NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF SOFTWARE ENGINEERING

CORONAVIRUS DETECTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS

M.Sc. THESIS

Mohammadreza KAKHBOD

Nicosia

February, 2022

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF SOFTWARE ENGINEERING

CORONAVIRUS DETECTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS

M.Sc. THESIS

Mohammadreza KAKHBOD

Supervisor

Assoc.Prof.Dr. Boran ŞEKEROĞLU

Nicosia

February, 2022

Approval

We certify that we have read the thesis submitted by Mohammadreza Kakhbod titled **"CORONAVIRUS DETECTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS"** and that, in our combined opinion, it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

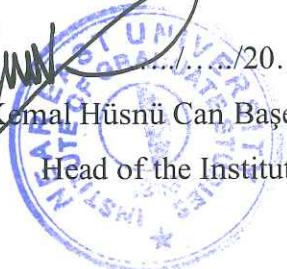| Examining Committee	Name-Surname	Signature
Head of the Committee:	Assoc.Prof.Dr. Yöney K. Ever	
Committee Member*:	Assoc.Prof.Dr. Kamil Dimililer	
Supervisor:	Assoc.Prof.Dr. Boran Şekeroğlu	

Approved by the Head of the Department

11/02/2022

Assoc.Prof.Dr. Yöney Kırsal Ever

Head of Department

Approved by the Institute of Graduate Studies

..../..../20...

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

Declaration

I hereby declare that all information, documents, analysis, and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of the Institute of Graduate Studies, Near East University. I also declare that, as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Mohamadreza Kakhbod

2022/02/11

Acknowledgments

I would like to thank my supervisor, Assoc.Prof.Dr. Boran Sekeroglu, who helped me a lot along this path, and due to my illness, understood me and was with me until the end, and I will never Forget dr. Boran's kindness. Many thanks to all the Professors of the Faculty of Engineering, including Assoc.Prof.Dr. Yoney K. Ever, Assist.Prof.Dr. Kaan Uyar, and Assist.Prof.Dr. Umit Ilhan. Finally, I would like to thank NEU university for its great contribution to the advancement of science and technology with its good facilities and environment, and I was honored to study at this university.

Mohammadreza Kakhbod

Dedication to my dear mother, who always prayed for me from a distance and helped me a lot

Abstract

CORONAVIRUS DETECTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS

Mohamadreza Kakhbod

MSc., Department of Software Engineering

February, 2022, 82 pages

Due to the high rate of spread of the coronavirus, the increase in the number of patients with this disease makes it necessary to supply fast and accurate methods to diagnose people infected with this virus. There are various methods, like PCR testing, CT scan imaging, etc., to screen people infected with COVID19, which are time-consuming, so providing a faster method can speed up the quarantine process of infected people and inhibit the spread of the virus. Therefore, during this study, epidemiological data were accustomed diagnose people with COVID19. The desired data were obtained from the database. Four different machine learning models were implemented, including an artificial neural network (with three forms of activation functions), a call tree, the closest neighbor, and a support vector machine (with three forms of kernels). The results showed that the ultimate accuracy, sensitivity, and accuracy of the choice tree in diagnosing people with COVID19 were 92.94%, 83.02%, and 98.69%, respectively. Also, the general accuracy of the three NN methods in classifying epidemiological data was 80.59%, which had a sensitivity of 68.73% in diagnosing patients with COVID19. In addition, the accuracy of the support vector machine for linear kernels, Gaussian radial basis function, and polynomials were 90.09%, 93.44%, and 96.73%, respectively. Overall, the results showed that the most effective diagnostic tool for coronary cardiovascular disease is a polynomial support-vector machine.

Key Words: Artificial Intelligence, Machine Learning, COVID19, Support Vector Machine

Özet

Covid 19 ile enfekte olmuş kişileri taramak için PCR testi, CT tarama görüntüleme vb. gibi zaman alıcı çeşitli yöntemler vardır. Bu nedenle daha hızlı bir yöntem sağlamak enfekte kişilerin karantina sürecini hızlandırabilir ve virüsün yayılmasını önlemeye yardımcı olabilir. Bu nedenle bu çalışma sırasında COVID19 enfekte olmuş kişilerin teşhisi için epidemiyolojik veriler kullanıldı. Veri tabanından istenilen veriler elde edildi. Yapay sinir ağı (üç tür aktivasyon fonksiyonu ile), bir çağrı ağacı, en yakın komşu ve bir destek vektör makinesi (üç çekirdek şekilli) dahil olmak üzere dört farklı makine öğrenme modeli uygulandı. Sonuçlar, seçilen ağacın COVID19 enfekte olmuş kişileri teşhis etmedeki maksimum doğruluğunun, hassasiyetinin ve doğruluğunun sırasıyla %92.94, %83.02 ve %98.69 olduğunu gösterdi. Ayrıca 3-NN yönteminin epidemiyolojik verileri sınıflandırmadaki genel doğruluğu %80,59 olup, Covid 19 hastalarının tanısında %68,73 hassasiyete sahipti. Ayrıca lineer çekirdekler, Gauss radyal fonksiyon tabanı ve polinomlar için destek vektör makinesinin nihai doğruluğu sırasıyla %90.09, %93.44 ve %96.73 olmuştur. Genel olarak, sonuçlar, koroner kalp hastalığı veya epidemiyoloji teşhisi için en etkili aracın polinom destek vektörünün kullanılması olduğunu göstermiştir.

Anahtar Kelimeler: Yapay Zeka, Makine Öğrenimi, COVID19, Destek Vektör Makinesi.

Table of Contents

CHAPTER I

CHAPTER II

List of Tables

List of Figures

<div align="center">

CHAPTER I

Introduction

</div>

Introduction

In December 2019, the global coronavirus epidemic broke out in Wuhan, China, causing the world a serious health problem (WHO, 2021). The virus, which caused the worldwide spread of COVID-19, was called coronavirus acute respiratory syndrome. Coronaviruses are a large family of viruses that lead to diseases such as Middle East Respiratory Syndrome (Coronavirus Mers) and Acute Respiratory Syndrome (Coronavirus SARS). Coronavirus is a new species that was discovered in 2019 that was not previously detected in humans. Diseases of the coronavirus family are common to humans and animals, with some studies showing that the SARS coronavirus has been transmitted from cats to humans and the coronavirus Mers from unicorn camels to humans. However, in Covid 19, human-to-human transmission of the disease through the respiratory tract has accelerated the spread of the epidemic.

Problem Statement

The symptoms of Covid 19 are mild in 82% of cases but severe or severe in the remaining cases. The total number of coronavirus cases is approximately 6,371,661, including 377,552 deaths and 636,629,922,922 cases of improvement. The results showed that about 95% of patients with this disease survived without any problems, and the remaining 5% were in a severe or critical state (WHO, 2021).

Symptoms include respiratory symptoms, fever, cough, and shortness of breath. The disease can cause pneumonia, acute respiratory syndrome, infectious shock, organ failure, and death in more severe cases. Studies have shown that men are more likely to get the disease than women, and no deaths have been reported in children between the ages of 0 and 9.

According to the most recent Chinese government guidelines, the diagnosis of Covid 19 must be confirmed by gene sequencing testing on respiratory or blood samples, which may well be a key sign of a reverse transcription-polymerase chain reaction or hospitalization. Within the current context of public health, the problem is so

important that Covid 19 patients are not identified quickly, so there is no appropriate treatment. Additionally, because this virus is inherently highly contagious, these patients may infect large populations if not identified.

One of the solutions recently proposed by researchers is the use of CT scan imaging of the chest area, which might be wont to identify patients with Covid 19 more quickly. During this method, rather than looking ahead to patients to check positive for the virus, all people whose pneumonia pattern is clear on chest imaging are now considered as Covid 19 infected. With this method, the authorities are going to be ready to quarantine and treat patients much faster, but the matter with this method is the time-consuming and expensive process of performing a CT scan, which leads the researcher to seek out the simplest way to diagnose coronary arteries more quickly—encouraged that one in each of these methods is the use of epidemiological data. Therefore, the present study aims to produce a technique for diagnosing coronary cardiovascular disease using data and machine learning methods.

Coronavirus is a type of virus that causes lung infection. Coronavirus diseases in humans range from a gentle cold to severe illnesses like respiratory syndromes, which might be fatal. The prevalence of coronavirus is throughout the system. Therefore, the disease can spread rapidly among people, and the incidence increases. On the opposite hand, in some cases, this disease causes death in about two to five, and some Covid 19 patients will need to spend the remainder of their lives with permanent lung damage (Yong et al., 2020). This is because the COVID19, like the SARS virus, makes honeycomb-like cavities in the lungs (WHO, 2021). Accordingly, if rapid methods do not seem to be provided to diagnose this disease, the entire world is going to be infected with this disease quickly, and because of the high number of infected people, not everyone can receive treatment, and also the death rate of this disease will increase. One of the methods wont to diagnose people with COVID19 is the reverse transcription-polymerase chain reaction method; however, it is time-consuming (Raban et al., 2020). Therefore, it is necessary to produce away with high speed and high accuracy for diagnosing patients with Covid 19 (Hussain et al., 2021).

