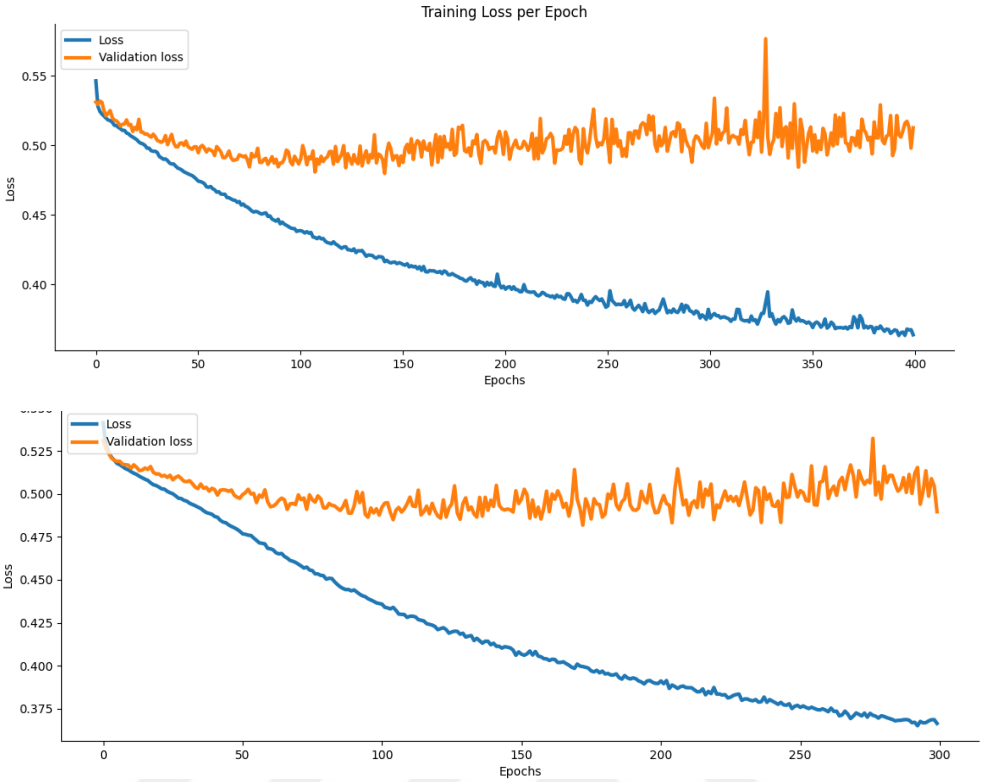

Figure 5.2: Schematic representation of training stages for epoch 300 and 400

There is an obvious healthy association between training loss and validation loss in this diagram. Both appear to diminish to a certain point and then remain constant. This shows that the model has been well trained and performs equally well on both training and latent data. The loss value diminishes as the period lengthens. The distinction between training and validation loss, on the other hand, is striking. The validation loss was larger than the training loss. As the number of period values rises, so does the difference in loss between training and validation. The loss reflects how well the model comprehends the problem. In this graph, we can observe that as the loss diminishes, so does the training process.

