REPUBLIC OF TURKEY

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

A NEW METHOD BASED CNN COMBINED WITH GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE FOR COVID-19 DETECTION BY ANALYZING X-RAY IMAGES

Karam Mohammed Dawood DAWOD

Master's Thesis

Supervisor

Prof. Dr. Osman Nuri UÇAN

Istanbul, 2022

A NEW METHOD BASED CNN COMBINED WITH GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE FOR COVID-19 DETECTION BY ANALYZING X-RAY IMAGES

Karam Mohammed Dawood DAWOD

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

Istanbul, (2022)

The thesis titled A NEW METHOD BASED CNN COMBINED WITH GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE FOR COVID-19 DETECTION BY ANALYZING X-RAY IMAGES prepared by KARAM MOHAMMED DAWOOD DAWOD and submitted on 09.02.2022 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering

Prof. Dr. Osman nuri UÇAN

the Supervisor

Thesis Defense Committee Members:

Prof. Dr. Osman Nuri UÇAN	Faculty of Engineering and Architecture, Altinbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Faculty of Engineering and Architecture, Altinbaş University	_____
Prof. Dr. Hasan Hüseyin BALIK	Faculty of Electrıcal And Electronics, Yıldız Technical University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to the Graduate Education Institute: ___/___/___

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Karam Mohammed Dawood DAWOD

DEDICATION

I dedicate this thesis to my supervisor, father, mother and my all friends.

ABSTRACT

A NEW METHOD BASED CNN COMBINED WITH GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE FOR COVID-19 DETECTION BY ANALYZING X-RAY IMAGES

Dawood, Karam Mohammed Dawood

MSc, Electrical and Computer Engineering, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UÇAN

Date: February 2022

Pages: 52

COVID-19 is an infectious disease caused by the newly discovered coronavirus. This new type of virus and disease was unknown before the outbreak that emerged in Wuhan, China in December 2019. COVID-19 poses a serious threat to public health. Older adults and people with pre-existing medical conditions such as diabetes, high blood pressure, heart disease, chronic lung disease and obesity are at increased risk for serious illness and complications. Last year, a computer scientist used various machine learning and deep learning techniques to detect Covid-19.

In this study, efficient Covid-19 detection framework presented to detect Covid-19 by analyzing x-ray tests. The proposed framework based CNN combined with genetic algorithm and SVM classifier. The main contribution in this study is combining CNN with genetic algorithm and SVM to detect Covid-19 with accurate estimation and minimum execution time. Several scenarios are executed to validated the presented method . Finally, the obtained results compared with several studies presented to solve this problem.

Keywords: Covid-19, Genetic Algorithm, SVM, CNN.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

SVM	:	Support Vector Machine
NN	:	Neural Network
RBF	:	Radial Basis Function
ESD	:	Energy Spectral Density
IoTs	:	Internet Of Things

1. INTRODUCTION

COVID-19 is a serious epidemic in Wuhan, China in December 2019 that infects humans from bats, spreads worldwide in a very short time, and is a pandemic and catastrophic health system in many countries. became. Severe Acute Respiratory Syndrome, COVID-19, causes severe respiratory failure in infected organisms with fatal consequences as the disease progresses. The main signs and symptoms of the disease are: It is known as fever, dry cough, sore throat, headache, malaise, malaise, diarrhea, and shortness of breath. In more advanced cases, the disease progresses to severe pneumonia, which can cause pneumonia due to oxygen imbalance and multiple organ failure. Especially for people with chronic illnesses, those with weak and immune defenses, smokers and the elderly, the disease has far more dangerous and debilitating consequences [1-4].

COVID-19 is usually spread by physical contact, for example. B. by breathing, hand or mucus contact with a person carrying the virus [5]. Antibiotics, antimalarials, antipyretics, cough suppressants, and pain relievers are often used to treat illness after infection. Hospitalization of infected patients depends on the degree, condition and severity of the disease. The number of patients infected with the COVID-19 virus around the world is increasing day by day. Even powerful countries like the United States, Italy and Spain have failed to protect themselves from the virus and are badly affected. With all of this information in mind, early diagnosis of the disease facing the healthcare system is essential to completely preventing a pandemic, or at least minimizing potential damage from the virus. In other words: at least suspected cases must be detected quickly, accurately and without error [6].

Coronavirus disease (COVID) was first reported by the World Health Organization (WHO) on 31 December 2019, with pneumonia cases of unknown origin being reported in Wuhan, China. Subsequently, it was announced that the identified cases were due to a new virus called 'novel coronavirus 2019' (2019-nCoV). After WHO declared this disease as an epidemic causing a global health emergency on January 30, 2019, COVID-19 reached the pandemic stage in March 2020. It has deeply affected all world health systems, economic and social mobility. The new Coronavirus 2019 (2019-nCoV) was later named as severe acute respiratory syndrome coronavirus (SARSCoV-2), describing its clinical features [1,2,3].

COVID-19 droplet infection can be transmitted by direct contact or fecal-oral route, and its clinical course is asymptomatic for the first 2-14 days, then upper and lower airway response lasting for a

few days, followed by hypoxia in the lung, ground glass opacities, ARDS. It has been stated that it can be examined in three different stages that can progress [4,5]. In the first asymptomatic stage, SARS-CoV-2 spikes glycoproteins and, unlike nasal cavity or other respiratory system viral infections, primarily binds to alveolar epithelial cells in the lower respiratory tract, especially through the ACE2 receptor, and enters various intracellular protein synthesis and viral RNA replication-transcription complexes. It is known that the virus multiplies by activating, the mature viral particle is released to infect other cells, but a limited innate immune response occurs [6,7,8]. In the second stage, where the virus continues to replicate and reaches the lower respiratory tract from the upper respiratory tract, the patient's clinical symptoms (fever, cough, sore throat, weakness, myalgia, headache, shortness of breath, loss of taste and smell, less frequently abdominal pain, diarrhea, etc.) in the airways, it is limited to mild to moderate symptoms . 3 In the remaining 15-20% of the patients, increased cytokine secretion, cellular apoptosis, inflammatory response, diffuse alveolar damage, fibinoid exudate in the alveolar space, giant cell reactions, especially type 2 pneumocyte cells in the alveoli, which are functional pulmonary units, are infected with the virus. At this stage, pulmonary ground glass/consolidation is detected as a radiological finding, and ARDS develops depending on factors such as the patient's immunity status, age, concomitant chronic diseases, and a possible genetic susceptibility that is not yet known. Mortality rates are lower (about 2%) compared to SARS and MERS, but mortality rates are reported to be higher (8-15%) in the population aged >80 years. It is also known that COVID-19 can cause infection and myocardial damage in the gastrointestinal system and central nervous system (encephalitis, necrotizing encephalopathy) with neurotropic effects.

Currently, the real-time reverse transcription polymerase chain reaction (RT-PCR) is widely used to diagnose and diagnose COVID-19. Chest x-rays, such as computed tomography (CT) and x-rays, are preferred for the early detection of COVID-19. The rapid spread of the disease and increasing death rates in many countries indicate the need for effective treatments. Therefore, treatment of the disease, including diagnosis, early quarantine, and follow-up care, is imperative. Right now, AI can contribute to the above perspective. Despite the strong similarities and severe shortage of skilled workers between COVID-19 and traditional pneumonia, artificial intelligence (AI) -based self-recognition models could be an important step towards significantly reducing testing times. There is sex [8]. In this context, effective Artificial Intelligence (AI) solutions to solve complex problems are essential. The research solution provides more accurate and less

expensive diagnostic treatment and diagnosis for COVID-19 and related diseases. Today, especially in the medical field, positive results are achieved with deep learning techniques that use image data sets such as retinal deep learning images, chest x-rays and MRI cerebral. It is used in many applications to extract, analyze and recognize data patterns [7].

In this study, three different conditions (COVID-19, viral pneumonia, and normal state) were used to diagnose COVID-19 using the widely used deep learning approaches, the New Structure Convolutional Neural Network (CNN) technique.) and the SoftMax method to diagnose. The aim of this study is to diagnose and identify patients with COVID-19. In the second part of the study, a bibliographic search is carried out; In the third part the details of the method and technique are presented, in the fourth part - the experimental data and finally in the fifth part - the results and suggestions that were received after the study.

1.1 THESIS QUESTIONS

In this part of the study, several questions are presented after deep literature in the field of detecting and recognizing covid-19 using machine learning and deep learning techniques. We determine number of questions that we will answer them in this study:

What is the Covid-19?

How can developed deep learning based framework to detect the Covid-19 by analysing the X-ray images?

Why convolutional neural network was used in most of studies and how can developed new structure without using transfer learning techniques?

1.2 RT-PCR TEST

SARS-CoV-2 is an enveloped virus whose genetic material consists of single-stranded RNA. Laboratory diagnostic tests used in COVID-19 infection include two main categories; nucleic acid hybridization-related techniques, the first of which is based on the detection of SARS-CoV-2 viral RNA by reverse transcriptase polymerase chain reaction (PCR) during the acute infection phase; the second one is the techniques based on the detection of immunoglobulin M and immunoglobulin G antibodies or antigenic proteins by serological immunological methods in the subacute-recovery

period in individuals exposed to the virus.7 RT-PCR test in which viral nucleic acid is detected as a diagnostic method for the detection of SARS-CoV-2 virus is considered the gold standard. It is a molecular technique based on amplifying very small amounts of viral genetic material to detectable levels in a nasopharyngeal swab sample (studies using serum, stool, and ocular secretions have also been reported) for testing. In the RT-PCR test, reverse transcriptase (RNA dependent DNA polymerase) enzyme recognizes the relevant region through DNA primers that bind to the target region of the viral genome to replicate and creates complementary DNA (cDNA) from the target segment of viral RNA, and this DNA strand is synchronized with the DNA polymerase enzyme in real time. reproduced [23,24]. The amplification process is repeated for about 40 cycles until the replicated cDNA is detected by fluorescent dye-labeled DNA probes.7 Most of the PCR-based molecular diagnostic tests developed are for SARS-CoV-2's nucleocapsid, spike protein, RNA-dependent RNA polymerase, or envelope. It targets genomic regions associated with proteins. Thanks to advances in molecular detection of the viral genome by replicating it, HMPV, endemic coronaviruses, HBoV, SARS (2003), MERS (2012) and 2019-nCoV (SARS-CoV-2) viruses have been discovered in recent years. See Figure 1.1.

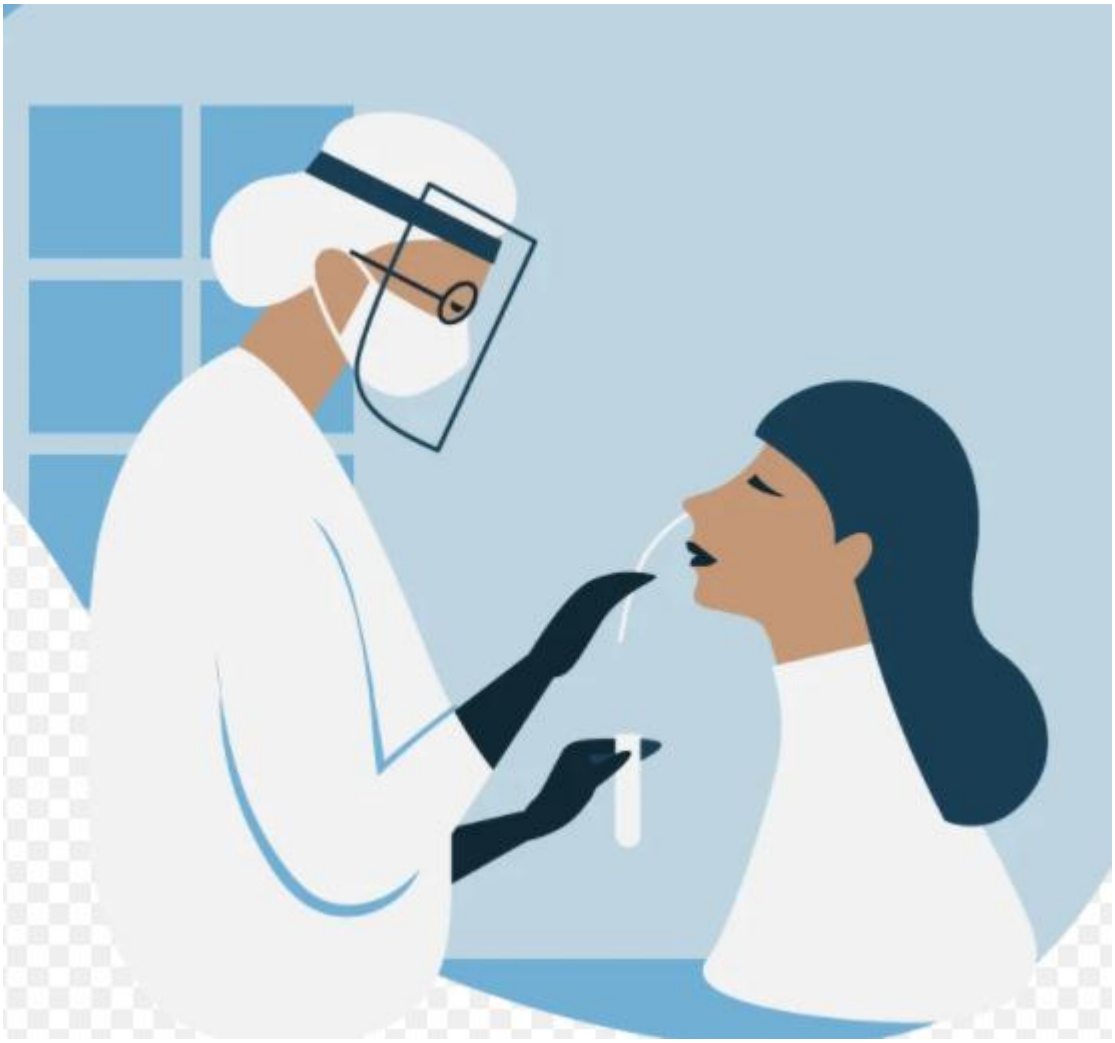


Figure 1.1: Rt-Pcr Test

1.3 X-RAYS

It is a very effective way of looking at bones and can be used to help detect a number of conditions. X-rays are usually performed by trained specialists called radiographers in hospital x-ray departments, but can also be done by other healthcare professionals, such as dentists. X-Ray is a type of radiation that can pass through the body. They are invisible to the naked eye and you cannot feel them. As X-rays pass through the body, they are absorbed at different rates by different parts of the body. A detector on the other side of the body takes the X-rays after they pass and converts them into an image. Dense parts of your body, such as bone, that X-rays find more difficult to pass

through appear as light white areas in the image. Softer parts, such as your heart and lungs, where X-rays can pass more easily appear as darker areas.