Innovation and novelty of the thesşs

An innovative aspect of this research is that the rapid diagnosis of coronary cardiopathy supported epidemiological data, which has not been exhausted by domestic research to diagnose coronary cardiovascular disease. Therefore, the innovative aspect of this research is the use of epidemiological data. On the other hand, a comprehensive study of machine learning methods for corona detection supported by this data set is another new aspect of this research that seeks out the simplest car windshield model for corona detection.

Research background

Preliminary studies in the field of COVID19

Screening of an outsized number of suspected cases followed by appropriate drug and quarantine is required to limit the transmission of Covid 19. The PT-PCR test is considered the gold standard of the test, but it has yielded significant false-negative results. Therefore, efficient and rapid analytical techniques are expected to be seriously wont to combat this disease.

It was shown that the standard method for diagnosing Covid 19 disease within the early stages features a low positive rate (Zhu et al., 2020). Understanding the fundamentals and differing types of Covid 19 was also an ongoing experiment and would be built and shared worldwide.

Use of CT Scans to detect COVID19

Researchers have recently concluded that individuals with Covid 19 might be identified by examining chest X-ray image patterns. The research summarized various studies that are drained this field (Ozsahin et al., 2020).

Kedia et al. (2021) designed a deep learning model called COVNet to extract the graphic features of CT images, chest scans, and Covid 19 detection.

Gozes et al. (2020) proposed a comprehensive system to detect COVID-19 in normal cases. The proposed system included lung segmentation, COVID-19 detection in CT slices, and marking the case as COVID-19 using a predetermined threshold based on the counted COVID-19 positive slices.

Ahuja et al. (2020) used data augmentation and pretrained networks to classify COVID-19 images. Data augmentation was performed using stationary wavelets, and the random rotation, translation, and shear operations were applied to the CT scan images.

Research objectives

The aim of this thesis is to present and implement a quick and accurate method for diagnosing coronary cardiovascular disease. Another goal of this dissertation is to cut back the time of diagnosis of coronary cardiovascular disease and speed up the method of quarantining someone and starting treatment. Other sub-objectives are considered during this research:

1. Compare different machine learning methods for corona recognition

2. Select the most effective car learning model for corona detection

3- Comparing the results of this research with the fundamental article

CHAPTER II

Literature Search and Research Background

History of the COVID19

On 29 December 2019, doctors in Wuhan, China, noticed unusual cases of patients with pneumonia. However, the primary cause of the disease was observed on 12 December. Subsequent studies showed that the first disease originated within the seafood, poultry, and live animals market in Wuhan City in central China's Huai province. An unusual outbreak of pneumonia was reported to the globe Health Organization. On 1 January 2020. The market was closed and decontaminated. After widespread speculation about the reason behind the disease, the Chinese section of the Centers for Disease Control and Prevention finally announced a replacement coronavirus called Covid 19.

Due to the simultaneous increase in demand for intensive care units, the health systems of most countries, even many developed countries, have collapsed. Intensive care units are full of patients with Covid 19. The global distribution of Covid cases up to 27 March 2020 is shown in Figure 1.

Due to the highly contagious nature of Coronavirus Disease 2019 (COVID-19) and also the lack of specific proven therapies and licensed vaccines that are effective against it, early detection of this disease is of particular importance. Joint chest CT imaging of COVID-19 confirmed cases are discussed here. It was suggested that a CT scan of the chest is important as an adjunctive diagnostic method because of its' sensitivity and might provide clear images of occurrence. Display pneumonia and, therefore, the extent of involvement of the patient's lungs. Therefore, repeat PCR tests and CT scans of the breast are often useful in people suspected of SARS-CoV-2 infection in whom the PCR test results are negative.

Figure 1

Epidemic curve of confirmed COVID-19, by date of report and WHO region through 27 March 2020 (WHO, 2020).
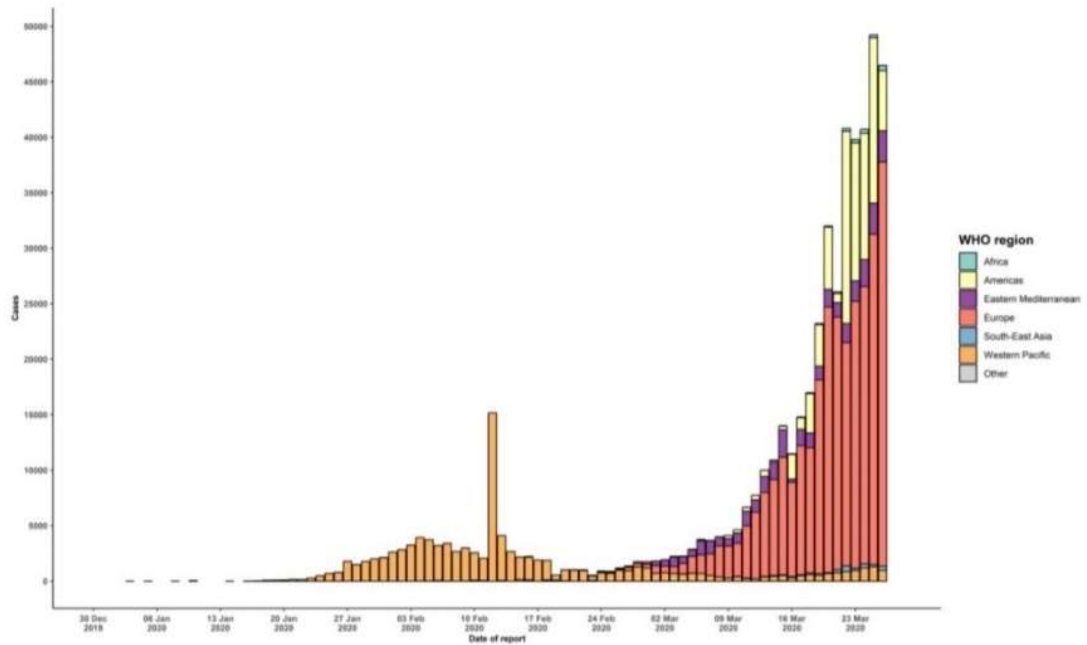
Performing a high-resolution CT scan to look at the lungs is important for early diagnosis and evaluation of the severity of Covid-19 infection in patients. Typical CT scans in Covid-19 patients show images of frosted enclosed parenchymal tissue on each side of the lung and established pulmonary opacity, and in some cases as circular shapes that stretch to the outer periphery of the lung tissue.

Although chest CT scans are considered a valuable diagnostic method for diagnosing Covid-19 disease, especially in areas where the SARS-CoV-2 virus is prevalent, there are shortcomings in this method, like other diagnostic methods. There is a scarcity of differentiation between different virus infections and weird CT images.

Lung tissue involvement is additionally seen in patients with SARS and Mers disease, and CT scan images of their breasts, like those with Covid-19 disease, are seen as fixed, opaque glass opacities. This reduces the specificity of the CT scan method for the definitive diagnosis of Covid-19 disease, and thru CT scan images of the breast alone, it is a challenging task to differentiate SARS, Mers, and Covid-19 diseases or

other viral diseases. Therefore, the RT-PCR test, together with the CT scan method, is critical for the definitive diagnosis of coronavirus and Covid-19 disease.

Artificial intelligence

Large computing systems are systems based on knowledge, events, and development of conventional computing and are referred to as fifth-generation computing. Using a database, experts can define rules that simulate the thought process and solve problems using easy conclusions and rules. The idea of large systems is that by preparing a list of logical propositions and making logical transformations on them, thinking can be modeled. Large systems are used to perform medical diagnoses and solve other diagnostic problems. In environments where ambiguity and uncertainty can help, this method helps a lot in forming predictions and decisions. For example, incorrect scientific models of human health and disease, and data that is insufficient and sometimes inaccurate about specific patients, often differ from medical practices, which can interfere.