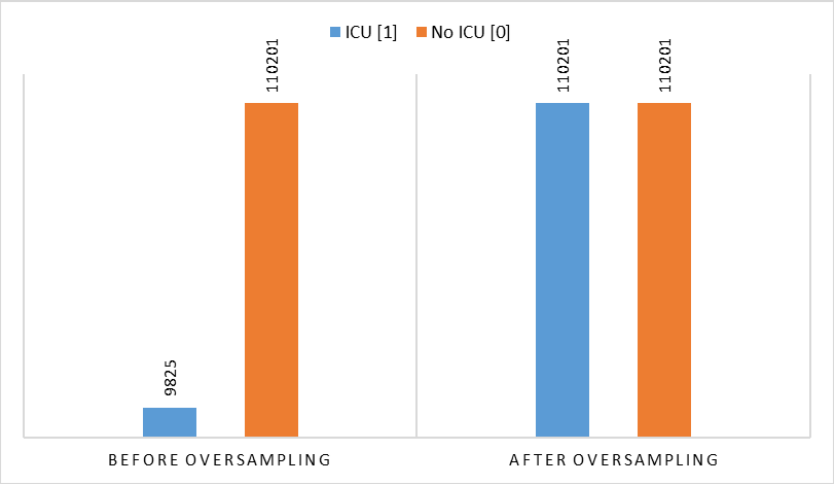

Figure 5.3: Distribution of patients regarding the need for ICU

The imbalance distribution among the 2 classes is balanced with the function SMOTE and better performance results are obtained. The ICU values earlier than oversampling have been 0: 110201, 1:9825, and the value of two classes have been synchronized with 110201 after oversampling copies of the created minority class as shown in Figure 5.3.

5.1 Discussion

It should be emphasized that the patient data extracted from the accessible database was very imbalanced. Of the 120,026 patients in the database, 9,825 required intensive care. The percentage of patients requiring ICU admission was 9%. This means that the model is handling an imbalanced class set.

With a 79 percent accuracy rate, the suggested ANN model predicted whether the patient would be admitted to intensive care unit. When evaluating the performance of models that cope with an unbalanced dataset, accuracy isn't the only measure to evaluate. Numerous machine learning techniques cannot give reasonable results with an unequal of classes when faced with unbalanced datasets. When attempting to reduce the total error rate, the resultant classifiers may overlook the minority class by classifying the dominating class. When working with uneven datasets, appropriately classifying the minority class becomes increasingly difficult.

If the data is significantly imbalanced, using the Synthetic Minority Oversampling Technique (SMOTE) increases the sample size of the minority class and improves the performance of the classification model ML. In this study, we used Python 3.9 software to build an ANN model and run the process. SMOTE is an important and useful oversampling method for balancing unbalanced data by interpolating the characteristics of randomly selected instances from existing monitor classes to generate a composite example of a minority class (Chawla et al., 2002). SMOTE is one of the most commonly used oversampling approaches to handle imbalanced classes, often improving the performance and generalization capabilities of ML models when raw data is imbalanced and it has been shown in studies (Gao & Elzarka, 2021; Tasmektepligil & Gunpinar, 2022).

In fact, one of the important contributions of this study is that it helps in systematic decision-making and management of resources in similar pandemic situations such as this one in the future together with the ANN model used. Since the intensive care units are the bottleneck of health facilities in the most difficult times of the COVID-19 pandemic, the need for the intensive care units was much more felt. During this period, excessive health workers had to make critical decisions about the lives of patients. Since ICU numbers are limited, this can be considered a source allocation problem. As in the world, many patients were waiting in ICU when they needed because of COVID-19 in Turkey. To avoid this situation in the future, this model can be predicted whether patients need ICU especially during the epidemic period. In this way, as a preparation for the inadequate ICU numbers that exist, hospital departments will be transformed into the intensive care unit urgently, and the medical staff working in those departments can receive the necessary training in advance and be prepared for a new epidemic.

6. CONCLUSION

Throughout the pandemic, there were times when the number of hospitalizations, positive cases, patients requiring intubation, and patients seeking medical attention was massive, and the disease was extremely serious. Physicians, medical staff especially ICU personnel were overworked, and hospital resources were frequently underutilized. When medical departments were overcrowded, clinicians made difficult decisions about patients in the early stages of the pandemic to save many lives. The ability to make quick and precise decisions is critical in such situations for effective resource management. Artificial intelligence and its importance in healthcare have been reviewed with this foresight and studies on this topic have been taken up.

6.1 Thesis Contribution

This study discussed the use of prescriptive analytics in epidemiological studies with COVID-19 cases. First, risk factors are investigated and defined, then the data is analyzed to create predictive models for future cases. The main reason to use prescriptive analytics is to take advantage of the increasing amount of data and provide practitioners with insights to make more informed decisions. The proposed decision support framework predicts a patient's ICU needs based on the information gathered at admission. Its 79% accuracy allows the model to assist decision-makers. This foresight can guide hospital resources, equipment, and medical staff to more efficient locations and save patients' lives through timely diagnosis and treatment. This allows hospital managers and governments to take precautions to contain the outbreak before planning resources.