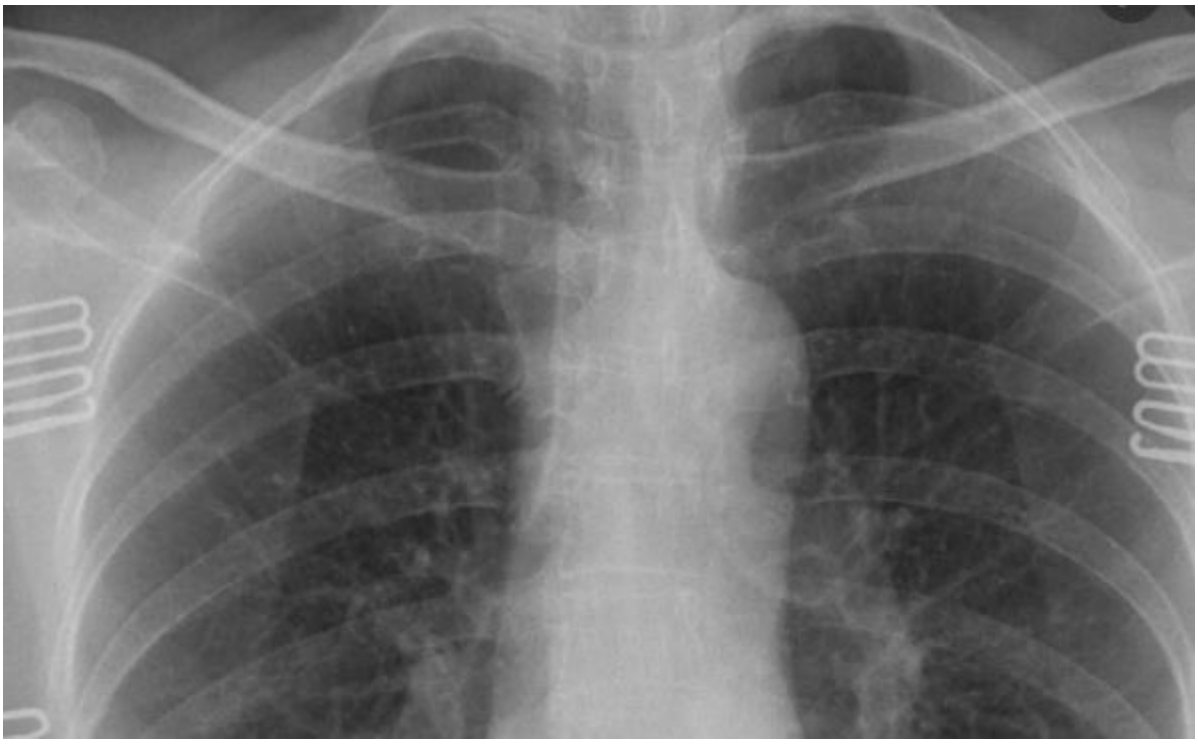


Figure 1.2: X-Ray

2. RELATED WORKS

2.1 LITREATURE REVIEW

In [8] Tawsifur Rahman et al. We proposed a new UNet model and compared it to the standard U-Net model for lung segmentation. The new U-Net model achieved 98.63%, 94.3% and 96.94% accuracy, lung segmentation transition index (IoU), or cube. The gamma-ray amplification method is superior to other methods for detecting COVID-19 on straight segment chest X-rays.. The proposed approach is extremely robust and has comparable performance, increasing the speed and reliability of COVID-19 detection by chest X-ray.

In [9] Stefanos Caracanis et al provide a new approach to COVID-19 detection that uses an artificial antagonist network to generate synthetic images and increase the amount of data available. It also follows a simple architecture and provides two deep learning models proportional to the total amount of data available. Our experiments focused on both the dual classification of COVID-19 and normal cases, as well as various classifications, including third-class bacterial pneumonia. Our model competed with other studies in the literature and the ResNet8 model. The most efficient binary model achieved 98.7% accuracy, 100% sensitivity, 98.3% specificity, and the tri-level model achieved an accuracy between 98.3% and 98.3% with 99.3% specificity. In addition, our model, which adopts the testing protocol recommended in the literature, is more reliable in detecting COVID-19 than the reference version of ResNet8, which prints COVID-19 on X-rays of the anterior sinuses. A good candidate for viewing.

In [10] Hao Quan et al. It provides a deep learning framework that integrates convolutional DenseCapsNet, a new deep learning framework, is based on a combination of dense convolutional network (DenseNet) and capsular neural network (CapsNet) and takes advantage of the respective advantages of convolutional neural networks and their dependence on large amounts of data. To reduce. This method can provide 90.7% accuracy and 90.9% sensitivity as well as an F1 value for COVID detection. 19 has 96% bandwidth. These results demonstrate that the DenseCapsNetDeepFusion neural network is suitable for the radiological detection of novel coronavirus pneumonia.

In [11], Muhammed Attika Khan et al. We present a detailed program for the classification of COVID-19 pneumonia infections using a conventional breast scanner. With this in mind, a 15-layer convolution architecture has been developed for neural networks that extract depth features from wireless image samples. Detailed features are collected at two different levels, a common intermediate group and a fully connected level, and combined using the highest level of detail (MLD) approach. The damage method is then integrated into the main project and the most distinctive features of the human group are selected.

In [12] Sert Panwar et al. We proposed a deep learning method based on the nCOVnet neural network, an alternative rapid detection method that can be used to detect COVID-19 by analyzing X-rays.

In [14] Turkish Tunceretal. A new approach was presented to classify fuzzy trees for the detection of Covid-19. Covid-19 disease is similar to pneumonia, so in this study B. B. We used three classes of datasets, including Covid-19, pneumonia, and traditional chest x-rays. This dataset represents a new machine learning model called a model model. First, we applied a fuzzy tree transform to each medium length image, which was used to derive 15 images from one medium length image (in this study, we created an F tree at three levels. Hat). .. Then there is a separation in the example of these images. A native multicore binary model (MKLBP) is applied to each instance and image to create a function. The most interesting functions are selected by the adjacent Recursive Component Function Selector (INCA). The INCA selected 616 most characteristic traits and transferred them to 16 classifiers in 5 groups.

[18] Placido L. Vidal et al .. Using a series of examples, he proposed a methodology to transfer knowledge from known disciplines to new and less complex disciplines. Using a pre-trained segmentation model based on pathologically independent MRI of the brain, a two-stage information transfer is carried out to create a robust system that enables computerized segmentation of the lung region without sampling. Notebook. Therefore, this methodology has achieved a satisfactory level of accuracy. COVID-19 patients have 0.9761 ± 0.0100, healthy patients with symptoms such as COVID-19 (e.g. pneumonia) have 0.9801 ± 0.014, and healthy patients have 0.9769 ± 0.0111 natural lungs .. .

[19] Emtiaz Hussein and others at CNN have proposed a new model called CoroDet that automatically detects COVID-19 using chest x-rays and computed tomography. The impact of the proposed model has been carefully compared to 10 existing methods for detecting COVID. In the model, which was significantly improved compared to the proposed model, successes were achieved in the 2nd year 99.1%, in the 3rd year 94.2% and in the 4th year 91.2%. It offers the most reliable information compared to existing COVID-19 detection methods. In addition, the X-ray image data set created for the evaluation of the procedure is the largest data set for the detection of COVID.

In [20] Ayan Kumar Das et al., Developing an Automated Covid-19 Detection Model to Identify Patients with This Disease on Chest X-ray.. Model performance was evaluated using three training models, showing that VGG-16 outperformed CNN and ResNet-50. The VGG-16 has an accuracy of 97.67%, an accuracy of 96.65%, a recovery factor of 96.54% and an F1 value of 96.59%. Performance reviews also show that this model is better than the two existing models for testing Covid-19.

[21] Sheetal Rajpal et al. used the ResNet-50 continuous learning function to train the network with a set of preprocessed images to obtain 2048 feature vectors. In the second module, the rate and frame rate were reduced to 64 functions using PCA. Then connect them directly to the neural network to create an array of 16 functions. The third module combines the features of the first and second modules and passes them to the high-density layer and then to the softmax layer to provide the desired classification pattern. In addition to the chest X-rays of Mendelli and Kagurus, chest X-rays of COVID-19 patients from four independent populations with pneumonia and generalized cases were also used. 10 cross-checks to determine the effectiveness of the proposed model. The model provided a sensitivity of 0.987 ± 0.05, 0.963 ± 0.05, and 0.973 ± 0.04 with an overall ranking accuracy of 0.974 ± 0.02 and a 95% confidence interval. .. Normal severity of COVID-19 and pneumonia.

In [22], Sobhan Sheykhivand et al. A new method of automatically identifying pneumonia (including COVID-19) using the proposed deep neural network is introduced. The proposed method uses a chest x-ray to classify grades 2 to 4 in 7 different functional scenarios (health, virus, bacteria and COVID-19). The proposed architecture uses a Contrast Network (GAN) which

combines data transmission and deep learning with LSTM networks for pneumonia classification without the need for feature extraction / selection. Over 90% accuracy in all failure scenarios.

In [23] Osama Shahid et al. We will discuss the role ML plays today in combating viruses, particularly in terms of detection, prognosis and vaccination. In this expedition, they provide a complete overview of algorithms and machine learning models that can combat viruses.

In [24], Fatih Demir proposes Sobel watershed signals and gradient segmentation are applied to the original image to improve performance. the proposed model in the process of preprocessing. Experimental studies were conducted using a publicly available dataset on COVID-19, pneumonia, and normal (healthy) chest radiographs. Overall, the proposed model significantly improves on existing radiological approaches and is very useful for radiologists and physicians to identify, quantify and monitor COVID19 cases during a pandemic. This could be a great app.

[25] Priyansh Kedia et al., Use key diagnostic skills to identify COVID-19 infected patients with chest x-rays to help radiologists and healthcare providers fight this epidemic. We proposed a deeply convolved neural network model, and the experimental results showed that the overall classification accuracy obtained using the proposed three-tier approach for classifying COVID-19, pneumonia, and normal conditions was 98.28% with average accuracy. the answer is 98%. 33% and 98.33% also clearly show that dual classification of COVID-free and COVID-free chest radiographs provides an overall accuracy of 99.71%.

[26] Ines Fekki et al. It is an integrated collaborative learning platform that enables multiple healthcare organizations to detect COVID-19 using chest x-rays without sharing patient information through deep learning. Sent. It describes some of the key features of a relational learning environment, such as providing identical (no IID) and disproportionate online and distributed data. They experimentally show that the proposed integrated learning framework provides competitive results in trained data exchange models based on two different model architectures. These results encourage healthcare professionals to implement a common process and use a wide variety of personal data to build robust models for COVID-19 detection.

[27] Danial Sharifrazi et al combined a convolutional neural network (CNN), a vector support machine (SVM), and a Sobel filter to collect new COVID-19 image datasets from X-rays and X-rays. J. made an offer. Package delivery service. A high-pass filter that captures the edges of the

image using a Sobel filter.. Show 95%. Sobel filters have been shown to improve CNN performance.

In [28] Soham Das et al. The functional surface learning classifier extracted from VGG19 from x-ray images was used to develop a two-step predictive model for healthy patients, pneumonia patients, and COVID-19 patients. This model is used for Normal, Pneumonia, and COVID 3168. We created the dataset by examining CXR images from various sources including the SIRM COVID-19 Database, Chest Imaging (Twitter) , the COVID chest x-ray dataset and x-rays. The experimental results confirm the superiority of the proposed model (precision of 99.26%) over the most efficient estimation method used here (precision of 96.74%).

In [29] E. Gotay et al. Collect global records from around the world, process and extract the number of confirmed cases for a given date. Supplied as a template for a training kit. The model is trained with monitored machine learning algorithms and the number of visitors is expected to increase in the coming days.

In [30] Prottoy Saha et al. Introducing the EMCNet Self-Identification System for Chest X-Ray in COVID-19 Patients. With an emphasis on the simplicity of the model, a convolutional neural network was developed to extract high-quality features from x-ray images of patients infected with COVID-19. Machine learning binaries (Random Forest, Support Vector Machine, Decision Tree, AdaBoost) are designed to detect COVID-19 using extracted functions. Finally, we combine the results of these classifiers to create the most appropriate classifiers for records of different sizes and resolutions. Compared to other new deep learning systems, EMCNet has improved performance: 98.91% accuracy, 100% accuracy, 97.82% response rate, and 98.91% response rate. , 89% F1. This system can provide highly automated COVID-19 detection with instant detection and few false negative results.

In [31] Adeyinka P. Adedigba et al. Introducing COVID-19 X-ray Diagnostic Techniques (CXR) and Explaining the Deeper Installation Issues Related to COVID-19 due to the small dataset and class imbalance of most CXR datasets available. This flexibility allows for quick retraction and prevents the seat post plate from tangling. In addition, we have solved the problem of the high computational load of deep models through the application of memory-based algorithms and efficient machine learning with varying degrees of precision. The model performed well despite

the paucity of data sets. And in general. The verification precision was 96.83%, the sensitivity and specificity 96.26% and 95.54%, respectively. If the model is tested with a whole new set of data, it achieves 97% accuracy without additional training. Finally, visual interpretation of the model results shows that the model helps radiology technicians quickly identify symptoms of COVID-19.

Fei Shan et al. developed a deep learning-based system for automatic segmentation from chest computed tomography images for the diagnosis of COVID-19 disease. They used the "VB Net" neural network for their deep learning-based segmentation. They conducted training using 249 COVID-19 images to segment the infection sites of the COVID-19 disease with computed tomography images. They adopted the strategy in the identification cycle of computed tomography images for education (HITL). They calculated their deep learning-based system from the results of membrane similarity coefficient, volume, percent differences, infection (POI), automatic and manual segmentation. The proposed system obtained the membrane similarity coefficient s as 91.6% and 10.0% in automatic and manual segmentation, respectively. Automatic and manual segmentation achieved a mean POI estimation error of 0.3% for s and all. Validation dataset usually takes 1 to 5 hours compared to full manual identification cases. The 3 model update iterations in their proposed loop strategy reduced the identification time to 4 minutes. The datasets include cases from anterior and posterior x-ray images of patients. As a result of the study using the publicly available dataset, they obtained 96.3% accuracy and 0.151 losses. They defined the deep learning model studied on the dataset images they created from different cases around the world as three false positives and one false negative case. They have automated the diagnostic model of COVID-19 disease [61].

Chuansheng Zheng et al. developed an automatic deep learning-based model for the diagnosis of COVID-19 disease from computed tomography images. They used 3D CT volumes to detect COVID-19. They fed segmented 3D lung regions with segmented 3D lung regions using a pre-trained UNet for each case with a 3D deep neural network to predict the probability of transmission of COVID-19 cases. 499 CT volumes collected from 13 December 2019 to 23 January 2020 were used for training. , 131 CT volumes collected between January 24, 2020 and February 6, 2020 were used for testing. Deep learning algorithms achieved an accuracy of 0.959 ROC AUC and 0.976 PR AUC. The classification algorithms as COVID-positive and COVID-negative provided

0.901 accuracy. They obtained a positive predictive value of 0.840 and a negative predictive value of 0.982. Their algorithm required 1.93 seconds to process the CT volume of a single patient using a dedicated GPU [62].

Joseph Paul Cohen and colleagues presented a score prediction model for the diagnosis of COVID-19 pneumonia from anterior chest X-ray images. Their work aimed to measure the severity of COVID-19 lung infections and pneumonia in general. Images taken from a publicly available COVID-19 database were scored retrospectively by three experts and tagged and created the datasets. Their work predicted the outcome of regression model training on a subset of outcomes from pre-training, with a geographic coverage score (0-8) mean of 1.14. Absolute error (MAE) and lung opacity scores (0-6) achieved an MAE of 0.78 [63].

Ophir Gozes and colleagues have created a dataset from a large number of international CT images for the diagnosis of COVID-19 disease. Using 2D and 3D deep learning models, they presented a system that modifies and adapts existing artificial intelligence models, combining them with clinical understanding. COVID-19 thoracic computerized They analyzed the performance of the system by determining the tomography features and used 3D skin examination. They performed a retrospective experiment to create a "Corona score" by evaluating the evolution of the disease over time in each patient. The datasets included a test set of 157 international patients. Classification results for coronavirus and non-coronavirus cases per thoracic computed tomography studies as a result of their study 0.996 AUC (95%CI: 0.989-1.00) accuracy. The probable study point in Chinese cases achieved 98.2% sensitivity, 92.2% specificity [64].

D. Javor et al developed a simplified deep learning-derived machine learning classifier for the diagnosis of COVID-19 disease. Training the open source dataset consisting of 6868 computed tomography lung images of 418 patients and validation subsets. They achieved 95.6% Area Under The Curve (AUC) overall accuracy in the independent test dataset of 90 patients [65].