Figure 2

Learning types of machine learning
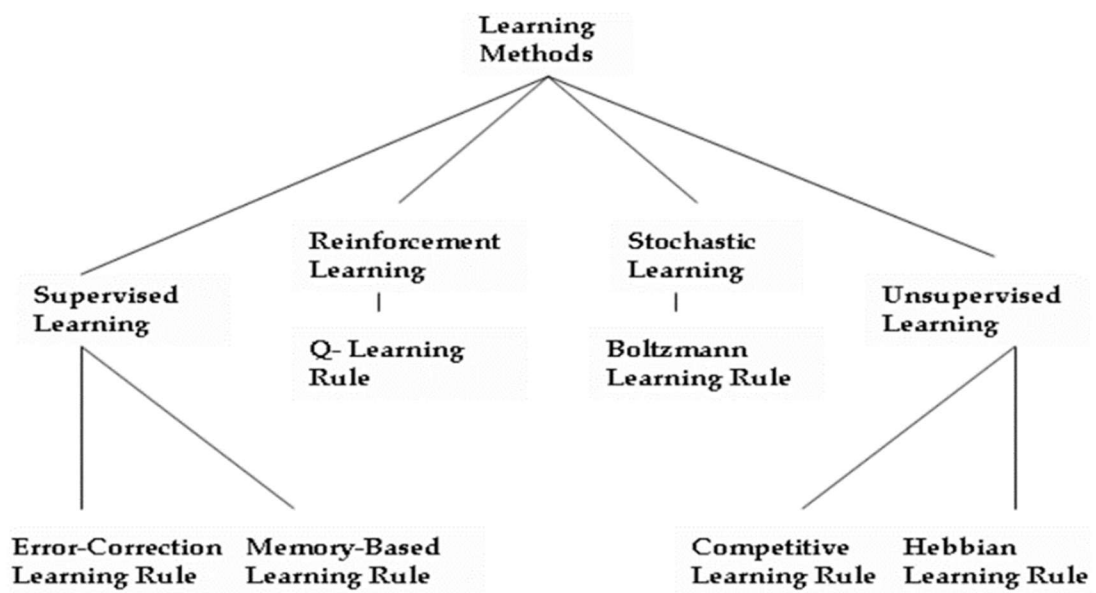
Learning methods

The machine learning framework is split into supervised learning, unsupervised learning, and reinforcement learning (Figure 2). In the supervised learning model, it is assumed that the trainer or supervisor categorizes the tutorial samples and uses the category membership information of each training data. However, in unsupervised learning models, pattern class information does not send to the model, and in reinforcement learning, learning is finished through interaction with the environment-supported trial and error (with reward/punishment allocation). Figure 3 shows the difference between supervised and unsupervised learning.

Unsupervised learning aims to cluster the input points near one another, so a label - cluster index - is assigned to every input point .

Reinforced learning is between learning with supervision and without supervision. Unlike unsupervised learning, it is a type of monitoring, but this monitoring is not an arbitrary output to any input of the data bank in the assigned variety. An augmentation learning algorithm receives feedback only after selecting an output for a particular input or view of the environment. Indicates numerical feedback that the output (known for reinforcing learning practice) meets the learner's goals. Reinforcement learning is used in sequential decision-making issues in which the learner adapts to his or her environment through sequential action (outputs) with observations (inputs) when receiving answers to each chosen action .

Artificial Neural Network

The artificial neural network may well be a computational model supported by biology and consists of the various individual units called artificial neurons that are connected by coefficients (weights) and form the body structure neurons are called processing elements (PE) (Sekeroglu et al., 2020). Each processing element has inputs, a transfer function, and an output. The processing element is actually an equation that balances inputs and outputs. Neural networks (ANNs) are called interconnected connection models because the connection weights represent the system memory. Although one neuron can perform simple IP functions, the ability of neural computation comes from the connection of neurons

during a network. Figures 4 and 5 present an artificial neuron and artificial neural network. The intelligence of artificial neural networks continues to be debated. Artificial neural networks rarely process over some hundred or some thousand elements, while the human brain has 100 billion neural cells; Thus, artificial networks are identical because the human brain's complexity is still aloof from the creative capacity of the human brain. Nevertheless, artificial neural networks are capable of processing large amounts of information and predictions (Dogruyol and Sekeroglu, 2020; Sekeroglu, 2004). Various neural networks are designed (Minsky, 1961).

Figure 3

Supervised vs. Unsupervised Learning (a) supervised learning and (b) unsupervised learning
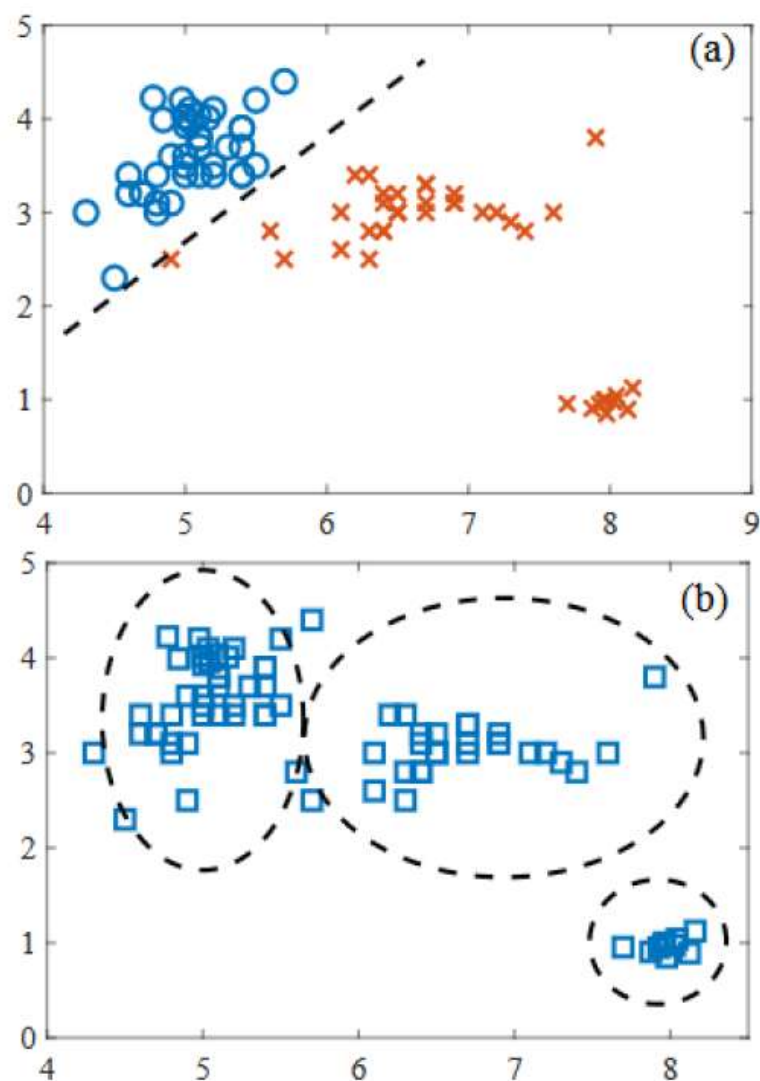
Artificial neurons are the building blocks of ANN, designed to simulate the function of biological neurons. Input signals are multiplied by the weights of the connection (modified), and the transfer function is applied to produce the neuron's output. The summation function refers to the sum of the weights of the neuronal inputs. The ANN-supervised learning paradigm efficiently solves linear and nonlinear problems (Alsmadi et al., 2009).

Figure 4

An artificial neuron
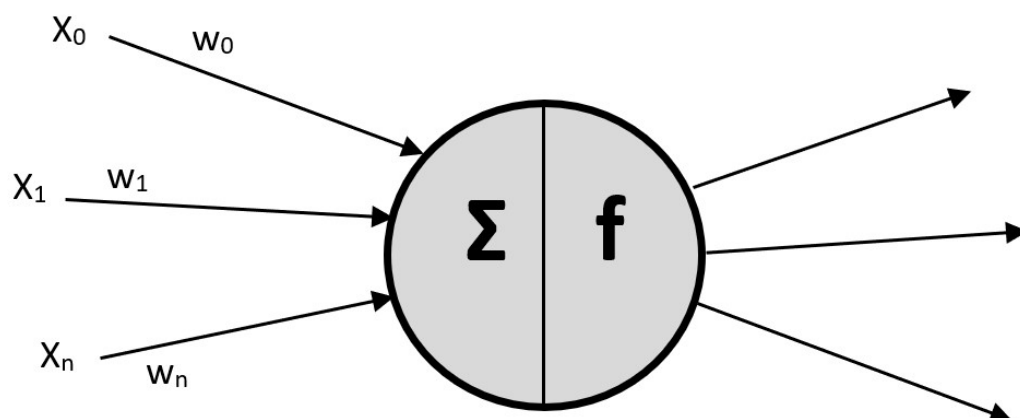
Decision Tree

The decision tree can be used for both classification and regression problems (Sekeroglu and Tuncal, 2021). As its name implies, this method uses a variety of nodes and branches. In each of the opposite nodes, a call is formed consistent with one or more specific attributes. The decision tree can also be used for describing the relationships between attributes. Decision trees are known among machine learning algorithms as a useful and effective method, and with speed and success, real-world problem solving is predicated on intelligence (Tuncal et al., 2020). Therefore, it has a high ability to resolve complex problems through graphical representation