6.2 Limitations and Future Work

This study, however, has certain drawbacks. First, the data has an unbalanced structure, needing a more in-depth assessment of performance measures. To balance the data set and examine the outcomes, another function other than SMOTE may be used. Second, acquiring patient data, which contains patient-specific information regarding COVID-19 patients' medical histories and behaviors, is challenging and legally protected. This is due to regulatory security requirements that prevent anybody from accessing legally protected health information (PHI). Because of the large quantity of missing data, the sample is reduced. Also, the risk factors explored in the literature have a higher amount, which may enhance the model further.

This pandemic shows that our healthcare system needs to be improved. In future investigations, we can identify the most relevant factors among the different causal factors that raise the probability of COVID19 infection. The relative impact of risk factors on COVID 19 cases and mortality can be estimated. The investigation was conducted using open access data, so the available attributes were limited. By including expert opinion in the ANN approach, new models can be developed, and new risk factors can be included.

Collecting data on critical medical conditions other than medical history can supplement these risk factors, which may also be collected from patients. By collecting more data using this approach, we can generate better insights. In other machine learning and neural network (CNN, RNN) models, Bayesian networks can be evaluated by comparing their accuracy to larger datasets.

REFERENCES

Aabed, K., & Lashin, M. M. A. (2021). Saudi Journal of Biological Sciences An analytical study of the factors that influence COVID-19 spread. *Saudi Journal of Biological Sciences*, *28*(2), 1177–1195. https://doi.org/10.1016/j.sjbs.2020.11.067

Albayati, N., Waisi, B., Al-furaiji, M., Kadhom, M., & Alalwan, H. (2021). Effect of COVID-19 on air quality and pollution in different countries. *Journal of Transport & Health*, *21*(March), 101061. https://doi.org/10.1016/j.jth.2021.101061

Alger, H. M., Williams, J. H., Walchok, J. G., Bolles, M., Fonarow, G. C., & Rutan, C. (2020). Role of Data Registries in the Time of COVID-19. In *Circulation: Cardiovascular Quality and Outcomes*. Lippincott Williams and Wilkins. https://doi.org/10.1161/CIRCOUTCOMES.120.006766

Aljaaf, A. J., Mohsin, T. M., Al-Jumeily, D., & Alloghani, M. (2021). A fusion of data science and feed-forward neural network-based modelling of COVID-19 outbreak forecasting in IRAQ. *Journal of Biomedical Informatics*, *118*(April), 103766. https://doi.org/10.1016/j.jbi.2021.103766

Allam, M., Cai, S., Ganesh, S., Venkatesan, M., Doodhwala, S., Song, Z., Hu, T., Kumar, A., Heit, J., & Coskun, A. F. (2020). COVID-19 diagnostics, tools, and prevention. In *Diagnostics* (Vol. 10, Issue 6). MDPI AG. https://doi.org/10.3390/diagnostics10060409

Azoulay, É., Beloucif, S., Beloucif, S., Guidet, B., Guidet, B., Pateron, D., Pateron, D., Vivien, B., Vivien, B., Le Dorze, M., & Le Dorze, M. (2020). Admission decisions to intensive care units in the context of the major COVID-19 outbreak: Local guidance from the COVID-19 Paris-region area. *Critical Care*, *24*(1), 1–6. https://doi.org/10.1186/s13054-020-03021-2

Blessy Trencia Lincy, S. S., & Suresh Kum, N. (2018). Transforming healthcare via big data analytics. In *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*. Elsevier Inc. https://doi.org/10.1016/B978-

0-12-813314-9.00015-3

Bohr, A., & Memarzadeh, K. (2020). Current healthcare, big data, and machine learning. In *Artificial Intelligence in Healthcare*.https://doi.org/10.1016/b978-0-12-818438-7.00001-0