Govardhan Jain et al. have carried out a four-stage deep network model design using existing resources and advanced deep learning techniques for rapid diagnosis of COVID-19 disease. They classified a total of 1832 x-ray images as healthy, bacterial and other viral origin. As a result of the study, they obtained 97.77% classification accuracy and 97.14% accuracy for COVID-19 disease diagnosis [66].

Shuo Wang and colleagues created a dataset containing 5372 images from seven different cities for the diagnosis of COVID-19 disease from lung computed tomography images. Performed operations on four external validation sets, yielding 86% AUC from viral pneumonia 87% AUC from other pneumonia. The systems succeeded in dividing patients according to high and low risk groups [67].

Y.Pathak et al. obtained 96.22% training and 93.1% test accuracy rate by using the CNN architectural model for the diagnosis of COVID-19 infected patients [68].

Zhua Tao et al. created a dataset by combining the aggregated datasets used in previous studies and new images from the healthcare institution, and conducted training and testing on different deep CNN networks [69].

M.Türkoğlu suggested the ELM-based Deep Neural Network (ELM-DNN) method of lung computed tomography images for the diagnosis of the disease, and performed deep feature extraction from CNN and CT images. The model suggested in the diagnosis of the disease is 98.28% achieved success [70].

Lu Huang et al. classified the lung computed tomography images for the diagnosis of COVID-19 disease as mild, moderate, severe and critical according to the laboratory and clinical picture of the patients. The study was compared with commercial software and follow-up computed tomography images [71].

Shervin Minaee et al. created a dataset from 5000 chest X-ray images from publicly available datasets. They used transfer learning on a subset of 200 radiograms to train four popular convolutional neural networks, including ResNet18, ResNet50, SqueezeNet, and DenseNet-121, to identify COVID-19 disease. As a result of their studies, most of their networks achieved a sensitivity rate of 98% (±3%), and the specificity rate reached approximately 90% [72].

Song Ying et al. created a data set from the computed tomography images of 88 patients diagnosed with COVID-19, 101 patients with bacterial pneumonia, and 86 healthy people from hospitals in two cities of China for the diagnosis of COVID-19 disease. They developed a computerized tomography diagnosis system with deep learning methods for the diagnosis of the disease. As a result of their studies, they achieved a sensitivity of 0.99 AUC and 0.93 [73].

Halgurd S. Maghdid et al. aimed to create a dataset with x-rays and CT scan images from many sources. In their study, they applied Convolutional neural network (CNN) and modified pre-trained AlexNet model to the datasets. Up to 98% accuracy and modified CNN on studied models using 94.1% accuracy [74].

Xuehai He et al. propose a self-supervised learning method with a synergistically integrating transfer learning Self-Trans approach in their work. Self-Trans approaches obtained F1 score and AUC values of 0.85 [75].

Ezz El-Din Hemdan et al. have run their x-ray images with seven different deep convolutional neural network models such as COVIDX-Net, modified Visual Geometry Group Network (VGG19) and the second version of Google MobileNet. VGG19 and dense convolution network (DenseNet) models obtained f1 values of 0.89 and 0.91 for normal and COVID-19, respectively [76].

Amine Amyar et al. presented an automated classification segmentation tool to aid in the diagnosis of COVID-19 disease. They aimed to evaluate the lesions in the segment and the severity of pneumonia from computed tomography images. Their architecture is derived from a common encoder for three decoded feature representations consists of. Datasets were created from 449 COVID-19, 425 normal, 98 lung cancer and 397 different pathology types. As a result of their studies, they obtained a coefficient higher than 0.88 for segmentation and a classification higher than 97% for the area under the ROC curve [77].

Muhammad Farooq and his friend used 3 different steps to fine-tune the ResNet-50 model of the CNN architecture pre-trained with open-access datasets, and they named these steps as COVID-ResNet. Their study achieved an accuracy of 96.23% across all classes [78].

Using deep learning techniques, Xiaowei Xu et al. created a data set from IAVP, healthy cases and cases diagnosed with COVID-19 through lung computed tomography images. They created the CT images they used in their dataset from three designated hospitals in Zhejiang. Datasets total 618 CT contains the sample. Their study achieved an overall accuracy rate of 86.7% for all computed tomography image cases [79].

Md Zahangir Alom et al. proposed the Inception Residual Recurrent convolutional neural network approach with transfer learning for the detection of COVID-19 disease. They used the NABLA-N network model to segment COVID-19 infected regions. The detection models achieved approximately 84.67% test accuracy from x-ray images and 98.78% accuracy from CT images. As a result of their studies, they obtained a superior sensitivity of 1.00 (95% confidence interval (CI) 0.95, 1.00) and an F1 score of 0.97 [80].

Aayush Jaiswala et al. used pre-trained deep learning in their article due to limited image datasets. They propose DenseNet201-based deep transfer learning (DTL) to classify diagnosed patients. The impact of their proposed DTL model on the diagnosis of COVID-19 disease. As a result of their experiments to evaluate its performance, they found that DTL-based COVID19 classification outperforms competitive approaches [81].

Eduardo Luza and colleagues extended their work by using the EfficientNet architecture for the diagnosis of COVID-19 disease from chest X-rays. They made classifications in datasets consisting of 13569 chest X-rays. As a result of their studies, they obtained 93.9% overall accuracy, 96.8% and 100% positive predictions [82].

Resul Tümer et al. aimed to diagnose COVID-19 disease in multi-class datasets consisting of different lung tomographies with artificial intelligence and machine learning techniques. They used convolutional neural networks and deep neural networks and k-nearest neighbor methods, which are machine learning algorithms, for the diagnosis of the disease [83].

2.2 MACHINE LEARNING

Considering the age we live in, it is seen that the use of technology in many fields such as health, education, finance, security and transportation has become quite common in almost every country. The widespread use of technology has resulted in an increase in the amount of data produced by real and legal persons. At the same time, with the development of technology, storing and accessing data has become both easier and less costly compared to the past. Depending on these, data has started to be considered as valuable as finance in our world. So much so that the new generation banks and intermediary institutions, most of which provide services in the electronic

environment, have started to carry out transactions such as fund transfer and stock purchase without commission fees. The reason for making the mentioned transactions free of charge is to expand the customer base because of the desire to have a large amount of data. Because, in case of having more data, it is possible to make strategic decisions and make forward-looking predictions. For example, the United States It is even possible to predict the election results in the country from the stocks, which are heavily demanded by investors in the United States [101].

In such a case, the data related to the customer and the customer's transactions, which enable the said estimation to be made, are more valuable than the fees such as commissions to be received from the customer. However, it should be noted right away that due to the increase in the amount of data as mentioned above, making the aforementioned estimations over the data is not in a position where people can do it alone. Essentially, making predictions on huge data requires a great deal of experience, expertise in the relevant field, and foresight ability. It is possible to say that this process will not be very effective when considering the time it takes to analyze huge data and make predictions by people, even if it is assumed that they have these qualifications, ignoring the fact that not every person can have this experience, expertise and ability. At this point, machine learning, which is based on the philosophy of learning from data and used in many fields, comes to the fore in analyzing data.

When the machines used in the early stages of industrialization and the following periods are compared with the machines used today, it is understood that the introduction of computers into our lives carries the concept of machine to a very different dimension. So much so that today, when the machine is mentioned, many hardware such as computers, telephones, air conditioners, cars and televisions, which contain software, come to mind. The point that needs to be emphasized in the context of machine learning is that the learning ability of the mentioned hardware is gained by computer software. In the software in question, instead of explicitly programming the hardware to do a task, the approach of developing programs that can learn to do the task is used. In other words, in traditional programming, every step necessary for the fulfillment of a task is coded by the programmers, while in machine learning, algorithms learn to solve the same task themselves by training on data. Similarly, machine learning by Blum; has the ability to learn rules from data, can update itself within the scope of changes, and improve its performance with the experience gained.

Considering the age we live in, it is seen that the use of technology in many fields such as health, education, finance, security and transportation has become quite common in almost every country. The widespread use of technology has resulted in an increase in the amount of data produced by real and legal persons. At the same time, with the development of technology, storing and accessing data has become both easier and less costly compared to the past. Depending on these, data has started to be considered as valuable as finance in our world. So much so that the new generation banks and intermediary institutions, most of which provide services in the electronic environment, have started to carry out transactions such as fund transfer and stock purchase without commission fees. The reason for making the mentioned transactions free of charge is to expand the customer base because of the desire to have a large amount of data. Because, in case of having more data, it is possible to make strategic decisions and make forward-looking predictions. For example, the United States

It was stated that he was interested in designing software (Blum, 2021). Therefore, a self-learning of software using data in machine learning is to be realized. In this context, in the field of machine learning, in summary;

i. Understanding computerized systems when it comes to learning machine and the fact that the element that enables learning in systems is computer software,

ii. In the definitions made, the subjects of learning with experience and increasing performance Emphasizing that the increase in performance with experience is accepted as learning to be made,

iii. Datasets in which the experience can be essentially the learning event to be,

iv. The learning, the information in the data set that will be useful according to the purpose of the work detection and self-adaptation of the software to the changes that occur that it can provide,

v. In order to test what level of learning has taken place, some using the criteria

vi. Parameters in the models created to improve the learning level making adjustment,

It is possible to count these issues as the important points of the subject Machine learning is basically supervised learning, unsupervised learning and It is divided into three categories as reinforcement learning. supervised learning, developed based on examples consisting of both input and output parameters. The model is the matching of new inputs with outputs. First to specify data representing these examples in machine learning should be consists of attributes similar to columns in the file. Attribute measurable is a feature and the type of attack in an application, such as the prediction of terrorist events, may include qualitative characteristics such as the type of weapon used in the attack, the type of target of the attack, In this case, the output variable may be the terrorist group that carried out the action. Therefore, In supervised learning, the attributes in the examples used in the development of the model It is analyzed and the relationship between these attributes and the output variable with a label, that is, a specific class, is tried to be understood. This process produces a function that can be used to predict outputs from samples that the model has not seen before. Then, using this function, estimations are made using the test data, which is not used in the development of the model but whose output values are known, and the obtained results are compared with the actual results. This comparison is made using loss functions, which measure how much the estimates deviate from the true values. As a result, if the predictions are found to be sufficiently accurate, it is interpreted that the algorithm generalizes. Unlike supervised learning, there is no labeled output of the training data to be used in creating the model in unsupervised learning. That is, in this method, there is only input data and the main purpose is to create data patterns is understanding. This method is mostly used in applications related to data clustering and detection of anomalies in the data set. In reinforcement learning, instead of directly programming the agents on which the software is installed to perform a task, the software in question learns by interacting with their environment. This learning process takes place by winning a reward or defeating a penalty depending on the action taken in the face of the situation and is basically based on the trial and error method [102].

Before evaluating the possible effects of machine learning explained above on forensic applications, it would be correct to talk about the general working mechanism of the forensic field. In this direction, just as the advancement of technology has contributed to the development of fields such as data science and machine learning, it also

It has also caused crime and criminal elements to become widespread in the electronic environment. In other words, with the introduction of devices such as computers into our lives, there are now crimes that take place in the electronic environment and leave traces in this environment, and the mentioned computer-like devices appear as both an element and a tool of the crime. At this point, forensic computing; focuses on proving the data in electronic devices with data storage capability such as computers, mobile phones and external disks by using methods accepted by legal authorities. Even if there is an accepted process for the work to be done for this purpose, there are differences in practice depending on the situation possible to see. However, the generally accepted forensic information process.

consists of stages. Therefore, it is aimed to obtain data related to the event studied by applying the processes specified in the mentioned stages in forensic informatics and to clarify the relevant event as a result, and at this point, the application area of machine learning methods in forensic informatics emerges.

In this direction, machine learning has a wide variety of applications related to forensic cases. In the most basic sense, machine learning algorithms; It can be used for purposes such as analyzing big data on crimes, revealing criminal behaviors, and detecting criminal structures. Because the use of machine learning algorithms for crime detection provides researchers with the opportunity to query data from many different sources such as social networks, wired networks and cloud environments and to analyze this data even if it reaches very large sizes. This essentially refers to the analysis of large amounts of data with machine learning algorithms in order to predict certain behaviors such as criminal intent and activity of certain elements such as criminals and terrorist groups.

For example, learning from past events can be performed in order to predict future criminal behavior in areas such as theft, money laundering and cyber attacks. In addition, it is possible to conduct remote data analysis through the software used in forensic events.

One of the most important advantages of machine learning for the people investigating the events is that the data related to the crime can be brought into a format that can be understood and easily interpreted. The important point here is essentially to obtain the data on which the researchers will work and to the fact that qualified data regarding the events is kept in the relevant institutions.

The possibilities that machine learning can provide in the field of forensic investigations mentioned above, profiling and estimation activities, predicting when and where crimes may occur, or predicting who the perpetrators may be in a crime that has occurred, similar opportunities to combat terrorism related to medicine, finance and forensic events. as well as for many fields. So much so that recently, machine learning algorithms are frequently used by relevant institutions to examine different factors of terrorism. In this thesis, machine learning models, the details of which are given in Chapter 3, were created on the data set related to terrorism in order to shed light on the work and future applications in the field in question.

2.2.1 Classification

In Section 2.2, it was stated that the category of machine learning where both inputs and outputs are together in the data set is supervised learning. In this section, after the general information about classification is mentioned, basic information about the classification algorithms belonging to the supervised learning category and generally used in the thesis study is given. Classification, which aims to assign data to one of the predetermined classes, has also previously been widely used in fields such as terrorism, intelligence, and homeland security.

It is one of the main areas of machine learning. In models created with classification algorithms, it is aimed to place a piece of data that actually belongs to a class but whose class is unknown by the model at the beginning, to one of the known groups. In order to fulfill this purpose, the samples in the training data set, which was used as the basis for the creation of the model, and whose labels are known, are used. Therefore, both input parameters and output values are known at the stage of creating the model in classification.

Various approaches are used to classify data sets, and it is possible to evaluate classification algorithms as probability-based and non-probability-based. Probability-based classifiers show the probability of any sample belonging to a certain class, while non-probability-based, that is, discrete classifiers, show the predicted class of the sample in question. In addition, the aforementioned classifiers can be considered as binary or multiple classifiers. Here, binary classification refers to the separation of data in a certain data set into two groups with the specified classification rule, while multi-classifiers refers to the separation of more than two groups.

i. Naive Bayes Classifier

The Naive Bayes algorithm is widely used because it is easy to implement and understand and is useful for large data sets, and it is a method in which classification is performed with the conditional probability approach. In this algorithm, which assumes that the existence of an attribute in a certain class is not related to the existence of any other attribute and is based on Bayestheorem, the probability calculation is made about which class a new data belongs to, by using the existing and already classified data. During this process, as in other classification processes, the learning process is carried out on the training data set. It is known that the Naive Bayes algorithm, in addition to being simple to apply, can even outperform sophisticated classification methods in some cases. The operation of the algorithm starts with the conversion of the data set to the frequency table. Then the probability table is created by calculating the probabilities. Finally, the formula given in figure 2.1 below is applied to calculate the posterior probability of each class. The estimation of which class the sample belongs to is made as the class with the highest posterior probability. $P(A/B) = \frac{P\left(\frac{B}{A}\right)XP\left(\frac{B}{A}\right)}{P\ (B)}$ (2.1)

It will be easier to understand if the mentioned operation is shown through an example. Accordingly, if the frequency table created from the training data set is assumed to be as in table 2.1 in terms of estimating which organization committed the terrorist act according to the type of weapon used, the probability table to be created using this frequency table will be as in table 2.1.