Figure 5

Artificial Neural Network
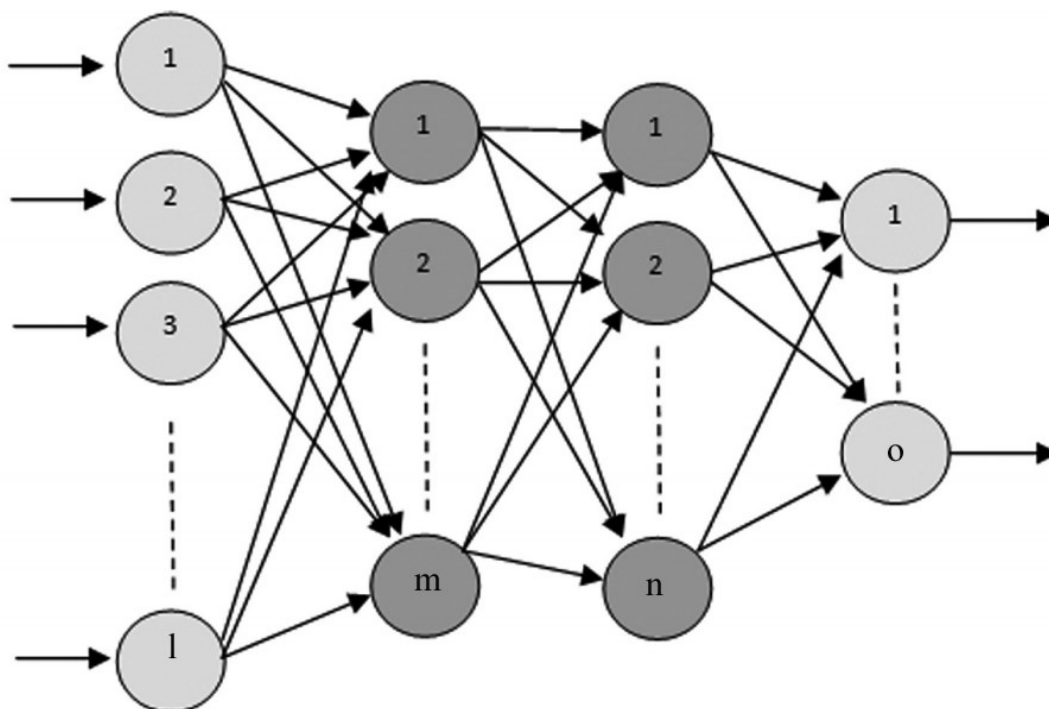
k- Nearest Neighbor

It is one of the foremost widely used algorithms within the field of classification. It supports approximation, and any calculation is delayed until the classification is completed. Within the kNN classifier, an untrained sample is identified that supports the similarity between known trained or labeled samples by calculating the gap between anonymous samples and labeled samples. The closest samples are then selected for classification by default, and also the unspecified sample is assigned to the category with the very best sample of the closest specimens.

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both categorized and regression problems (Sekeroglu and Emirzade, 2018;

Alpan and Sekeroglu, 2020). However, it is mostly used in classification issues. In the SVM algorithm, each data instance is plotted as a degree in the next n space in the data scatter diagram (where n is the number of attributes that a knowledge instance has). It then categorizes different and distinct data by drawing a decision line.

The SVM considers the support vectors, which are the closes data points of each class, to draw a decision line. Figure 6 shows the support vectors and the hyperplane to separate the classes.

Figure 6

Visualization of SVM (Manjrekar and Dudukovic, 2019)
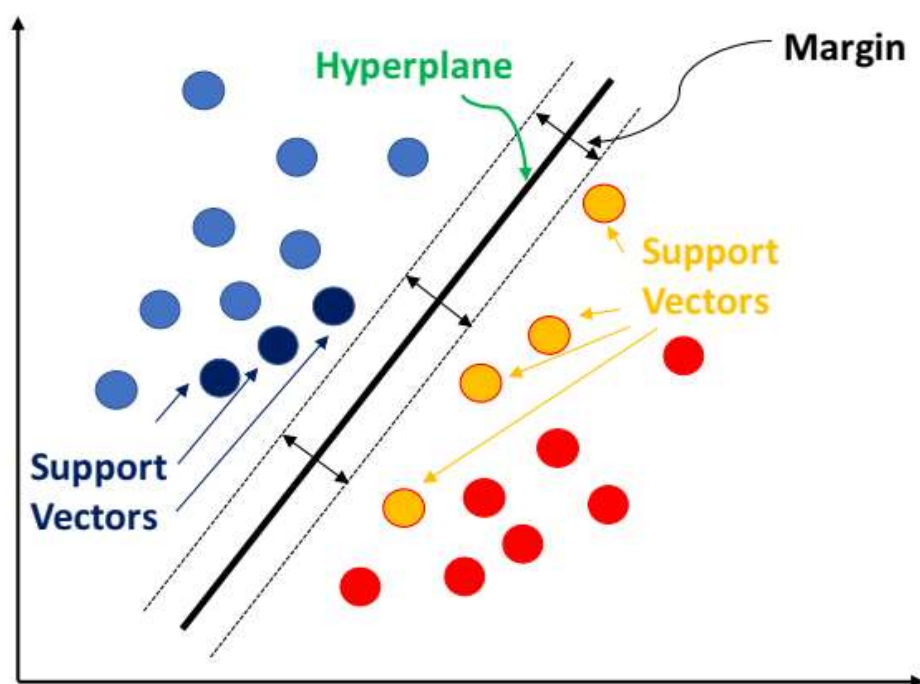
Kernel functions of SVM are used to map the data points into m-dimension to provide separable data points which cannot classify linearly. Figure 7 shows an example plot of nonlinear mapping of data points into high-dimensional feature space. Table 1 presents the frequently used kernel functions of SVM.

Application of AI in COVID19 Detection

Muhammad et al. (Muhammad et al., 2021) used machine learning to spot COVID-19. These models include logistic regression, decision tree algorithm,

support vector machine, and artificial neural networks. They used a dataset called Cui-Mexico to implement the models, which used 80% of the data for training and 20% for testing. Evaluation of the performance of models and algorithms has shown that the choice tree model with 94.99% has the most effective accuracy, and also the support vector machine model with 93.34% can show the simplest sensitivity. Recently, Roy et al. (2020) proposed a deep learning technique for classifying lung ultrasound (LUS) images.

Chen et al. (2020) performed a study on collected 46,096 images from 106 patients (Renmin Hospital of Wuhan University–Wuhan, Hubei province, China). The proposed system was based on segmenting CT scans using UNet++ and predicting the COVID-19 lesions.

Liu et al. (2020) proposed another deep neural network model, namely, lesion-attention deep neural networks, where the backbone of the model used the weights of pretrained networks such as VGG16, ResNet18, and ResNet50. Instead of deep learning approaches, Barstugan et al. (2020) considered machine learning algorithms to classify 150 COVID-19 and non-COVID-19 images. The feature extraction process considered varied methods such as grey-level size zone matrix (GLSZM) and discrete wavelet transform (DWT), and the extracted features were classified using a support vector machine.

Figure 7

SVM with a nonlinear kernel, (a) binary pattern classification problem in input space, and (b) nonlinear mapping into high-dimensional feature space (Khan et al., 2016)
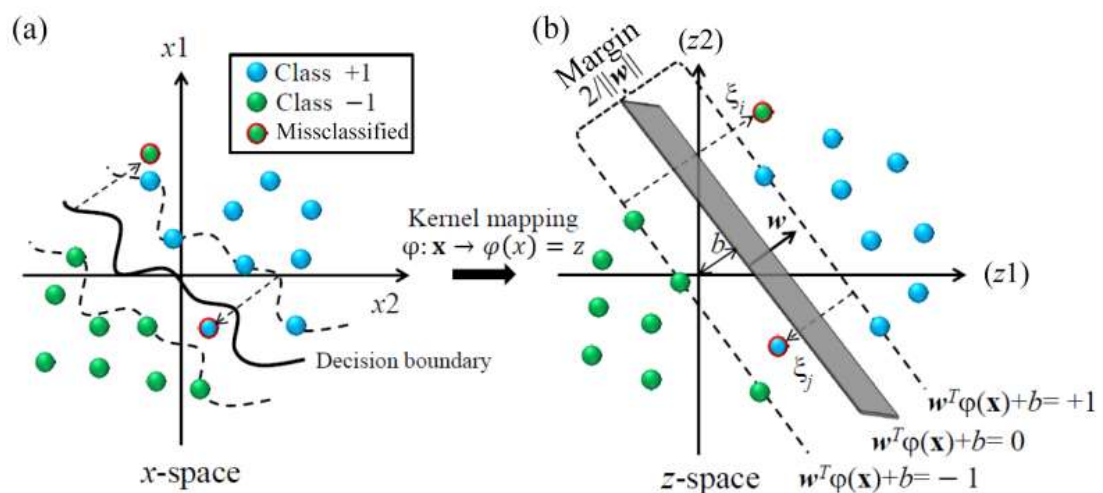
Table 1

The most common kernel functions

Kernel formula	Kernel type
$K(x, x_i) = (x^T x_i)$	Linear core
$K(x, x_i) = ((x^T x_i) + 1)^d$	Multi-degree core
$K(x, x_i) = \exp(-\gamma \| x - x_i \|^2)$	Radial base core
$K(x, x_i) = \tanh((x^T x_i) + b)$	Sigmoid nucleus

Many researchers are engaged in image processing techniques on Covid 19. The classification studies were performed with chest X-ray images to detect and classify different pneumonia and COVID19 using a convolutional neural network (CNN) (Sekeroglu and Ozsahin, 2020). Using lung scan images, Shan et al. (2021) implemented a deep learning-based system to segment all healthy and contaminated lung areas automatically. Sekeroglu and Ozsahin (2020) implemented CNN and pre-trained networks to detect influenza Covid 19. The CNN model was accustomed predict COVID19 and outperformed other pretrained models. Table 2 summarizes some of the recent research on detecting and diagnosing COVID19.