Brandt, J. S., Hill, J., Reddy, A., Schuster, M., Patrick, H. S., Rosen, T., Sauer, M. V., Boyle, C., & Ananth, C. V. (2020). Epidemiology of coronavirus disease 2019 in pregnancy: risk factors and associations with adverse maternal and neonatal outcomes. *American Journal of Obstetrics and Gynecology*, 1–9. https://doi.org/10.1016/j.ajog.2020.09.043

Cao, W., Chen, C., Li, M., Nie, R., Lu, Q., Song, D., Li, S., Yang, T., Liu, Y., Du, B., & Wang, X. (2021). Important factors affecting COVID-19 transmission and fatality in metropolises. *Public Health*, *190*, e21–e23. https://doi.org/10.1016/j.puhe.2020.11.008

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE : Synthetic Minority Over-sampling Technique. *Artificial Intelligence Research*, *16*, 321–357. https://doi.org/https://doi.org/10.1613/jair.953

Elhag, A. A., Aloafi, T. A., Jawa, T. M., Sayed-Ahmed, N., Bayones, F. S., & Bouslimi, J. (2021). Artificial neural networks and statistical models for optimization studying COVID-19. *Results in Physics*, *25*, 104274. https://doi.org/10.1016/j.rinp.2021.104274

Gao, C., & Elzarka, H. (2021). Advanced Engineering Informatics The use of decision tree based predictive models for improving the culvert inspection process. *Advanced Engineering Informatics*, *47*(May 2020), 101203. https://doi.org/10.1016/j.aei.2020.101203

Hao, Y., Xu, T., Hu, H., Wang, P., & Bai, Y. (2020). Prediction and analysis of Corona virus disease 2019. *PLoS ONE*, *15*(10 October). https://doi.org/10.1371/journal.pone.0239960

Holzinger, A. (2016). Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, *3*(2), 119–131. https://doi.org/10.1007/s40708-016-0042-6

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-

000101

Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). International Journal of Information Management Prescriptive analytics : Literature review and research challenges. *International Journal of Information Management*, *50*(April 2019), 57–70. https://doi.org/10.1016/j.ijinfomgt.2019.04.003

Levy, T., Richardson, S., Coppa, K., Barnaby, D., McGinn, T., Becker, L., Davidson, K., Cohen, S., Hirsch, J., & Zanos, T. (2020). Development and Validation of a Survival Calculator for Hospitalized Patients with COVID-19. *MedRxiv : The Preprint Server for Health Sciences*. https://doi.org/10.1101/2020.04.22.20075416

Lopes, J., Guimarães, T., Santos, M. F., Lopes, J., Guimarães, T., & Santos, M. F. (2020). ScienceDirect ScienceDirect Predictive and Prescriptive Analytics in Healthcare : A Survey Predictive and Prescriptive Analytics in Healthcare : A Survey. *Procedia Computer Science*, *170*, 1029–1034. https://doi.org/10.1016/j.procs.2020.03.078

Medicine, P., & Wars, W. (2021). *Personalized Medicine in Psychiatry impact of the pandemic on mental health*. *28*, 25–28. https://doi.org/10.1016/j.pmip.2021.100077

Mehta, N., Pandit, A., & Shukla, S. (2019). Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *Journal of Biomedical Informatics*, *100*(October), 103311. https://doi.org/10.1016/j.jbi.2019.103311

Misiunas, N., Oztekin, A., Chen, Y., & Chandra, K. (2016). DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. *Omega (United Kingdom)*, *58*, 46–54. https://doi.org/10.1016/j.omega.2015.03.010

Mohammadi, F., Pourzamani, H., Karimi, H., Mohammadi, M., Mohammadi, M., Ardalan, N., Khoshravesh, R., Pooresmaeil, H., Shahabi, S., Sabahi, M., Sadat miryonesi, F., Najafi, M., Yavari, Z., Mohammadi, F., Teiri, H., & Jannati, M. (2021). Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran. *Biomedical Journal*, *March*. https://doi.org/10.1016/j.bj.2021.02.006