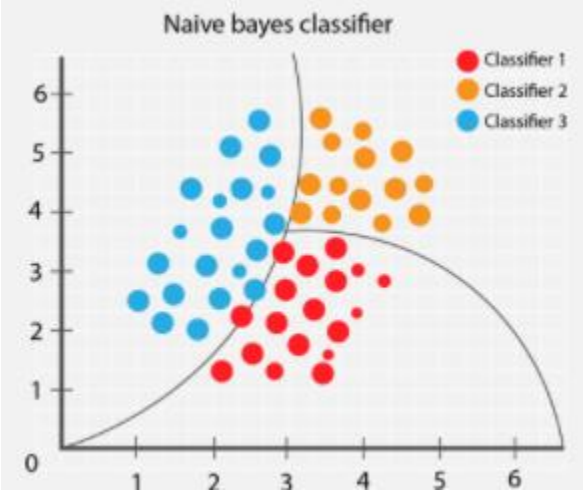


Figure 2.1: Naïve bayes [84]

ii. Logistic Regression

The logistic regression algorithm, although there is a regression expression in its name, is an algorithm used in both classification and regression. With this algorithm, it is tried to estimate the classes of the samples, that is, the dependent variable, with the attributes that are the input parameters, that is, the independent variables. In other words, it is aimed to find the relationship between dependent variables and independent variables that should be categorical in the logistic regression algorithm and to model it in the best way. In addition, the algorithm in question, the dependent variable to be estimated; It is called binary regression if it has two categories, ordinal logistic regression if it has more than two categories with a row between them, and multi-category logistic regression if it has more than two categories with no rows between them. As stated above, in the logistic regression used to understand the relationship between the probability of occurrence of the dependent variable and the features, the output value between 0 and 1 is obtained by using the sigmoid function, whose formula is given in Figure 2.2 below. The classification process is completed according to the output value in question and the threshold value determined. For example, if the output value obtained by using the attributes in the sample for a class is found to be 0.80, which indicates that the sample in question belongs to the tested class with 80% probability, and if the threshold value is determined as 0.5 at the same time, the relevant sample is included in this class. will be.

One of the issues to be considered regarding the logistic regression algorithm is that the input parameters, which are called independent variables, should also be independent from each other, that is, there should not be a high level of correlation between them. Otherwise, if the features can be written as functions of each other, collinearity will occur. In addition, this algorithm does not require scaling of features. However, the logistic regression algorithm does not perform very well when there are too many features or when there are too many categorical variables in the features.

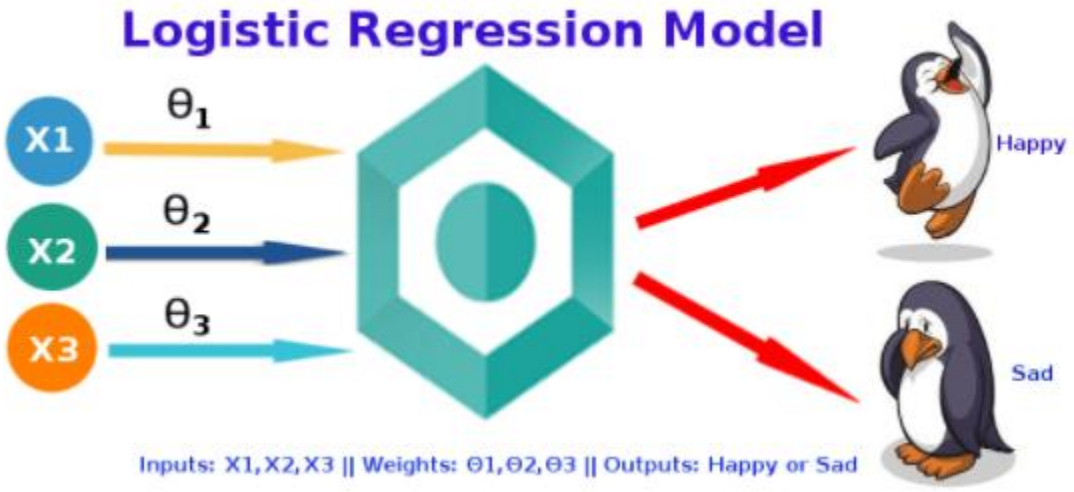


Figure 2.2: Logistic regression [85]

iii. K-Nearest Neighbors (KNN)

KNN is a simple and easy-to-control machine learning algorithm that can solve classification and regression problems. The ANN algorithm assumes that these objects are in the immediate vicinity. In other words, they are close [39]. Look at the picture above. In most cases, similar data points are related. It is based on the assumption that the ANN algorithm is valid enough to make the algorithm useful. The ANN diagram reflects the concept of similarity (also called Proximity Proximity) in some calculations that can be learned as a child in calculating the distance between points. The mathematical model presented in Eq (2.2)

The K-nearest neighbor algorithm is widely used for classification and is an algorithm based on distance calculation. The algorithm in question performs the classification process with the lazy learning method. Namely, in the implementation of the algorithm, all training data is stored and the model is not created until a new sample needs to be classified. of a new sample when it is necessary to classify and predict, unlike other algorithms, instead of using the parameters determined by preprocessing, the classification process is completed by using the training data set each time. Thus, the training dataset

No probability assumptions are made regarding the distribution. Therefore, the k-nearest neighbor algorithm is a non-parametric algorithm. In order to use the algorithm, first of all, a training data set is created from the examples with certain class values and the distance function to be used with

the k parameter is determined required. Then, the distance of the sample to be classified to the data in the training set is calculated using the distance function determined for the classification of any sample using the k-nearest neighbor algorithm. The k samples determined after the calculation are selected from the training set. This new selected dataset is called the classification dataset. Finally, the classification process is completed by assigning the class of the sample, which is tried to be estimated to belong to which class, as the most common class in the classification data set.

$$R_R\left(C^{Bayes}\right)\left\{B_1\frac{1}{k}+B_2\left(\frac{k}{n}\right)^{4/d}\right\}\{1+0(1)\} \tag{2.2}$$

The distance criteria used in the implementation of the k-nearest neighbor algorithm may vary depending on the features in the data set to have categorical, numerical and binary values or consist of all of them. Euclidean and Manhattan criteria are the most frequently used criteria in the aforementioned algorithm. Choosing the above-mentioned k-value is a problematic issue as well as being very important to obtain a high level of accuracy. In the literature, it is recommended to cross-validate this process, that is, to try different k values for the same data set. The importance of choosing the K value is illustrated in figure 2.3 below. In fact, it is seen that the result differs if the k value is selected as 1, 5 or 10 for the classification of a new shape as a circle or a square circle. Also, in the k-nearest neighbor algorithm, which is very simple to understand, there may be more computation time since all the computation for making the prediction is done during testing rather than training. In addition, if there is an imbalance between the classes, for example, if a class is more in the training set compared to the other classes, it may be necessary to apply the normalization process in order not to have a tendency for this class.

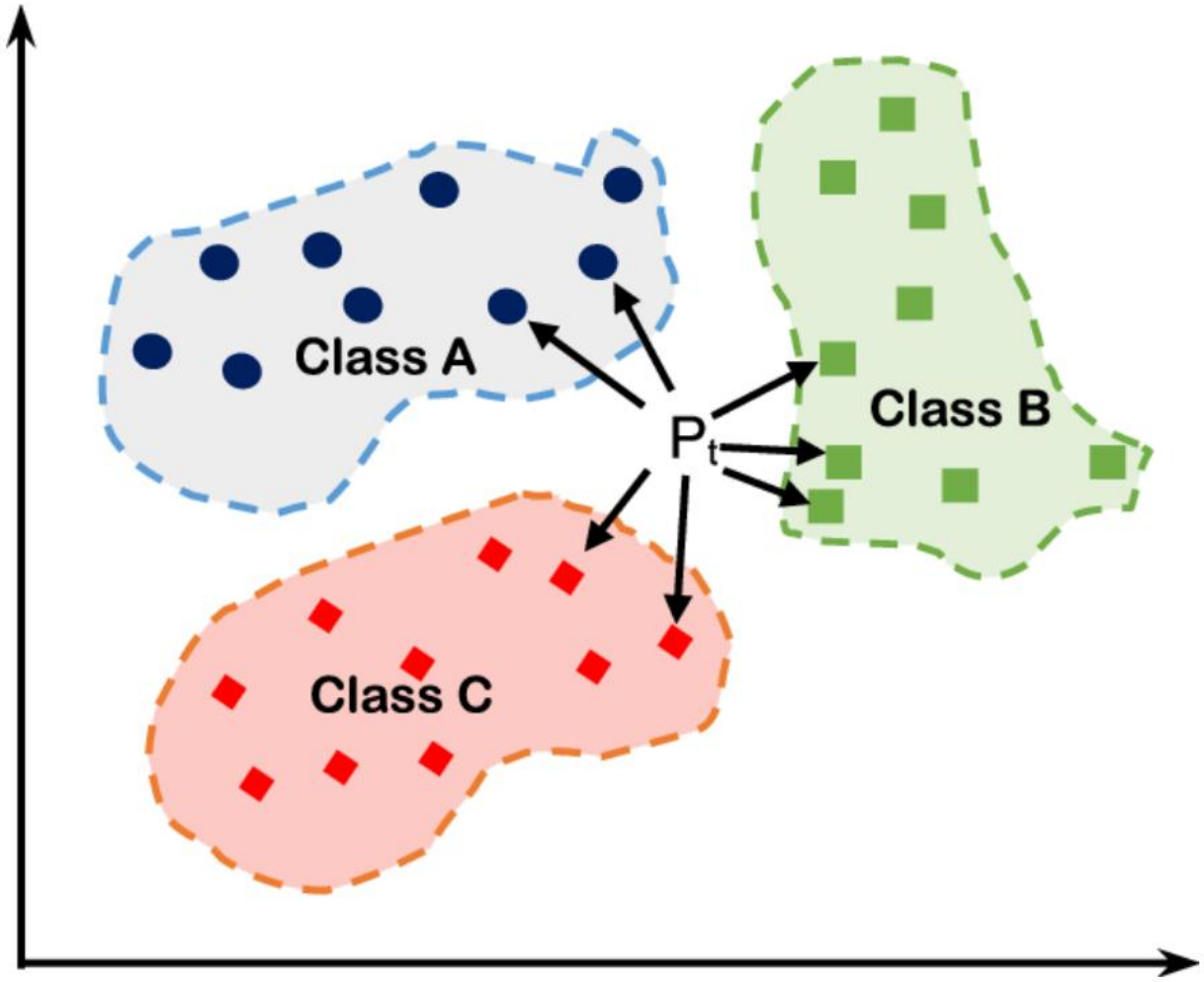


Figure 2.3: KNN [86]

iv. Decision Tree

Decision trees consist of roots, nodes and leaves, similar to the tree structure used in data structures. The purpose of the decision tree algorithm in this structure is to develop a model that can predict the value of the target variable by learning the decision rules extracted from the features. For this purpose, dependent and independent variables are represented in a tree-like structure in which a set of test questions and conditions are arranged. In the tree structure in question, the leaf nodes will include the class labels to which the samples to be tested will be included, the non-leaf nodes, that is, the decision nodes, including the node located at the top of

the tree and called the root node, the test process applied on the attributes, and the aspects to be acted upon as a result of the test process, each branch specifies the result of the test operation, namely the class labels. In the decision tree algorithms, which include the attribute test conditions to separate nodes with different attribute values, the inputs are entered starting from the root node and the attribute values of the tested sample are asked questions determined in the tree structure, and the appropriate branches are followed in line with the answers received. This process is continued until the leaf node is reached, and when the leaf node is reached, the tested sample is assigned to the class label represented at the node in question and the classification process is terminated. Decision in progress as specified Since trees can be easily visualized, the decisions taken can be monitored and understood effectively. For this reason, decision tree algorithms are widely used in classification problems is used. Decision tree algorithm, discrete and non-discrete can use data. However, it is observed that the classification error increases if the training data set is small compared to the number of classes in the data set, that is, if there are too many leaf nodes compared to the branches. Also, the algorithm in question is prone to the problem of over-learning by rote learning. In order to prevent memorization and to prevent decision trees from being too complex, pruning can be done, and in this process, leaves are simply replaced with the lower branch. One of the most important aspects of decision tree algorithms is deciding how to build the tree. At this point, there are various approaches to obtain a flexible and highly accurate decision tree, and these approaches generally involve a gradual observation of the tree. In the implementation of these approaches, the ID3 algorithm comes to the fore with entropy and information gain. In general, the ID3 algorithm uses the technique of determining the closest state to the result by moving down from the root. Entropy focuses on the possibility of unexpected situations and the training set

It takes 0 if it is homogeneous, and 1 if the values in the training set are equal to each other. The small entropy value calculated for any feature indicates that the said feature is an important feature for the ID3 algorithm. On the other hand, information gain means the opposite of entropy. During the creation of decision trees, the features with the highest information gain value can be selected. In addition, it is possible to have information about the importance degree of the features with the decision tree algorithms that act with the logic of induction.

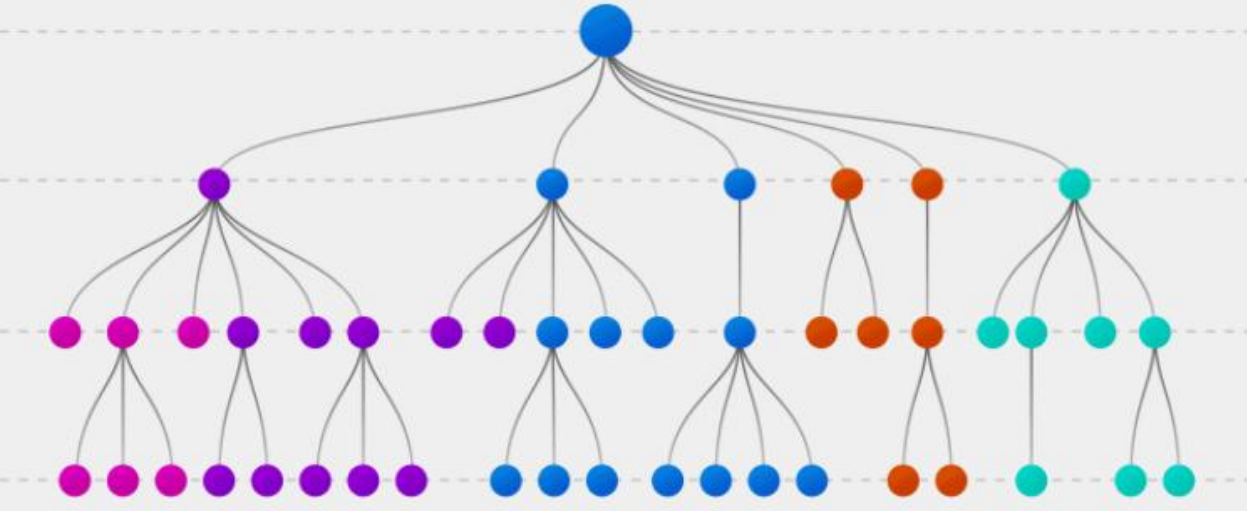


Figure 2.4: DT [87]

v. Random Forest Algorithm

In section 2.2.4 above, it was stated that one of the biggest problems with the decision tree algorithm is memorization by the algorithm in question. In the random forest algorithm, which can simply be expressed as a collection of decision trees, it is aimed to prevent the over-learning problem mentioned. For this purpose, a large number of decision trees are created by randomly selecting subgroups from both the data set and the features. These decision trees created are combined in a structure called forest. In other words, while randomly selected features are used in the creation of the aforementioned decision trees, sub-datasets obtained by substituting the original dataset are used in their training. Therefore, in the random forest algorithm, as in the decision tree, instead of obtaining a single tree by branching the nodes by using the best branch among all the features, instead of obtaining a random tree.