Table 2.

Summary of recent COVID19 research

| Study	Aim
Vaid et al., 2020	Deep learning COVID19 detection bias
Sekeroglu and Ozsahin, 2020	Detection of COVID19 using CNN
Salman et al., 2020	Covid-19 detection using artificial intelligence
Ismael and Sengur, 2021	X-ray and Deep learning for COVID19 detection
Xu et al., 2020	Deep learning system for screening COVID19
Shan et al., 2020	Abnormal lung quantification in chest CT images of COVID-19 with Deep Learning
Ardakani et al., 2020	Deep learning technique to manage COVID-19 in routine clinical practice using CT images

CHAPTER III

Materials and Methods

Introduction

This chapter presents the data normalization, selection, and feeding of the data to the machine learning models. Figure 8 shows the general block diagram of the study.

Figure 8

General block diagram of the study.
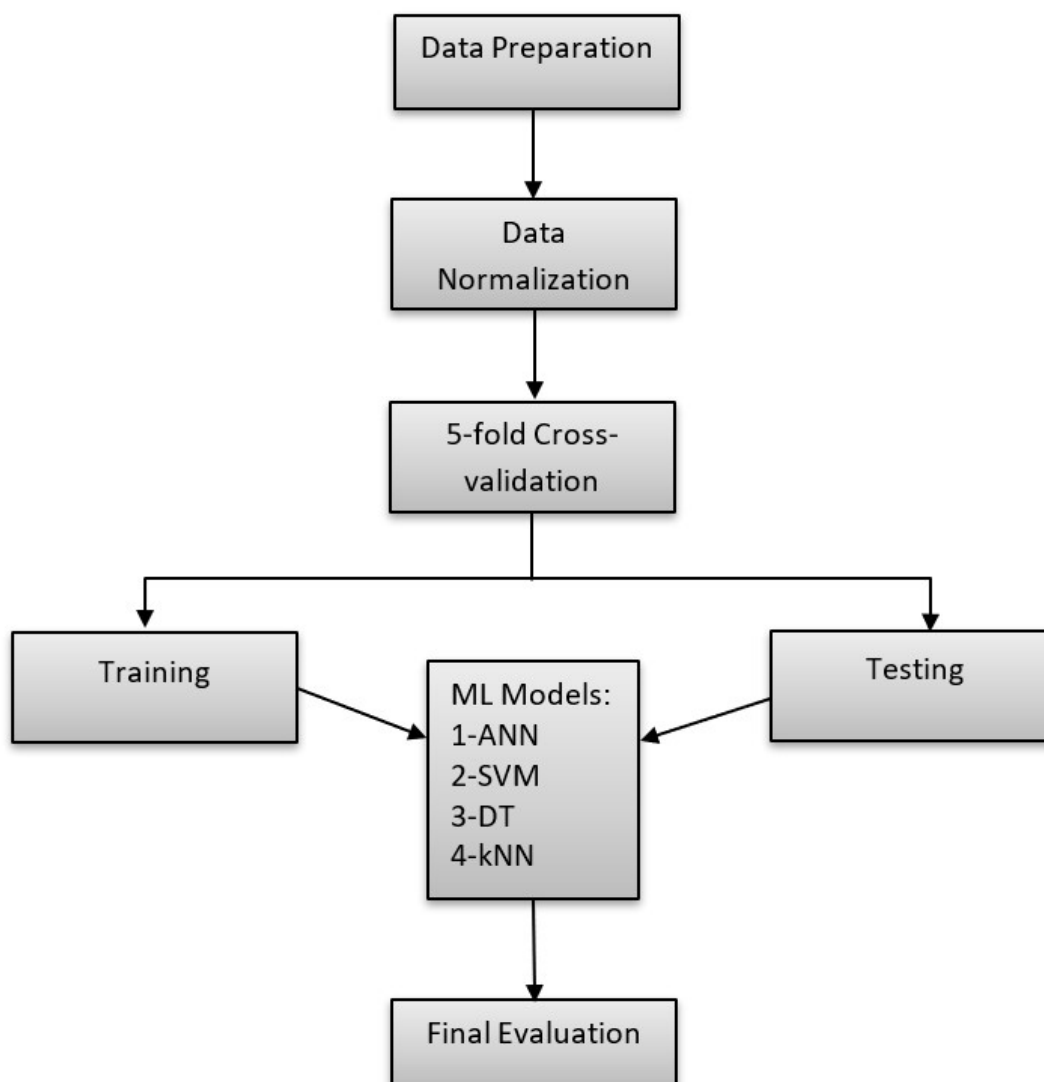
Dataset

Mexico COVID-19 clinical data (Mexico Dataset, 2020) is considered in this thesis. A dataset consisting of positive test results further as COVID-19 negative test results from all Mexican hospitals was employed in this study. Database supported Respiratory Viral Disease Information Surveillance Cameras, described by 475 Viral Respiratory Disorders Monitoring Units throughout Mexico. The dataset contains 263,007 items or files with 41 attributes.

Preparation of dataset

Mexico's COVID-19 positive and negative case data set has 41 attributes/columns. Of the 41 characteristics for every individual, only two demographic characteristics, including age and sex, and eight clinical characteristics, including pneumonia, diabetes, asthma, hypertension, anxiety, obesity, chronic renal impairment, and a high-risk factor, were tested. The considered attributes and statistical description can be seen in Tables 3 and 4.

Machine Learning Models

This study used four machine learning methods to diagnose people with coronary artery disease. These four methods are artificial neural network, support vector machine algorithm, decision tree algorithm, and k-nearest neighbor algorithm.

Table 3

Dataset attributes considered in this thesis

Attribute	Number
Age	1
Gender	2
Pneumonia	3
Diabetes	4
Asthma	5
High blood pressure	6
Cardiovascular diseases	7
Obesity	8
Chronic kidney disease	9
smoking	10
Result	11

Table 4.

Statistical descriptions of the considered attributes.

standard deviation/mean	Median	Maximum	Minimum	Attribute	Number
16/90	45/59	120	0	Age	1
0/5	0/49	1	0	Gender	2
0.81	0.17	99	0	Pneumonia	3
6/07	0/51	98	0	Diabetes	4
5/80	0/38	98	0	Asthma	5
5/84	0/52	98	0	High blood pressure	6
5/91	0/38	98	0	Cardiovascular diseases	7
5/92	0/53	98	0	Obesity	8
5/84	0/37	98	0	Chronic kidney disease	9
5/98	0/46	98	0	smoking	10
0/49	0/39	1	0	Result	11

Evaluation

The confusion matrix was used for evaluation. Figure 9 shows the confusion matrix.

Figure 9

Confusion Matrix

True positive is the correctly classified samples that have a disease, while True Negative is the correctly classified samples that do not have a disease. False-positive

and false-negative values are the samples predicted as having a disease but not having a disease and samples predicted as not having a disease but having a disease, respectively.

If the data is balanced, accuracy is the major metric for model evaluations in classification problems. The formula of accuracy is given in Equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+F} \qquad (1)$$

The precision, sensitivity, and specificity are the other metrics used to measure the ability of models to detect positive or negative samples correctly. The formulae of precision, sensitivity, and specificity are shown in Equations 2, 3, and 4.

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (4)$$

The k-fold cross-validation method is used for both hyperparameter tuning and the final performance evaluation of the models. It divides the dataset into k subsets; one is used for testing while the others are considered for the training. A different fold is assigned as the test set in each training, and the training process is repeated k times. In this thesis, 5-fold cross-validation is applied, and the mean of all metrics was considered as the final results.