Mosavi, N. S., & Santos, M. F. (2020). How prescriptive analytics influences decision making in precision medicine. *Procedia Computer Science*, *177*, 528–533. https://doi.org/10.1016/j.procs.2020.10.073

Oakes, M. C., Kernberg, A. S., Carter, E. B., Foeller, M. E., Palanisamy, A., Raghuraman, N., & Kelly, J. C. (2021). Pregnancy as a risk factor for severe coronavirus 2019 (COVID-19) disease using standardized clinical criteria. *American Journal of Obstetrics & Gynecology MFM*, *3*(3), 100319. https://doi.org/10.1016/j.ajogmf.2021.100319

Poornima, S., & Pushpalatha, M. (2020). A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks*, *1*(August), 76–84. https://doi.org/10.1016/j.ijin.2020.07.001

R. Wadhera, P. Wadhera, Gaba, Figueroa, Yeh, Shen, J. M. (2020). Characteristics of Hospitalized Adults With COVID-19 in an Integrated Health Care System in California. *JAMA - Journal of the American Medical Association*, *323*(21), 2–5. https://doi.org/10.1001/jama.2020.0757

Roedl, K., Jarczak, D., Thasler, L., Bachmann, M., Schulte, F., Bein, B., Weber, C. F., Schäfer, U., Veit, C., Hauber, H. P., Kopp, S., Sydow, K., de Weerth, A., Bota, M., Schreiber, R., Detsch, O., Rogmann, J. P., Frings, D., Sensen, B., … Kluge, S. (2020). Mechanical ventilation and mortality among 223 critically ill patients with coronavirus disease 2019: A multicentric study in Germany. *Australian Critical Care*, *xxxx*. https://doi.org/10.1016/j.aucc.2020.10.009

Saba, A. I., & Elsheikh, A. H. (2020). Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Safety and Environmental Protection*, *141*, 1–8. https://doi.org/10.1016/j.psep.2020.05.029

Schwartz, I. M., Nowakowski-sims, E., Ramos-hernandez, A., & York, P. (2017). CE Do Not Cite or Quote Without Permission. *Children and Youth Services Review*. https://doi.org/10.1016/j.childyouth.2017.08.020

Srinivas, S., & Ravindran, A. R. (2018). Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. *Expert Systems with Applications*, *102*, 245–261. https://doi.org/10.1016/j.eswa.2018.02.022

Summers, J., Cheng, H., Lin, H., Telfar, L., Kvalsvig, A., Wilson, N., & Baker, M. G. (2020). *The Lancet Regional Health - Western Pacific Potential lessons from the Taiwan and New Zealand health responses*. *4*(August). https://doi.org/10.1016/j.lanwpc.2020.100044

Tasmektepligil, A. A., & Gunpinar, E. (2022). Advanced Engineering Informatics SplineLearner : Generative learning system of design constraints for models represented using B-spline surfaces. *Advanced Engineering Informatics*, *51*(December 2021), 101478. https://doi.org/10.1016/j.aei.2021.101478

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. In *American Journal of Hematology* (Vol. 95, Issue 7, pp. 834–847). Wiley-Liss Inc. https://doi.org/10.1002/ajh.25829

Toğa, G., Atalay, B., & Toksari, M. D. (2021). COVID-19 Prevalence Forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. *Journal of Infection and Public Health*. https://doi.org/10.1016/j.jiph.2021.04.015

Toraman, S., Alakus, T. B., & Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons and Fractals*, *140*. https://doi.org/10.1016/j.chaos.2020.110122

Trewartha, A., Dagdelen, J., Huo, H., Cruse, K., Wang, Z., He, T., Subramanian, A., Fei, Y., Justus, B., Persson, K., & Ceder, G. (2020). COVIDScholar: An automated COVID-19 research aggregation and analysis platform. *ArXiv*, 1–25.