It is possible to create a desired number of trees by dividing the nodes into branches by choosing among the selected attributes. In these trees in the forest, which consists of different trees, pruning is not performed as in the decision tree algorithms, which contributes to the increase in the accuracy of the algorithm. The reason why pruning is not done is that, as it can be understood from the above-mentioned points, since learning is carried out on different subgroup training sets in this algorithm, the over-learning problem is already reduced. On the other hand, the negative side of this structure is that in addition to the increase in computational complexity, a single output that can be interpreted for the trees created, as in decision trees, cannot be given. At the stage of

estimating over the sample, each decision tree created makes an estimation independently, and the class with the most estimation, that is, the majority of the vote, is determined as the class of the sample. Another feature of the random forest algorithm, which is one of the ensemble methods based on the philosophy that many weak estimators can come together to form a strong estimator, is that it provides information about the importance of features, which is also discussed in the section on feature selection of the thesis.

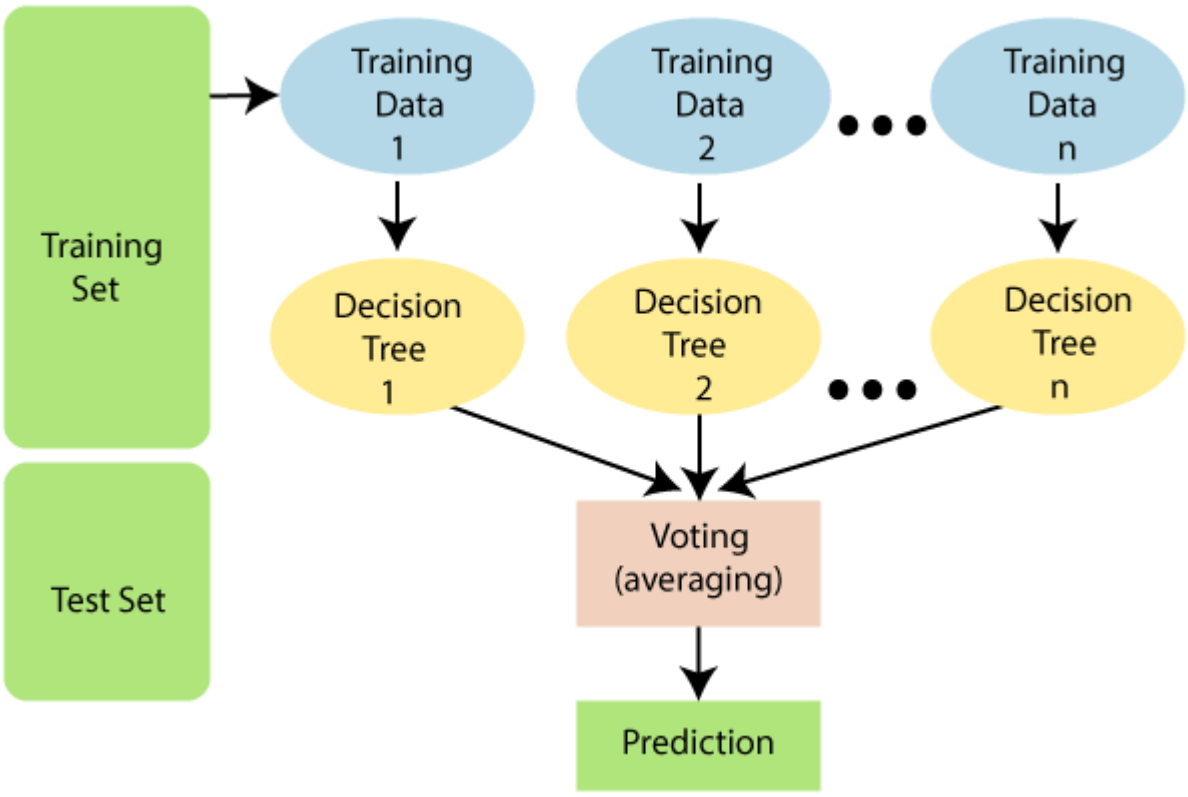


Figure 2.5: RF [87]

vi. Support Vector Machines

Support vector machines in regression and classification operations in general, it aims to find the best classification function that can be used to distinguish members of two classes. In other words, the support vector machines minimize the classification error by choosing the line with the highest margin, while using the features to separate the two classes, they maximize the classes.

It tries to find a plane or hyperplane that can separate it properly. As an example, as it can be seen in Figure 2.4 below, in order to apply the aforementioned algorithm for the two classes, first of all, two parallel vectors that minimize the error are selected, and the distance between them is maximized.

This method is generally used in cases where the data can be linearly separated, but it can also be applied in cases where there are data that cannot be linearly separated, since it is possible to make the data linearly separable by means of kernel functions such as sigmoid, polynomial, linear and radial basis. That is, a larger size is carried out for nonlinear data sets using kernel functions. For example, in figure 2.5 below, the mapping of data from 2-dimensional space to 3-dimensional space is shown, and as can be seen from the figure in question, the data has become linearly separable.

When we look at the above definition of vector machines and other explanatory expressions, it is especially mentioned about the process of separating into two classes. However, this does not mean that it is not possible to use support vector machines for applications with more than two classes. Therefore, support vector machines can also be applied to problems with multiple classes. At this point, it is possible to act with a one-versus-all approach for the separation of classes. Namely, firstly, a class label is selected and this class is labeled as 1 and other classes as -1, that is, a comparison of a class with all other classes is made. This process is applied separately for all classes and a number of classifiers equal to the number of classes are obtained. In other words, as many as the number of classes

The support vector machine is trained, and in the estimation phase, the output of each support vector machine is compared and the safest result is preferred. Support vector machines can be applied in a wide variety of fields because they can work with high generalization performance without the need for a priori knowledge. However, it should be taken into account that support vector machines are inefficient in terms of computational cost.

vii. Artificial neural networks

Artificial neural networks, inspired by the human brain and known as universal function estimators, can map both linear and nonlinear functions. In other words, artificial neural networks are used to predict any function that represents the relationship between an x input and y output,

provided that there are enough hidden layers between them. Therefore, it can be easily applied to classification problems.

Artificial neural networks consist of computational units inspired by biological neurons, called artificial neurons, and the connections between these neurons. Each neuron in the network takes several input values to produce an output value. In artificial neural networks, each input of neurons, namely connections, has coefficients called weights. These coefficients show how the inputs together with the bias value affect the output. Thus, the formula used to obtain the output from the inputs is given in figure 2.6 below, where x represents the inputs, w represents the weights of the inputs, and b represents the bias value. In the mentioned formula, the weighted sums of the inputs are calculated by using the bias value and the weights. Since the bias value is independent of the inputs, it is a parameter used for the translation of the line represented by the formula, and its number in an artificial neural network is equal to the number of neurons [117].

After calculating the weighted sum with the above formula, a nonlinear activation function σ is used to calculate the neuron's output. This activation function makes the weighted sum non-linear. The nonlinearity of the activation function is very important in terms of learning the nonlinear complex tasks of the network. artificial nerve

activation functions that are generally used in networks; sigmoid, hyperbolic tangent (tanh) and Rectified Linear Unit (ReLU) functions. The graph of the sigmoid function, which is the most widely used of these functions, is given in figure 2.6 below, and as can be seen from the figure, this function converts the inputs it receives into a real number in the range of 0 and 1. Therefore, it can be easily used in artificial neural networks that only need to give positive results, such as probability calculation. Although the hyperbolic tangent function is similar to the sigmoid function, the outputs of this function are between -1 and 1. The RELU function assigns the input value for positive inputs and 0 for negative inputs to the output.

Figure 2.8 below shows a simple neural network model with input and output layers and a hidden layer. There is no restriction to have a single hidden layer in artificial neural networks, and it is possible to increase the number of hidden layers depending on the designed architecture. In this case, it should be taken into account that as the number of layers increases, the complexity of the artificial neural networks will increase and the training time will increase.

In the training process of artificial neural networks, the weights of the neurons are adjusted. In other words, it is tried to find the weight values that will give the closest result to the real outputs by using sample data in the training process. It is important that the function, which is obtained as a result of the training process and can define the y = f (x) relationship between the inputs and the output, can be generalized to the data not included in the training data set. Accordingly, the ability of an artificial neural network to predict the correct output for a given input needs to be measured effectively, and this is done with functions called loss functions. The loss function, in its simplest terms, calculates the difference between the actual output and the output predicted by the network. When we look at the commonly used loss functions, it is seen that the mean square error function, which calculates the square of the difference between an actual and estimated value, comes to the fore. In this context, it can be said that the purpose of training a neural network is to minimize the value of the selected loss function by adjusting the weight and bias values of the neurons. The basis of the training process is mostly back propagation algorithm. A training process in which back propagation algorithm is used starts by assigning the weights to a random value. Afterwards, the process of moving forward and backward in the network is repeated until the value of the loss function is minimized. In the forward movement phase, the output value of the network and the value of the loss function are calculated. In the step backwards, how should the weights be used to reduce the value of the loss function.

There is a gradient descent algorithm that needs to be adjusted. In line with this process, the weights are updated in small amounts within the framework of the learning rate parameter of the artificial neural network. In the gradient descent algorithm used to determine how to update the weights to reduce the value of the loss function, the gradient is defined as the change in the value of the loss function when a weight parameter is updated. The algorithm in question calculates the gradient of each network parameter using the chain rule and uses the gradient direction to decide how to reduce the value of the loss function. However, it should be noted that for large data sets, calculating the gradient of each sample will be time consuming. For this reason, stochastic gradient descent algorithm can also be used to speed up the process. In this algorithm, the input set is divided into heaps and the gradient calculation is made for the heaps. Stochastic gradient descent algorithm, together with variants such as Adam, Adagrad, RMSprop, are widely used algorithms for training [118]. The training process for reducing the value of the loss function, whose steps are explained above, continues until the value of the relevant loss function can no longer be reduced,

and when this point is reached, the training process is completed. Thus, the created artificial neural network becomes available for new samples.

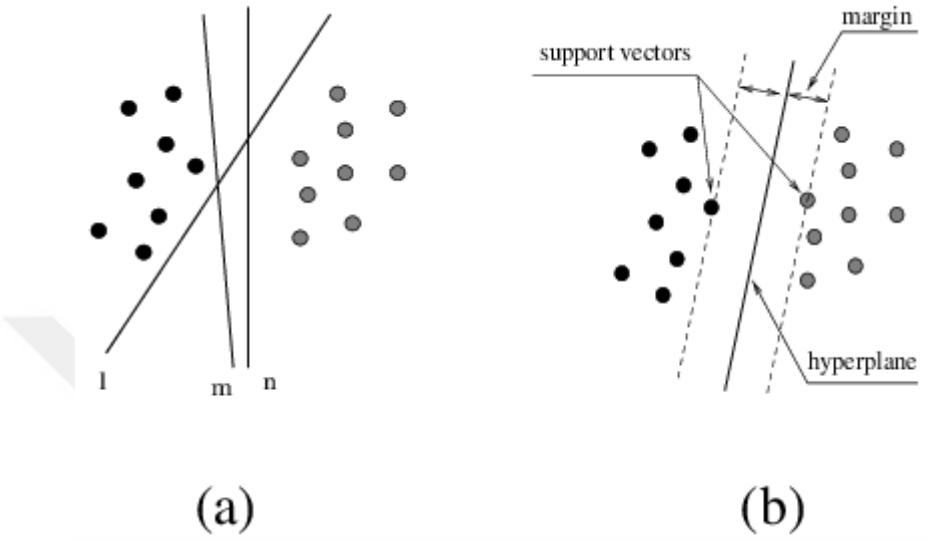


Figure 2.6: RF [88]

2.2.2 Feature Selection

Machine learning algorithms are capable of learning from data without the direct intervention of the user. Therefore, the performance of these algorithms is highly dependent on the data set used. Considering that the information is represented using features in the data set, it is necessary to determine the correct features for the purpose Machine learning algorithms are capable of learning from data without the direct intervention of the user. Therefore, the performance of these algorithms is highly dependent on the data set used. Considering that the information is represented using features in the data set, it is important to determine the right features suitable for the purpose. At this point, in the simplest sense, the data set It is necessary to focus on feature selection, which allows to decide on the features that may have an impact on the classification result and to eliminate other features. It is aimed to reduce the size of the data set, to minimize the working time of the algorithms, to prevent the algorithms from memorizing and to increase the accuracy rates as a result of removing irrelevant features from the data set depending on the variable that is planned to be estimated or classified by feature selection. After feature selection, in addition to achieving the mentioned objectives, especially in complex data sets, the management of the data set becomes

easier, the relationships between the features can be easily defined and the intelligibility of the models increases. There are various algorithms and methods for feature selection, which are mentioned below [119].

The methods used in feature selection; filter methods, wrapper methods and embedded methods can be divided into 3 main groups . In the filter method, the importance of the features is calculated by considering the connection between the target variable and the features, and the features are filtered by considering this importance. As a result, a new feature group is obtained as a result of the filtering process. While the positive side of this method is that it can operate independently of the classification algorithms, it is often encountered that it is less successful when compared to other methods. In the wrapper method, sub-attribute groups are determined and classification is performed with the algorithm to be used. The sub-feature selection and classification process continues until the highest accuracy rate is obtained, and the features with the highest accuracy rate are decided on. In this method, easy application and high success rate stand out as positive sides, while the high transaction cost constitutes the negative side of the method in question. In the embedded method, the feature selection process is carried out by using the features provided by some machine learning algorithms such as random forest during the model creation phase. In this method, the low transaction cost and high success rate are the positive sides.

On the other hand, it is possible to apply different approaches in addition to the methods described, taking into account the purpose of the study and the available data set. In this thesis, features were examined with an approach in which permutation and column removal methods and random forest algorithm were used together. For this reason, information on these matters is briefly mentioned below. First of all, it should be noted that it is very important to develop a model that can accurately predict the results of the classification process. However, this alone will not be enough because it is also important that the developed model is interpretable. Therefore, choosing among the features and examining the importance levels of the features for this purpose constitute an important part of the studies in the field of machine learning. At this point, information about the importance degrees of the features can be provided in applications related to the random forest algorithm. Therefore, the permutation method or column removal method, which can be applied in many models, and the analysis of the importance of the features can be easily applied with the random forest algorithm. Essentially, information is provided on the assumed severity levels in the random

forest algorithm. However, it is known that there is a tendency to inflate the importance levels of continuous variables or categorical variables with a high number of elements during the calculation of the importance of the features by the algorithm in question. are available. For this reason, it would be appropriate to use the techniques such as permutation and column removal mentioned above, instead of selecting features by considering the assumed importance levels of the random forest algorithm.

In order to calculate the importance of an attribute with the permutation method, a baseline accuracy rate is obtained by first passing the OOB samples provided by algorithms such as a validation set or random forest through the random forest algorithm. Then, by changing the values of a single attribute, a new accuracy rate is calculated in the same way. The importance of the mentioned feature is found as the difference between the basic accuracy rate and the accuracy rate obtained by replacing the feature values. In the last step, a selection can be made by looking at the relative importance of the features. In the permutation method, the model does not need to be retrained, only the feature values are replaced and new accuracy rates are calculated with the previously trained model. However, this method is still computationally more costly than the default method of the random forest algorithm. However, the accuracy of the results of the permutation method is higher than the default method. Although the mentioned method can be used with any machine learning algorithm, the random forest algorithm is preferred because it provides OOB samples and therefore eliminates the need to create validation sets. In general, the permutation method used to measure the effect of mixing the attribute values on the model accuracy does not require retraining of the determined model, which is a significant gain when the cost of training the model is taken into account, but also causes a potential bias towards the features that have a correlation with each other. At this point, if the computational cost of retraining the model is ignored, it is possible to obtain the most accurate feature importances with the column removal method. In the column removal method, the basic starting point is similar to the permutation method, and in this method, a basic performance value is obtained first. However, a column, or attribute, is then completely removed from the dataset and a new performance value is calculated with the retrained model. The degree of importance of the feature extracted from the data set is found as the difference between the newly obtained performance value and the basic performance value. Therefore, in this model, the focus is on how much an attribute has an impact on the performance of the model in general. Column removal method gives more accurate results

compared to permutation method, but is more costly. The said cost can be reduced to some extent by using subsets obtained from this training dataset instead of the entire training dataset.