CHAPTER IV

Results and Discussion

Introduction

The aim of this thesis is to present a model based on machine learning to diagnose people with coronary arteries. Several machine learning models described above are employed.

Results

Artificial neural network results

An artificial neural network with three different activation functions was accustomed to detecting and classifying positive and negative corona results supported the epidemiological database. The functions accustomed create the artificial neural network model included the Levenberg-Marquardt optimization (*Trainlm)*, Scaled conjugate gradient backpropagation (*Trainscg)*, and Resilient Backpropagation (*Trainrb)*. For all different models of the artificial neural networks, the number of neurons within the single hidden layer of 10 neurons and its structure was 2-10-10.

Figure 10 shows the change in training error of the neural network model trained with the Scaled conjugate gradient backpropagation. This figure shows that the error value of the trained model has converged to its optimal value, which indicates the right training of this model. Figure 11 presents the confusion matrix of the epidemiological data classification by the artificial neural network trained with the Scaled conjugate gradient backpropagation. The overall accuracy of the model in classifying the type of epidemiological data was 90.2%, which is that the sensitivity and accuracy of this model in detecting the positive results of Covid 19 were adequate at 80.1% and 96.7%, respectively. In addition, the sensitivity and accuracy of the ANN-Trainscg model in diagnosing a healthy person through epidemiological data is up to 98.3% and 88.5%, respectively.

Figure 10

Training curve of the ANN with Scaled conjugate gradient backpropagation
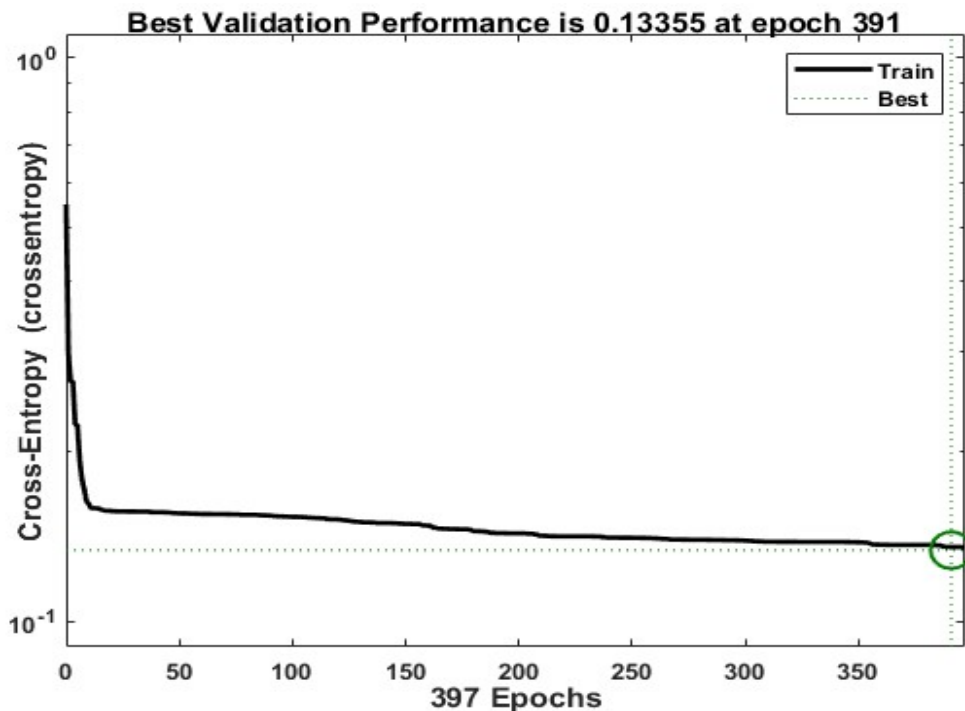

Figure 11

Confusion Matrix of ANN with with Scaled conjugate gradient backpropagation
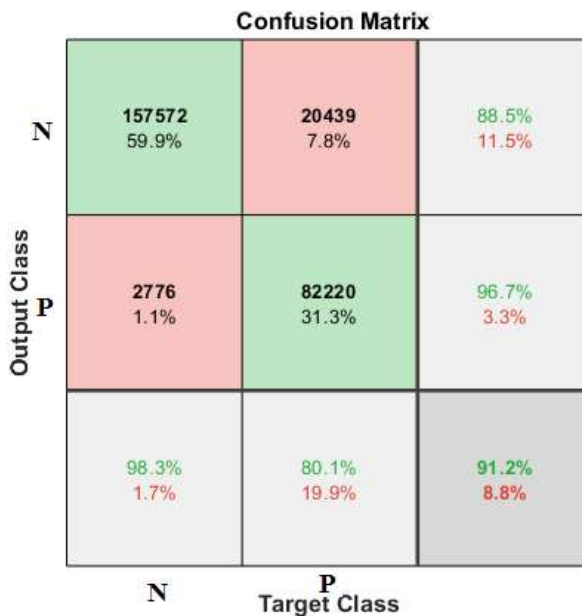

Figure 12 shows the convergence process of the error value of the model trained with the Trainlm activation function to spot individuals with Covid 19 to point to the validity of the model training. Figure 13 shows the ANN-Trainlm model confusion matrix. The results show that the ultimate accuracy of this model in classifying the

kind of epidemiological data is 92.5%, which has increased by about 1.3% compared to the Train-scg model.

Figure 12

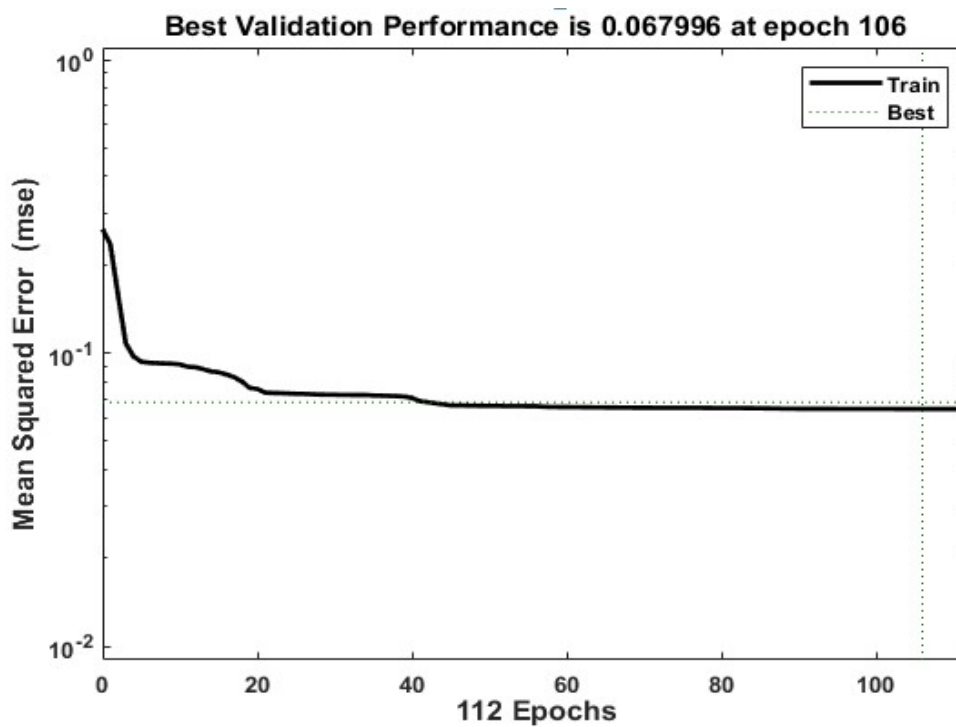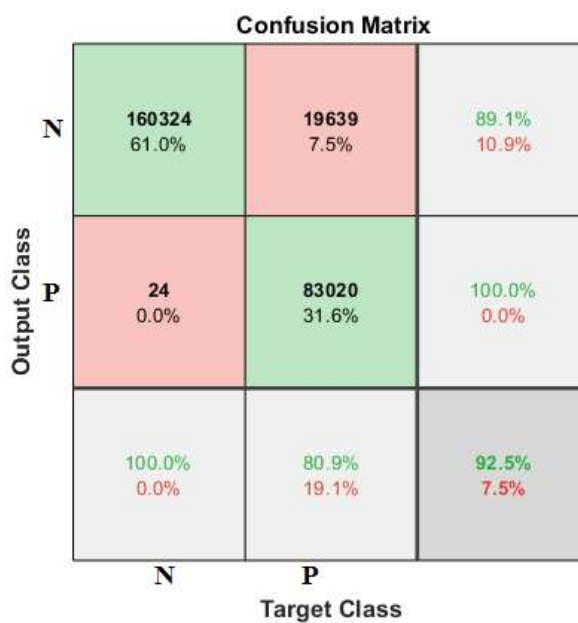*Training curve of the ANN with Levenberg-Marquardt optimization.*

Figure 13

Confusion Matrix of ANN with LM.