Tyrrell, C. S. B., Mytton, O. T., Gentry, S. V., Thomas-Meyer, M., Allen, J. L. Y., Narula, A. A., McGrath, B., Lupton, M., Broadbent, J., Ahmed, A., Mavrodaris, A., & Abdul Pari, A. A. (2021). Managing intensive care admissions when there are not enough beds during the COVID-19 pandemic: A systematic review. *Thorax*, *76*(3), 302–312. https://doi.org/10.1136/thoraxjnl-2020-215518

Vepa, A., Saleem, A., Rakhshan, K., Daneshkhah, A., Sedighi, T., Shohaimi, S., Omar, A., Salari, N., Chatrabgoun, O., Dharmaraj, D., Sami, J., Parekh, S., Ibrahim, M., Raza, M., Kapila, P., & Chakrabarti, P. (2021). *Using Machine Learning Algorithms to Develop a Clinical Decision-Making Tool for COVID-19 Inpatients*. *March 2020*, 1–22.

Vranas, K. C., Golden, S. E., Mathews, K. S., Schutz, A., Valley, T. S., Duggal, A., Seitz, K. P., Chang, S. Y., Nugent, S., Slatore, C. G., Sullivan, D. R., & Hough, C. L. (2021). The Influence of the COVID-19 Pandemic on ICU Organization, Care

Processes, and Frontline Clinician Experiences: A Qualitative Study. *Chest*. https://doi.org/10.1016/j.chest.2021.05.041

Wang, Q. Q., Xu, R., & Volkow, N. D. (2021). Increased risk of COVID-19 infection and mortality in people with mental disorders: analysis from electronic health records in the United States. *World Psychiatry*, *20*(1), 124–130. https://doi.org/10.1002/wps.20806

Wooding, D. J., & Bach, H. (2020). Treatment of COVID-19 with convalescent plasma: lessons from past coronavirus outbreaks. *Clinical Microbiology and Infection*, *26*(10), 1436–1446. https://doi.org/10.1016/j.cmi.2020.08.005

Xu, Z., Su, C., Xiao, Y., & Wang, F. (2021). Artificial intelligence for COVID-19: battling the pandemic with computational intelligence. *Intelligent Medicine*, *2*(1), 13–29. https://doi.org/10.1016/j.imed.2021.09.001

Yadaw, A. S., Li, Y.-C., Bose, S., Iyengar, R., Bunyavanich, S., & Pandey, G. (2020). Clinical predictors of COVID-19 mortality. *MedRxiv : The Preprint Server for Health Sciences*. https://doi.org/10.1101/2020.05.19.20103036

Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., & Peng, J. (2020). *Since January 2020*. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of Infection and Public Health, 81, e16-e25.*

BIOGRAPHICAL SKETCH

Yeliz Çotoy studied at Şişli Anatolian High School, from which she graduated in 2013. In the same year, she started her undergraduate studies at Doğuş University, Department of Industrial Engineering, and graduated in 2018. She also has a second major in Mechanical Engineering. She started her postgraduate studies at Galatasaray University, Department of Industrial Engineering in 2019. She is about to finish her Master of Science Degree in the Industrial Engineering at Galatasaray University Institute of Science. She plans to continue her career in Industrial Engineering Ph.D. program at Galatasaray University Institute of Science. She is a research assistant at Doğuş University in the Industrial Engineering Department. Her research interests are decision making, artificial intelligence, machine learning, and decision support systems.

PUBLICATIONS

- Cotoy Y., Ozaydin O. (2019) Using GeoMarketing criteria for Retail Site Location Selection Problem: Case of a Sports-Goods Retailer in Turkey EURO 2019, 30th European Conference on Operational Research, Dublin, Ireland.
- Çotoy Y., Özaydın Ö. (2019) Analitik Hiyerarşi Sürecine Dayalı TOPSIS ve VİKOR Yöntemleri ile bir Spor Ürünleri Perakendecisi için Mağaza Yeri Seçimi