While the column removal method can be used with any algorithm similar to the permutation method, it is mostly used with the random forest algorithm because there is no need for an additional validation data set in the random forest algorithm and the OOB score provided by the algorithm can be used directly for performance measurement. On the other hand, each attribute is handled separately in both the permutation method and the column removal method. This should not be a problem if each attribute is independent of each other and there is no correlation between them. However, these methods may yield unexpected results if there are features in the data set with correlations between them. For example, considering the column removal method, the importance of the attribute in question will appear to be quite low, since removing an attribute that is correlated with another attribute will not make a huge change in the baseline performance. Therefore, attributes that are correlated with each other will appear to share importance, and even if they are important, they will appear to be insignificant. In this context, when using both the permutation method and the column removal method, the features that contain correlation should be examined and it is important to determine the features that are very similar in order of importance, even if they differ in value in both methods.

i. Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a statistical and computational technique used to identify hidden elements at the root of a measure or set of signals of random variables. The CIA is defining a model for generating multidimensional observable data. It is usually presented as a large sample database. The model assumes that data variables are a linear mixture of unknown and latent variables and that the mixed system is also unknown. Hidden variables are considered independent, not external variables and as independent components of the observed data. The CIA can find these independent components, also called sources or factors. At first glance, the ICA is involved in the analysis of fundamental components and factors. However, ICA is a much more powerful method and can be used to find the underlying factor or source when these traditional methods completely fail. The data analyzed by the ICA can be obtained from a variety of applications, including digital imagery, documentation databases, economic indicators, and psychometrics. In many cases, measurements are provided as a series of parallel signals or time series. The term blind source

propagation is used to describe this problem. Typical examples are simultaneous voice signals recorded by multiple microphones, brain waves recorded by multiple sensors, radio interference signals from cell phones, or parallel time series in industrial processes. The mathmarical model presented in Eq (3.3)

$$L(W) = \frac{1}{N}\sum_{i}^{M}\sum_{t}^{N} \ln(1 - \tanh(w_i^T x_t)^2) + ln/\det w/ \tag{4.3}$$

ii. PCA

In the case of using high-dimensional data sets with a large number of features during the creation of the models, the number of samples in the data set should also be large in order to discover sufficient permutations regarding the input variables. In order to combat this problem, size reduction techniques that aim for simplicity versus some accuracy rate and the subject of examining the features that are correlated with each other, which is their focus, comes to the fore. Feature extraction, which is one of the dimensionality reduction techniques, provides a new feature space with a smaller size than a high dimensional feature space. By applying dimension reduction techniques such as feature extraction, it is possible to improve the performance of algorithms, reduce computational complexity, reduce the memory required for models, generalize the models more easily, and make them more reliable. Because, as stated, the new feature space created as a result of reducing the size of the data set contains the most relevant features. In addition, it is possible to visualize data sets with lower dimensions and analyze them faster by machine learning algorithms. In this thesis, Principal Component Analysis (PCA), which is one of the frequently used size reduction methods, was preferred, and the information about this method is mentioned below.

With the PCA method, it is aimed to find a lower-dimensional representation of the data set by preserving the information in the original data set as much as possible. For this purpose, PCA collectively explains and summarizes the variance represented by the large number of related variables in the original data set with a smaller number of representative variables called principal components. The amount of information in PCA expresses the variance rate explained for each

principal component, and the variance rate in question takes a value between 0 and 1. The first step in the application of the PCA method is to standardize the continuous variables in order to avoid biasing the dominant variables. As the second step, the covariance matrix is calculated, and it is aimed to understand whether there is unnecessary information representation in the data set by examining whether there is a correlation between this step and the features in the original data set. In the next step, the eigenvalues and eigenvectors of the covariance matrix are calculated to determine the principal components. These principal components are new variables created as linear combinations or mixtures of initial variables. There is no correlation between these principal components, and most of the information in the original dataset is compressed into the first principal component. If the said situation is to be stated geometrically, the principal components represent the aspects of the data that explain the maximum amount of variance, that is, the lines that capture the information of most of the data. Therefore, since the PCA method involves discarding variables with low information, size reduction can be done without much information loss. While the eigenvectors of the covariance matrix give the directions of the axes containing the most variance, that is, the principal components described above, the eigenvalues give the amount of variance carried in each principal component. Principal components can be obtained in order of importance by ordering the eigenvectors from the largest to the smallest according to their eigenvalues. In this context, finally, eigenvectors are selected depending on the desired size reduction amount.

One of the most important issues in the PCA method, the application steps of which are given above, is to decide on the number of fundamental components, that is, sufficient subspace. Because while the feature size is reduced, the information in the original data set should not be lost too much. In order to understand whether this requirement is fulfilled or how much information is lost, first the variance in each principal component is observed according to the variance in the original feature space, and then the variance in the original feature space is compared with the variance in the new subspace. Currently, there is no accepted method for determining the number of principal components to be used in the PCA method. However, one of the commonly used methods in this regard is the Elbow method. The method in question is mainly used to decide the appropriate number of clusters in unsupervised clustering algorithms, but it can also be used to determine the number of principal components to be found in the PCA method. In the Elbow method, the point where the variance ratio starts to decrease is determined in general. In order to make the said

determination, a scree chart is created in which the variance is a function of the number of principal components, and the point where the rate of change starts to decrease, that is, the elbow point, is observed. Therefore, the basic logic of the Elbow method is based on finding the elbow point by examining the graph in question. It is possible to observe the effect of each principal component from the aforementioned scree plot, and the first principal component contains the highest amount of variance. That is, this principal component causes more variability than any principal component.

3. MATERIAL AND METHODS

3.1 MACHINE LEARNING

machine learning; It is the general name of algorithms that allow a particular problem to be modeled with data specific to that problem. The model, which is designed with the data set and the appropriate algorithm, is designed to have the highest performance. [10] The purpose of machine learning is to create a model that is trained with certain inputs and to make predictions for new values by learning from data such as machine learning and human-specific competencies such as analyzing and making inferences. It focuses on teaching an algorithm to progressively improve on a given set of tasks. In a nutshell, machine learning designs a model of how a process works and applies the model to practically create applications that offer iterative development. When a new sample arrives, it processes the information from the past data and determines which path it will follow, and continues its development for each new sample. Machine learning basically functions as learning and self-improvement. Machine learning is a sub-branch of artificial intelligence studies. While all machine learning problems are analyzed as an artificial intelligence problem, not all artificial intelligence problems can be considered as machine learning problems. Machine learning is an applied form of artificial intelligence. It is aimed to learn like humans by enabling the machine to develop the ability to make decisions on its own by methods such as data collection, storage, parsing, and data analysis. All available data must be analyzed to reveal predictable patterns or meaningful data. After that, using statistical analysis, a model is created to make predictions against new data with the ability to make inferences based on historical data [8]. Success and error rates are determined by testing the created model with various metrics, and it is aimed to reduce the error rate and increase the success rate by improving the learning process with each new data [103].

3.1.1 Machine Learning Basic Concepts

In machine learning, certain algorithms are used and methods are followed to create a self-learning dynamic model. The basic terminology that should be known while applying these algorithms and methods is given below.

summarized: Samples: Each piece of data used for machine learning and evaluation is expressed as a sample or observation. Features: Data that represent an observation. In a rental house price estimation problem, there may be features such as the location of the house, the number of rooms, and its age, while in a spam e-mail classification problem, there may be features such as words in the e-mail content and e-mail length. Labels: Categories in observation in a classification problem. Not spam/spam in the email example, or cat/dog/bird etc in the problem of recognizing animal pictures. Classes are examples. Training data: It is the data set consisting of analyzed and pre-prepared samples used for training the machine learning model to be created.

Test data: Success of machine learning model trained with training data It is the test data to be tested in order to increase the rate of error and reduce the error rate. Overfitting: It is the situation where the created model starts to memorize on the data set. If the training set is created uniformly and the label weights are not correctly distributed, the probability of over-learning of the model will increase. If high scores are obtained in the training set and low scores are obtained in the test data set, it means that the model has memorized the values in the training set. This usually happens when there are too many variables in the model relative to the number of observations. Underfitting: It is the opposite of excessive learning. If learning is insufficient in a model, it means that the model does not progress in accordance with the training data and cannot generalize the new data. Models with incomplete learning problems have a high error rate in both training and test data sets.

Machine learning is progressing in a certain workflow. In this workflow, data collection comes first. The more examples for the model to be created in machine learning, the higher the success rate. It is therefore important to collect as many samples as possible from different sources. The collected data will go through the preparation process in the second step, which is to create meaningful data very important to your name. The data collected from different sources completes the preliminary preparation process by going through steps such as sorting, normalizing, filling in the missing data with various methods in order to serve the same purpose and to make sense of the algorithms to be used. The normalized clean data obtained in the next stage should be divided into two as training and test data with various methods to be used in model training and testing. In some studies, validation data is also needed for validation studies, depending on the type of problem. In such cases, the data set is divided into three at certain rates and training, validation and test data

sets are obtained. The split data is primarily used to train the model with the training data. Afterwards, the model is tested with test and validation data sets and controlled with evaluation metrics. By ensuring that the created model is used, it is ensured that the model is constantly improved with new data collected over time.

Figure 3.1: Machine learning procedures [104]

3.1.2 Types Of Learning

i. Supervised Learning

Supervised learning is a type of machine learning in which a model is trained on labeled data. In this type of learning, it is known what the data set is and what the desired output should be from this data. In supervised learning, the data set consists of both inputs and outputs, and the purpose is to provide the algorithm to create a model by giving outputs along with the inputs to the algorithm. It is requested that the algorithm learns from the existing data and creates a function for itself and makes predictions by subjecting the new data to this function. The algorithm that learns patterns from all the features and outputs in the training data set and reveals a function starts to make predictions and stops when it reaches an acceptable level of success. Supervised learning is

a frequently used type of machine learning, and depending on the output of the problems, it is grouped into regression and classification problems [9]. Classification is to determine which class a data belongs to according to its properties is to be determined. In the classification algorithms, patterns are obtained from the existing data and it is predicted in which class the new future data will take place. The obtained data can be labeled and classification algorithms can be used in the form of a data set that can be divided into categories. The most common classification algorithms are; decision trees, Naive Bayes classification algorithm, k nearest neighbor (K-NN) and an example is support vector machines (SVN). Regression tries to determine the strength of the relationship between a dependent variable and independent variables and uses a statistical analysis that makes predictions based on this power is the measurement. Regression algorithms are used to predict a continuous value. Input values are tried to be expressed with a continuous function. For second-hand vehicle price estimation, the vehicle's make, model, km, year, etc. in the vehicle data set. Obtaining a function according to its properties (inputs) and obtaining a price output as a result of estimation is an example of regression. Linear regression, multiple linear regression, polynomial linear regression, decision tree regression, Random Forest regression are examples of regression algorithms.

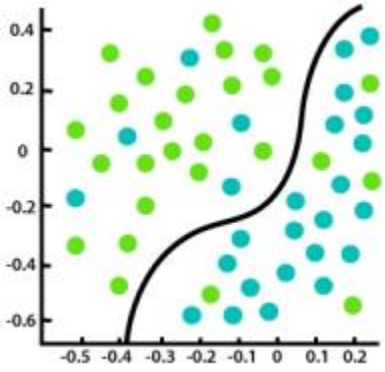


Figure 3.2: Classification [105]

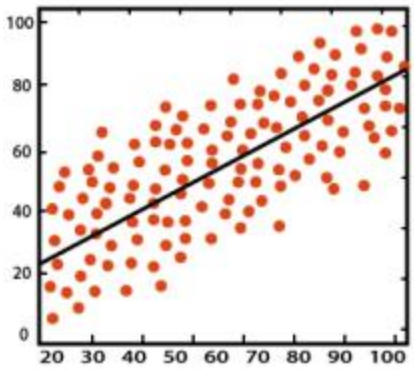


Figure 3.3: Regression [106]

ii. Unsupervised Learning

In the unsupervised learning system, there is no labeled data in the training data, the system works without a teacher. It is a type of machine learning that works to transform unlabeled raw data into organized data. The goal in unsupervised learning is to collect a lot of information about the data and model the distribution of this information on the data. Unsupervised learning algorithms are useful when we do not know what we are looking for in the data. A few unsupervised learning applications in our lives: Music/series/movie/video recommendation systems such as Spotify, Netflix, YouTube, Facebook friend/group/page recommendation system, E-Commerce sites product recommendation system. [9] Unsupervised learning systems can be grouped as clustering and association:

Clustering is algorithms applied to make natural grouping among data. A grouping can be made according to the distribution of the data, or by specifying a certain number of clusters, the algorithm can be divided into the desired number of groups. K-clustering, hierarchical clustering and probabilistic clustering are among the clustering types.

Association Rule is learning association rule between data such as people who buy product A tend to buy product B as well. It is a type of unsupervised learning that is operated to increase sales, especially in markets.

Figure 3.4: Clustering [107]

iii. Semi-Supervised Learning

Semi-supervised learning is a type of machine learning between supervised and unsupervised learning. It can also be said that it is a form of supervised learning that enables the use of unlabeled data in education. In this type of learning, the number of unlabeled data is usually much higher than the labeled data. It has been observed that unlabeled data can significantly improve learning accuracy when used with small amounts of labeled data [108]. Labeling data in a learning problem often requires an expert or a physical experiment. Semi-supervised learning may be preferred when the cost of a fully labeled dataset may be higher than a partially labeled dataset. Self-training is a frequently used semi-supervised learning technique. The classifier used in this technique is first trained on a small labeled data set. Then, classification is attempted on an unlabeled data set, and the most secure labeled points are added to the training set with their predicted labels. These processes are repeated and the model continues to be trained.

iv. Reinforcement Learning

Reinforced learning shows how a system that perceives its environment and can make decisions on its own will learn to make the right decisions in order to realize its purpose . In reinforcement learning, there is an instructor as in supervised learning, but it cannot give much detail. Instead of giving details, when the learning system makes a decision, if that decision is the right one, it rewards the system and guides the system by punishing a wrong decision. The purpose here is to check whether the possible situations that the learning system tries are the target it wants to achieve and to ensure that all tried true or false situations are remembered. Reinforced learning is often used with fields such as robotics, game programming, factory automation. Q learning is one of the most commonly used reinforcement learning algorithms. In a random environment, the system is shown how it can learn the optimal policy. It is difficult for the system to learn the optimal policy

directly because there is no training data available. The system's available training data are only instant rewards. Using this training data, it is much easier to learn a numerical evaluation function and then determine the most appropriate policy with the help of this function. The purpose of the Q learning algorithm is to examine the next movements, see the reward according to the movements that will take place, and act in a way that increases this reward. In the algorithm, in a sense, the agent is expected to plan for the future. This algorithm's maze, search, etc. It is frequently used in problems [109].

3.2 DEEP LEARNING

Deep learning is a machine learning version which uses many layers of nonlinear processing units to extract and transform functions. Each subsequent level accepts the output of the previous level as input [32]. Algorithms can be controlled (eg classification) or uncontrolled (eg analysis of samples). Deep learning has a structure based on learning more than one level of function or data representation. By referring to a view for an image; You can consider a vector of intensity values per pixel or features like edge groups, custom shapes. Some of these properties better represent the data. At this point, deep learning methods use efficient algorithms to extract hierarchical functions that represent data better than manual functions. Figure 3.2.