The sensitivity and accuracy of the trained model in diagnosing an individual with Covid 19 from epidemiological data were adequate at 80.9% and 100%, respectively, and the model's sensitivity and accuracy in diagnosing a healthy person were 100% and 89.1%, respectively.

Figures 14 and 15 show the training process of the ANN with the resilient backpropagation model and the confusion matrix, respectively. Evaluation of the results shown within the confusion matrix shows that the accuracy of this model in classifying the sort of epidemiological data is adequate at 91.8%. The model's sensitivity for diagnosing a healthy person and someone with Covid 19 is up to 99.7% and 79.5%, respectively, and the accuracy of the model is capable 88.4% and 99.4%. In general, the Yan model was less accurate than the opposite two neural network models in diagnosing someone with artery disease.

Decision Tree Results

Table 4 shows the decision tree confusion matrix for diagnosing Covid 19 with supported epidemiological data. The general accuracy of the choice tree in data type classification was 92.94%, which was the overall artificial neural network model.

Figure 14

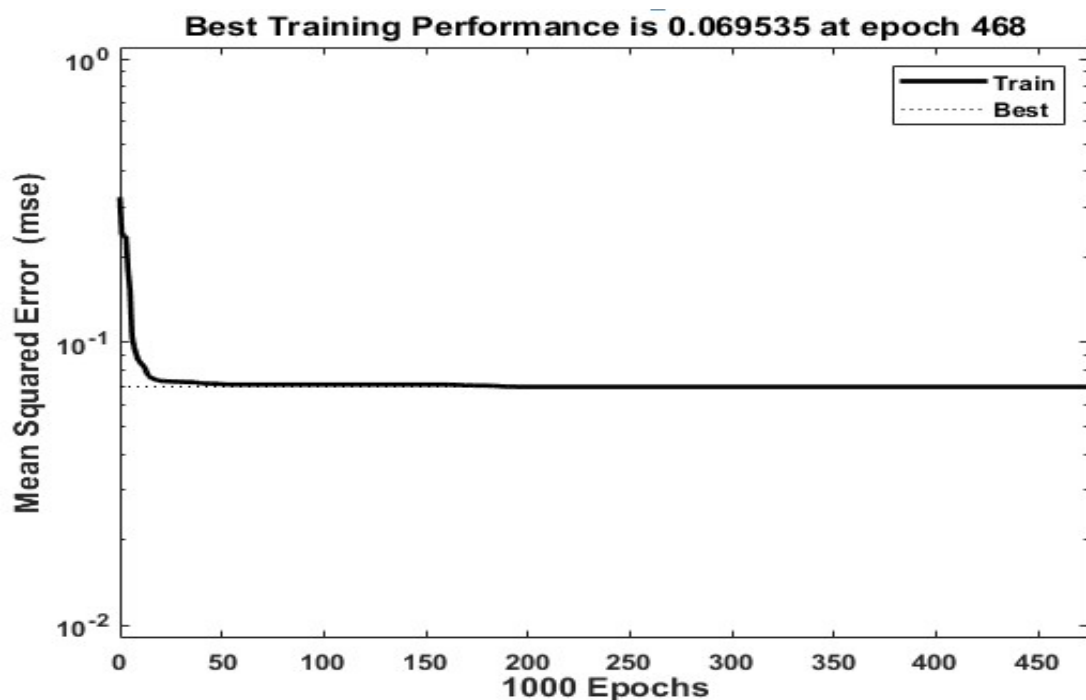*Training curve of the ANN with the Resilient Backpropagation.*

Figure 15

Confusion Matrix of ANN with Resilient Backpropagation.

The sensitivity and accuracy of the decision tree model in diagnosing people with Covid 19 supported epidemiological data were 83.02% and 98.69%, respectively. The specificity and accuracy of the decision tree model in diagnosing a healthy person were 99.29% and 90.13%. Table 5 shows the confusion matrix for DT, and Figure 16 presents the final decision tree for diagnosing COVID19.

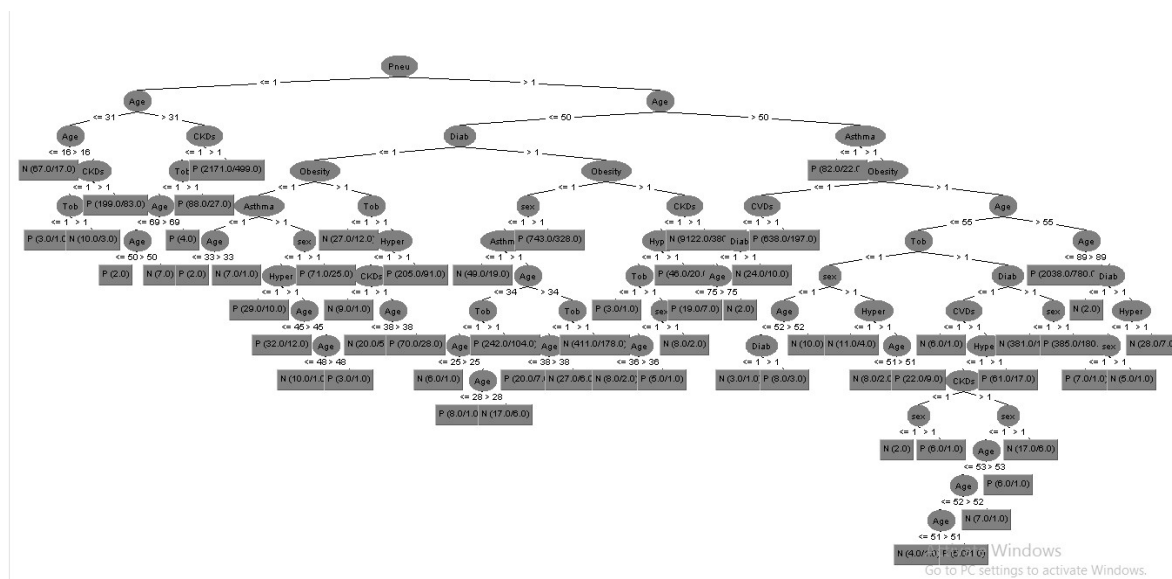Table 5

Confusion Matrix of DT

	Negative	Positive	(%)
Negative	159216	17435	90.13
Positive	1132	85224	98.69
(%)	99.29	83.02	92.94

Figure 16

Constructed Decision Tree

k-Nearest Neighbor Results

This section presents the results of detecting Covid 19 by the K-nearest neighbor method. Three neighborhoods were considered for data type classification. Table 6 presents the kNN confusion matrix for the detection of Covid 19. The overall accuracy of the kNN method in classifying epidemiological data was 80.59%. Also, the sensitivity and accuracy of the model in diagnosing a person with Covid 19 were equal to 68.73% and 78.83%. The sensitivity of the kNN model for diagnosing a healthy person was 88.18%, and its accuracy was 81.50%. The accuracy of the nearest-neighbor model was the lowest relative to the neural network and decision tree accuracy.

Table 6

Confusion Matrix of kNN

	Negative	Positive	(%)
Negative	141403	32097	81.50
Positive	18945	70562	78.83
(%)	88.18	68.73	80.59

Support Vector Machine Results

This section presents the results of implementing a support vector machine to diagnose individuals with Covid 19 by three different kernels. The kernels employed in this section consisted of linear, polynomial, and Gaussian radial basis function kernels. The confusion matrices for all kernel functions considered in this thesis are presented in Tables 7, 8, and 9. the accuracy of knowledge type detection and classification for various support vector machine models for linear, Gaussian radial, and polynomial kernels were 9.09%, 93.44%, and 96.73%, respectively, with the very best accuracy. The backup vector machine with the kernel was polynomial. Also, the sensitivity of SVM-linear, SVM-RBF, and SVM-polynomial models in diagnosing an individual with Quaid 19 was 86.61%, 89.70%, and 93.24%, respectively, with a big difference between vector accuracy Backup with polynomial kernel was observed compared to other models. Examination of the results of this section showed that the support vector machine observed the best accuracy in diagnosing an individual with COVID19 with the polynomial kernel, which was more accurate and sensitive than all the methods studied during this study.

Table 7

Confusion Matrix of SVM with Linear Kernel.

	Negative	Positive	(%)
Negative	150674	13750	91.64
Positive	9674	88909	90.19
(%)	93.97	61.86	91.09

Table 8

Confusion Matrix of SVM with Radial Basis Function Kernel.

	Negative	Positive	(%)
Negative	153650	10563	93/57
Positive	5698	92096	94/17
(%)	96/44	89/70	93/44

Table 9

Confusion Matrix of SVM with Polynomial Kernel.