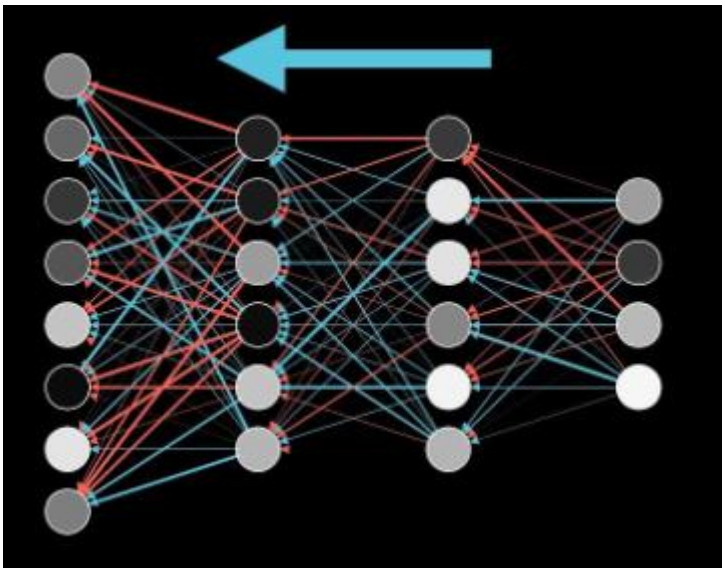


Figure 3.5: Backpropagation [35].

Ivankhnenko's first "Neocognitron" deep learning architecture was proposed by Fusui in 1979. An inexpensive design was introduced, developed, developed and developed in an institution inspired by the visual nervous system of vertebrates. Fucus nets contain several curved pool levels, modern nets. The most important learning gap is the propagation of mistakes at different levels, deep levels, deep levels. Although backpropagation algorithms have been introduced in recent years, the first successful application for deep neural networks was by Yann LeCune et al. on items in a mailbox. Although the network worked successfully, the application turned out to be insufficient as the training took 3 times 3 times. After this work, Ian LeCann reimplemented convolutional networks and network backpropagation for handwritten digit classification (MNIST) using "Lenet" [37]. The advantages of ANN algorithms in this period - because of the computational burden. years. expand | The computational speed has increased almost 1000 times over a 10-year period. During this time, ANN began to compete again with support vector machines [38]. The term "deep learning" was first introduced in 2000 in relation to children and adolescents. In a 2006 paper by Jeffrey Hinton, he showed how a direct-acting multilayer neural network can efficiently train one layer (a layer with a Boltzmann machine layer) and also improve it using a controlled back propagation method. As GPU speed increases, Ciresan et al. | Krishevsky, Suckever, and Hinton developed similar ones in 2012. In their GPU-based work, a normalization technique called "dropout" 1 was used to reduce overfitting 1. ILSVRC 2012. One of the classes of machine learning (artificial learning), deep is learning. Deep learning, machine learning; machine learning can be expressed as a sub-branch of artificial intelligence. In this learning, many layers of nonlinear processing units are used for feature extraction and transformation. one after the other, that is, each The successive layer takes the output of the previous layer as input (Deng & Yu, 2014). Deep learning, multiple features of data it is a construct based on learning the level or representation of Lower a hierarchical representation is created by deriving the higher-level features from the features found at the levels. of abstraction this representation learns multiple representation levels corresponding to different levels. It is a learning based learning. How much for the learning process if there is data entry then much success is achieved. Multiple data passes through the layer. The upper layers reveal more detail. are layers. Deep learning can be performed supervised, semi-supervised or unsupervised. Image and natural language processing for weapons and security systems carried out in ASELSAN's R&D center activities in the fields of deep learning can be given as examples of deep learning studies in Turkey. In the OttOCR project, the Ottoman

character identification system and the OpenZeka project image and video identification deep learning APIs are available for Multilayer Perceptron (Multilayer Perceptrons), Convolutional Neural Network (Convolutional Neural Networks), Recurrent Neural Networks There are three main deep learning models. Deep learning is a type of ANN with many (2 or more) hidden layers and a field of study of machine learning. By increasing the number of layers in ANN architectures, the interpretation and extraction of nonlinear information has improved. Deep learning architectures. The only difference from basic ANN architectures is that it contains more neurons and layers. The most important features of deep learning architectures are that they can learn the features in the data by themselves without the need of an expert, and they can make feature extraction according to the related problem. Although ANN architectures were introduced in the 1950s, deep-layer ANNs architectures could not be used for a certain period of time due to limitations such as lack of sufficient data, insufficient computing power in computer systems, vanishing gradient, overfitting problems. Today, many Ease of access to data, advanced computing power with graphics processing units (Graphics Processing Unit; GPU) and developments in hardware have enabled deep learning to be used in the field of artificial intelligence. Deep learning architectures can be trained as supervised and unsupervised as in ANN architectures. In supervised learning, labeled data it is possible to learn the weight values that minimize the error in the target value estimation of the classification or regression problems. Labeled data is not used in unsupervised learning. Unsupervised learning is generally used for clustering, feature extraction and dimensionality reduction.

3.2.1 Deep Learning Techniques

3.2.1.1 Convolutional neural network

Convolutional Neural Networks (CNN) A type of multilayer perceptron (MLP). The cell in the visual center is split to cover the entire image, with single cells with border elements and complex cells with larger receptors that are supposed to focus the entire image. CNN's algorithm, a positive neural network, is inspired by the visual center of the animal. Here, mathematical convolution can be considered as the response of the neuron to a stimulus from the stimulus field. A CNN consists of one or more convolutional layers, a downsampling layer. The best (latest) results have been obtained, particularly in the field of image processing. In a study of the MNIST dataset, Chireshan was able to use CNN to reduce the error rate to 2%. In another study, Cireşanetal [41]. In the

ImageNet Competition in 2014, all of the teams that received the best scores in object classification and detection with millions of images and hundreds of object classes used modifications of CNN algorithms. In a 2015 study, CNN demonstrated success in capturing faces in wide-angle ranges, including reverse faces. This network is trained on a database containing 200,000 images containing faces and another 20 million images without faces at various angles and orientations. CNN algorithms have also been used in drug discovery. AtomNet, developed by Atomwise in 2015, was the first deep neural network developed for drug design. Trained with 3D representations of chemical reactions, the system has been used to discover new biomolecules in diseases such as Ebola and sclerosis. CNN was also used for the Go game, and a pre-trained 12-layer CNN model defeated the GNU Go algorithm, which was developed with traditional methods, in 97% of the games. CNN-based AlphaGo, developed by Google DeepMind, was the first program that could beat a professional player (human). CNN is a general artificial neural network (ANN) model that consists of an input and an output layer combined with multiple hidden layers. The input data presented to the model is taken by a neuron and subjected to some functions in the layers to reach the output data. Here, X represents the input data vector and Y represents the output vector. W is the weight vector of the connection between neurons of two adjacent layers. The weight vector is used to represent the strength of the link. Classification is also performed with this weight vector. It is the most used model in classification processes due to its ability to classify data in line with the contextual information obtained with the help of the CNN weighted vector. The preferred ESA model is prepared with AlexNet architecture. AlexNet architecture consists of successive convolution and pooling layers. Rectified Linear Unit (ReLu) activation function is used as the activation function in the convolution layer. Regional weight vectors were obtained by scanning the visual data presented to each convolution layer using a 3x3 filter. In this way, the most important regions for classification are determined by weight vectors. The layer output data obtained with the help of weight vectors in the convolution layer is processed in the next layer, the pooling layer. With the help of a 2×2 filter, it stores the most important data as a result of the interaction of the data in the filter with the weight vector. It is used to reduce computational complexity because it hides important regions. In this process, the loss of some information is inevitable, as the filter prepared to process the input data keeps the most important data in it. CNN consist from several functions as shown below:

Convolutional layers : The convolutional layer is the main building block of CNN, and this is where most of the computation takes place. The input requires several components such as a filter and a map. The filter is then applied to the region of the image and the dot product is calculated between the input pixels and the filter. This scalar product is then passed to the output array. The filter then moves one more step and repeats the process until the kernel has scanned the entire image. The end result of a family of entry and filter points is known as a feature map, activation map, or collapsed feature. It is the basic layer of convolutional neural network. The input image presented to this layer is treated as a matrix. In the layer, the features of the image are obtained using a filter. Generally, filters can be selected in 3×3 or 5×5 dimensions. As a result, it reveals the feature map of the image. In this study, the convolution layer can be expressed as in Figure 3.3. Some neurons in successive layers in the convolutional layer are connected to each other through related fields. This area of interest is linked to a specific region in the previous layer by the weight vector, which means that the weight vector has the same value in all neurons within the specified region. Through this similarity, similar features in different parts can be detected. The regional features of the input image are obtained in the convolution layer. A convolution layer in the ESA architecture performs feature extraction using a combination of linear and nonlinear (convolution and activation) operations. With the convolution layer, new images called feature maps are generated. The feature map highlights the unique features of the original image. The convolutional layer is the convolution that creates the feature map of the images instead of the connection weights and weight sum used by the neural network layers.

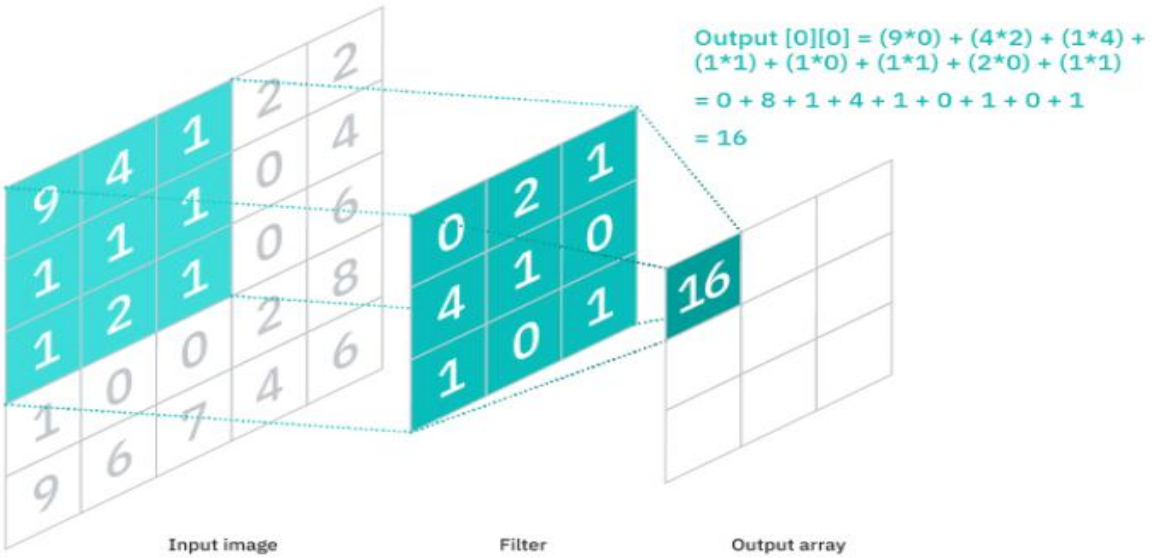


Figure 3.6: Convolutional layer operation [42].

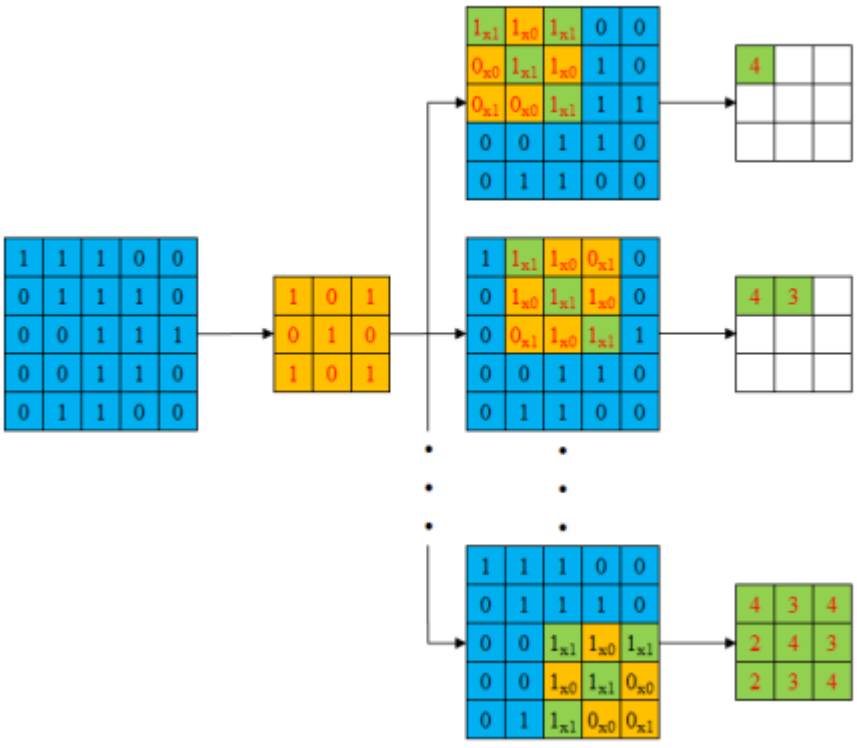


Figure 3.7: Convolutional layer [42].

uses filters. The convolution layer generates the same number of image feature maps as the number of convolution filters used. For example, as shown in Figure 3.19, four different 5*5 size

convolution filters were used and four different feature maps were obtained from an image. While the filter sizes in the convolution layers are generally 5*5 or 3*3 matrix sizes, two-dimensional matrices, 1*1 matrix size has also been used in recent applications. The matrix values of the convolution filters are determined during the training process of the network and the values are updated to the appropriate values during the training process.

3.2.1.2 Parameter sharing

The parameter decomposition system is applied to manage the several of parameters at the convolutional level. Using the example above, you can see that the first coexistence layer contains 55 * 55 * 96 = 290,400 neurons, each with 11 * 11 * 3 = 363 weights and 1 slope. In total, this only makes 290400 * 364 = 105705600 parameters on the first level of ConvNet. There are obviously a lot of them.

It turns out that with reasonable assumptions the number of parameters can be significantly reduced. If the function is useful for calculations in one spatial location (x, y), it is useful in a different location (x2, y2). That is, define a 2D depth segment as the depth segment to constrain neurons to each depth segment (for example, a volume [55 x 55 x 96] has 96 depth segments [55 x 55]). Use the same weights and offsets.[43]. The parameter sharing shown in Figure 3.4.

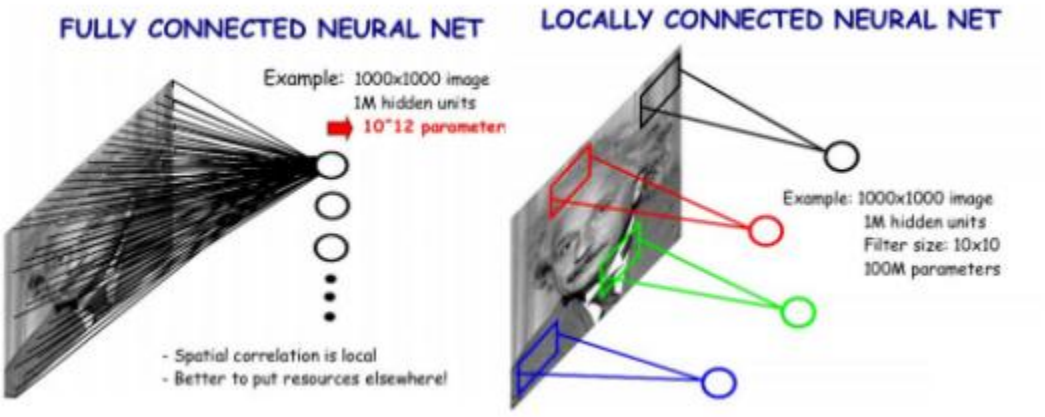


Figure 3.8: Parameter sharing [42].

3.2.1.3 Fully connected layers

Lastly, after applying multiple convolutional and max pooling and at the end of convolutional and pooling layers, the high-level features in the neural network are finished by fully connected layers.