	Negative	Positve	(%)
Negative	158695	6943	95/81
Positve	1653	95716	98/30
allergy (%)	98/97	93/24	96/73

Comparison of the results

This study presented four different machine learning models to diagnose Covid 19 from epidemiological data, showing that the support vector machine model's accuracy was higher than other methods. Table 10 compares all implemented models for diagnostic purposes. Overall, the results showed that the employment of a support vector machine in identifying people with Covid 19 increased the accuracy of the model compared to the bottom article, so using the polynomial kernel, the accuracy of this model increased by about 2%.

Table 10

Comparison of the obtained results.

Accuracy (%)	allergy (%)	Accuracy (%)	Model
96.7	80.1	91.2	ANN-Trainscg
100	80.9	92.5	ANN-Trainlm
99.4	79.5	91.8	ANN-Trainrb
90.19	86.61	91.09	SVM-Linear
94.17	89.70	93.44	SVM-RBF
98.30	93.24	96.73	SVM-Polynomial
98.69	83.02	92.94	Decision Tree
78.83	68.73	80.59	kNN

CHAPTER V
Conclusion

This thesis implemented different machine learning methods to diagnose people with COVID19 based on epidemiological data. A total of four machine learning models, including an artificial neural network, decision tree, support vector machine, and k-nearest-neighbor method, were implemented.

The final accuracy of the artificial neural network with Scaled conjugate gradient backpropagation was 91.2%, which was 80.1% and 96.7%, respectively, in detecting the positive results of COVID19. The final accuracy of the ANN with the Levenberg-Marquardt model was 92.5%, and the sensitivity and accuracy were 80.9% and 100%. The final accuracy of the ANN with Resilient Backpropagation was 91.8%, and the sensitivity was 79.5% for diagnosing Covid 19.

The decision tree's final accuracy, sensitivity, and specificity in diagnosing individuals with Covid 19 were 92.94%, 83.02%, and 98.69%, respectively. The overall accuracy of the kNN method in classifying epidemiological data was 80.59%, with sensitivity and accuracy of 68.73% and 78.83%. The final accuracy of the support vector machine for linear kernels, Gaussian radial basis function, and polynomials was 9.09%, 93.44%, and 96.73%, respectively, and the sensitivity of SVM-Linear, SVM-RBF, and SVM-Polynomial models were obtained. In the diagnosis of a person with Covid 19, the obtained results were 86.61%, 89.70%, and 93.24%, respectively. The SVM with polynomial kernel achieved superior results and outperformed the other models in diagnosing COVID19 from epidemiological data.

It is recommended that chest CT scan data and epidemiological data be used simultaneously to diagnose Covid 19 disease. In addition, the use of a differential evolution algorithm to optimize the structure of the support vector machine.

References

Ahuja, S., Panigrahi B. K., Dey N., Gandhi T., & Rajinikanth V. (2020). Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices, Applied Intelligence.

Alimadadi A., Aryal S., Manandhar I., Munroe P.B., Joe B., & Cheng X. (2020). Artificial intelligence and machine learning to fight COVID-19. Physiological Genomics 2020 52:4, 200-202

Alpan, K., & Sekeroglu B. (2020). Prediction of Pollutant Concentrations by Meteorological Data Using Machine Learning Algorithms, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences; Gottingen Vol. XLIV-4/W3-2020, : 21-27. Gottingen: Copernicus GmbH.

Alsmadi, M., Omar K.B., & Noah S.A. (2009) Back propagation algorithm: the best algorithm among the multi-layer perceptron algorithm. IJCSNS International Journal of Computer Science and Network Security, 9:4.

Ardakani A.A., Kanafi A.R., Acharya U.R., Khadem N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Computers in Biology and Medicine. 121:103795

Barstugan M., Ozkaya U., & Ozturk S. (2020). Coronavirus (COVID-19) classification using CT images by machine learning methods, ArXiv, https://arXiv:2003.09424

Chen J., Wu L., Zhang J. et al. (2020). Deep learning-based model for detecting 2019 novel coronavirus pneumonia on highresolution computed tomography: a prospective study. medRxiv, 2020

Dogruyol, K., & Sekeroglu, B. (2020). Absenteeism Prediction: A Comparative Study Using Machine Learning Models. In: Aliev, R., Kacprzyk, J., Pedrycz, W., Jamshidi, M., Babanli, M., Sadikoglu, F. (eds) 10th International Conference on Theory and Application of Soft Computing,

Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. Advances in Intelligent Systems and Computing, vol 1095. Springer, Cham.

Gozes O., Frid-Adar M., Greenspan H. et al. (2020). Rapid AI development cycle for the coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXi v, https://arXiv:2003.05037

Hussain E., Hasan M., Rahman A., Lee I., Tamanna T., & Parvez M.Z. (2021) CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images, Chaos, Solitons & Fractals, 142, 110495.

Ismael, A.M. & Şengür A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Systems with Applications. 164: p. 114054

Kedia P., Anjum, & Katarya R. (2021) CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. Applied Soft Computing. 104:107184

Khan, F.& Enzmann, F. & Kersten, M. (2016). Multi-phase classification by a least-squares support vector machine approach in tomography images of geological samples. Solid Earth. 7. 481-492.

Liu B., Gao X., He M., Liu L., & Yin G. (2020). A fast online COVID-19 diagnostic system with chest CT scans. in Proceedings of KDD 2020, New York, NY, USA, 2020.

Manjrekar, O., & Duduković, M. (2019). Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique. Chemical Engineering Science.

Mexico COVID-19 clinical data, Franklin, R.M. (2020). Epidemiology data for positive and negative COVID-19 cases in Mexico. Available at

https://www.kaggle.com/datasets/marianarfranklin/mexico-covid19-clinical-data/metadata

Minsky, M., Steps toward artificial intelligence. Proceedings of the IRE, 1961. 49(1): p. 8-30

Muhammad, L., Algehyne, E., Usman, S.S., Abdulkadir A., Chakraborty, C., & Mohammed, I. A. (2021). Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. SN computer science, 2021. 2(1).

Rabaan, A. A., Al-Ahmed, S. H., Sah, R., Al-Tawfiq, J. A., Haque, S., Harapan, H., Arteaga-Livias, K., Aldana, D., Kumar, P., Dhama, K., & Rodriguez-Morales, A. J. (2020). Genomic Epidemiology and Recent Update on Nucleic Acid-Based Diagnostics for COVID-19. Current tropical medicine reports, 1–7. Advance online publication.

Roy, S. et al. (2020). Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE Transactions on Medical Imaging, 39(8).

Salman F.M., Abu-Naser S.S., Alajrami E., Abu-Nasser B.S., & Ashqar, B.A.M. (2020). Covid-19 detection using artificial intelligence. International Journal of Academic Engineering Research (IJAER), 4:3(18-25).

Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., Shen, D., & Shi, Y. (2021). Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. Medical Physics 48, 1633–1645.. doi:10.1002/mp.14609

Sekeroglu B. (2004). Classification of sonar images using back propagation neural network. *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, 5:3092-3095.

Sekeroglu, B., Hasan, S.S., & Abdullah, S.M. (2020). Comparison of Machine Learning Algorithms for Classification Problems. In: Arai, K., Kapoor, S.

(eds) Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing, vol 944. Springer, Cham.

Sekeroglu B. & Emirzade E. (2018). A Computer-Aided Diagnosis System for Lung Cancer Detection Using Support Vector Machine. International Workshop on Pattern Recognition, China, May, 2018.

Sekeroglu B. & Tuncal K. (2021). Prediction of cancer incidence rates for the European continent using machine learning models. Health Informatics Journal. doi:10.1177/1460458220983878

Tuncal K., Sekeroglu B., Ozkan C. (2020). Lung Cancer Incidence Prediction Using Machine Learning Algorithms, Journal of Advances in Information Technology, 11(2) : 91-96.

Vaid, S., Kalantar R., & Bhandari M. (2020). Deep learning COVID-19 detection bias: accuracy through artificial intelligence. International Orthopaedics, 2020. 44: (1539-1542).

World Health Organization (2021). WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int.

Xu, X., et al. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering, 6(10): 1122-1129.

Yong, S.E.F. et al. (2020). Connecting clusters of COVID-19: an epidemiological and serological investigation. The Lancet Infectious Diseases, 20(7): 809-815

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., & Tan, W., (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine 382, 727–733.

Appendix A

Turnitin Similarity Report

Assoc.Prof.Dr. Boran Sekeroglu

Appendix B

ETHICAL APPROVAL DOCUMENT

Date: 10/02/2022

To the Institute of Graduate Studies;

For the thesis project entitled "CORONAVIRUS DETECTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS," the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

Title: Assoc. Prof. Dr.

Name Surname: Boran Şekeroğlu

Signature:

Role in the Research Project: Supervisor