Where in fully connected layer each single neuron is connected to each single neuron from the preceding layer. There are no layers after a fully connected layer [44]. Also, not anymore, they have spatial located (it can be visualizing as one-dimensional. The main goal behind the Fully Connected layer is to utilize their features for identifying the fed date into different classes dependent on dataset for training, for the identification work, a lot of the features from convolutional and pooling layers may be best, likewise may be blends of those features stunningly better.

$$E(W) = \left[\sum_{i=1}^{m}\sum_{r=1}^{N} 1\{y^{(i)} = r\} \, log \frac{\exp(w_r^T x^{(i)})}{\sum_{j=1}^{N} \exp(w_r^T x^{(i)})}\right] \tag{3.2}$$

Two important parameters required in the convolution process are the matrix values of the image and the convolution filters required for the feature map of the image. Convolution filters are shifted on the image in certain steps to create a feature map. The overlapping points of the matrix values of the image and the matrix values of the filter are multiplied and all product values are added. The sum result is determined as the matrix value of the newly created feature map. The matrix values of the feature map are completed by the end of the shifting of the filter on the image. Afterwards, the obtained feature map is processed with a non-linear activation function (ReLU, sigmoid and tanh functions) and the convolution layer output is created.

3.2.1.4 Non-linear Activation Function

Activation functions are really important to learn complex and non-linear mappings between the input and output. CNNs without non-linear activation functions would simply be a linear regression model, unable to model complicated non-linear functions [45]. The main purpose of using CNN is to make sense of something which is complex, high dimensional and non-linear such as videos, audio, images etc. This shows the importance of activation function in CNNs. There are different kinds of non-linearities such as Sigmoid, Tanh-hyperbolic tangent and Rectied linear units (ReLU). Figure. 3.3 shows the plot of all the three functions.

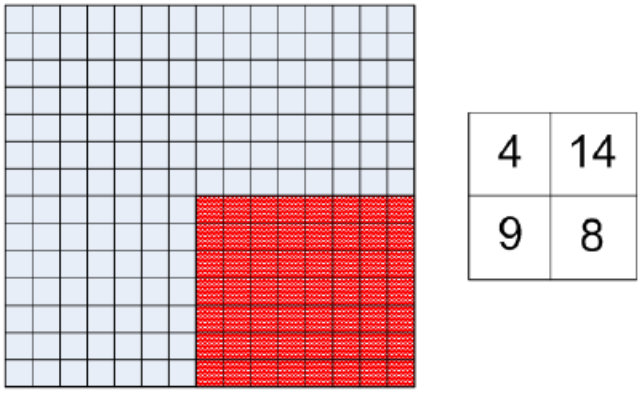


Figure 3.9: Activation function [46]**.**

This function mathematically can represented as shown below:

$$f(x) = max(0, x) \tag{3.3}$$

Why ReLU is important: The goal of ReLU is to give ConvNet non-linearity. I want ConvNet to know the real world data, so there are no negative linear values. The activation layer is the layer that subjects the outputs of the convolution layer to non-linear processing. Linear rectified for this study from the layer with various activation functions.

ReLu (Rectified Linear Unit-ReLu) activation function known as the unit was selected. The ReLu activation function sets negative values to zero instead of dealing with negative values. That is, for values less than zero, the result is zero, while for values greater than zero, it returns its own value as the result. The most preferred activation function in convolutional neural network structure is the sigmoid activation function. However, it has proven that ReLu is better than the sigmoid function in terms of ease of calculation of partial derivatives and preventing gradients from being reversible. In addition, considering the training times, the sigmoid function is slower than the ReLu function. One of the operations in the convolution layer is the convolution operation, while the other is the use of the activation function. The convolution operation is a linear operation, and therefore the outputs are passed through a nonlinear activation function to impart nonlinearity to its outputs. While regular nonlinear functions such as the sigmoid, whose graph is shown in Figure 3.21a, or the hyperbolic tangent, whose graph is shown in Figure 3.21b, were used as the activation function, the nonlinear ReLu function, whose graph is shown in Figure 3.21c, is now commonly used (Yamashita et al. 2018).

3.2.1.5 Stride

Stride is the amount of pixels in the input image. If the step is 1, they change the filters 1 pixel at a time. If the step is 2, we move the filters 2 pixels each, and so on. The picture below shows that a two-step step will work [47].

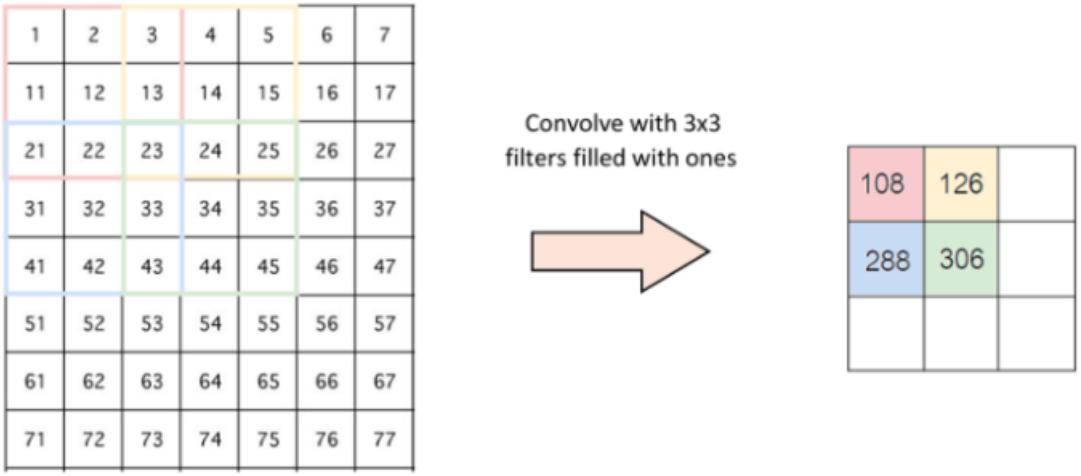


Figure 3.10: Stride [48]

3.2.1.6 Fully Connected Layer

Convolutional neural networks are structured on two bases: classification and feature extraction. While the convolution and pooling layers are active in the feature extraction stage, the softmax activation function layer with one or more fully connected layers is active in the classification process. The fully connected layer structure in the ESA architecture is like the interconnection structure of neurons in a traditional neural network. Each neuron (node) in a fully connected layer is directly connected to each neuron in the next and previous layer with a learnable weight value. The feature maps obtained as two-dimensional matrices from the image data as a result of convolution or pooling layers in the ESA neural network are connected to each neuron in the first layer of the fully connected layer (dense layer) by transforming them into a one-dimensional array of numbers (or vectors) through the smoothing layer . The training phase in the ESA architecture is the same as in other neural network architectures. ESA uses an optimization algorithm to update the weights in its structure and the values in the filters used in the convolution operation. The optimization algorithm takes the classification error (missing value) and propagates the error (differentiated) back to the network to update the filters and weights (Ismael et al. 2020). A fully

connected neural network The number of neurons in the output layer is determined equal to the number of classes in the classification task. Generally, the softmax activation function is used to solve multiple classification problems in the output layer. The Softmax classifier normalizes the real values (pulls the values between 0 and 1) and produces a probability value for each class with a total probability of 1.

3.3 TRANSFER LEARNING

The small amount of data set to be used for training deep neural networks makes it difficult to train the model effectively and may cause problems with memorizing the training data. Transfer learning method is widely used to avoid over-fit problem of deep neural networks (Wang et al. 2020). This method,

It allows the learned weights of models previously trained with large datasets (eg ImageNet, COCO) to be reused in the process of training the target model with a smaller number of target datasets. With transfer learning, the target model trains using weights learned from a trained base network instead of training from scratch. In this way, the features learned by the previously trained network are transferred and reused in another network designed to perform a similar task. Thus, while the training time of the target model will be reduced, it will be trained with a small amount of training data.

The accuracy performance of the models also increases. Transfer learning can be used with two approaches: using a pre-trained model with fixed feature extractor part or using a fine-tuned whole model. In the first approach, pre-trained weights are used directly to extract features from the image, while the last fully connected layer (classifier layer) is modified according to the target task and only this last layer is sloped with the target dataset. Thus, only the classifier layer is fine-tuned while the feature extraction layers of the model remain constant. This method is generally effective when the target dataset is small but similar to the original dataset. In the second approach, the last fully connected layer (classifier layer) is replaced with a new fully connected layer, and the entire network is retrained with the target dataset by back propagation. In this method, the parameters of each layer of the model are updated.

Transfer learning is essential machine learning tool for solving the underlying problem of poor training data. [49].

There are several categories of transfer learning listed below:

- Instances-based deep transfer learning:
  Instance-based portable deep learning is a group of training courses in the target area, limited to the resource area by applying a specific approach to weight correction and applying the appropriate weighting principles to these selected examples. Possibilities to choose as an add-on for.

- Mapping-based deep transfer learning: Mapping-Driven Deep Learning describes how to map source and destination field examples to new functional areas. Examples for both domains in this new data area are and are potentially suitable for corporate DNNs.

- Network-based deep transfer learning: Reconditioning of a restricted network previously trained in the domain of the deep learning provider of the distribution network, including the parameters related to that network configuration, indicates that it will be broadcast as part of the DNN applied to the target domain [50].

3.3.1 GoogleNet

Imagenet is a large image database that has been instrumental in the development of computer vision and deep learning research. The dataset contains approximately 1.2 million labeled images distributed over 1000 classes. Since 2010, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has been held annually, in which deep learning models compete for object detection and image classification based on the ImageNet dataset. Some of the winners
The weights of ESA architectures are recorded and these weights, together with transfer learning methods, enable new tasks to be performed quickly and with less data (Russakovsky et al. 2015).
A major improvement in GoogleNet uses a framework called Inception [51]. Inception is typically a network architectured network, and the optimal low-density LAN architecture repeats itself spatially from start to finish. Three inception architectures are presented for different environments. Inception usually calculates the layers in front of the expensive 3x3 and 5x5 turns with 1 × 1 turns. The structure of GoogleNet represented in Figure 3.5.

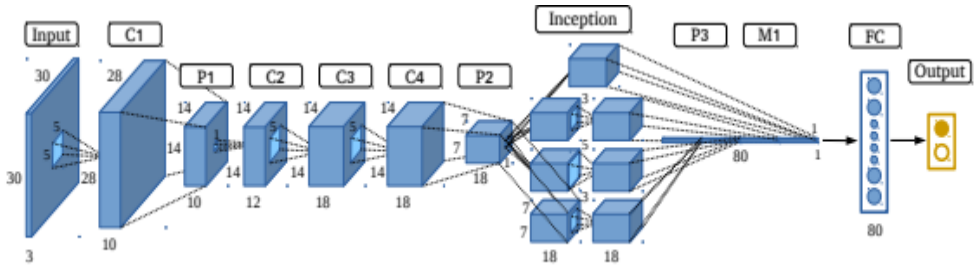


Figure 3.11: GoogleNet structure [46]

Furthermore, the structure of the GoogleNet presented in detail dorm in Table 3.1.

Table 3.1: GoogleNet structure.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

3.3.2 ResNet101

The thread depth has increased, but the accuracy will not necessarily improve. However, there are some near-clean issues in ResNet. Greater depth, which requires a change in weight at the edges of the net, slightly improves the estimate of the stock layer [52]. Add-ons are required parameters. The structure of ResNet101 presented in Figure 3.8.

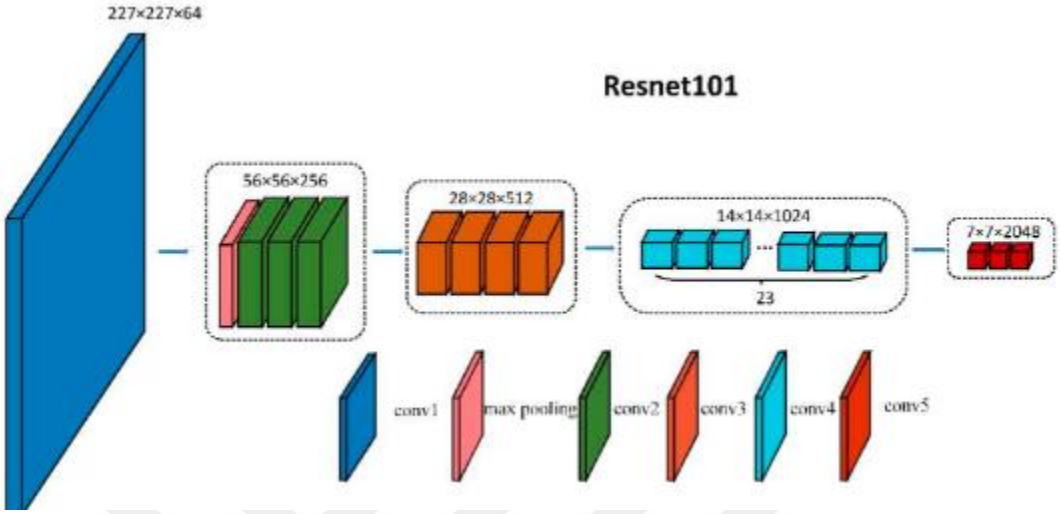


Figure 3.12: ResNet101 structure.

3.3.3 VGG-16

Oxford University Simonyan and Zisserman are developing sixteen floors, sixteen of which three are fully connected at the folding level. A CNN using only a path with a 3 × 3 filter and pad, and the path is up to VGG-16 with a 2 × 2 filter of 2. To reduce the number of parameters in the CNN, a small filter used 3x3 is applied to all convolutional layers with a filter with an error rate of 7.3%. VGG-19 is capable of recognizing various images, classifying videos, finding objects, etc. The design of VGG-16 presented in Figure 3.7.

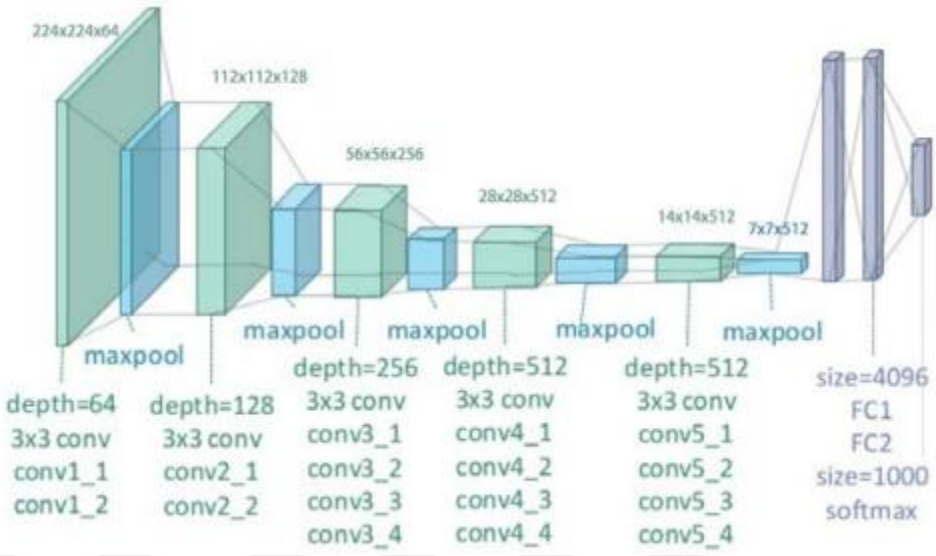


Figure 3.13: VGG16 structure

3.3.3 VGG-19

The deep features obtained with the Vgg-19 deep learning model were used in classification methods such as SVM, kNN, RO, KA and NB, and the results were compared. As seen in Figure 28, RO method showed the highest performance (87.778%), while CA showed the lowest performance (72.222%). It is seen that NB and DVM methods have the same performance (85.926%).

3.4 HYPER PARAMETERS

In many studies, it is stated that two important factors affecting the performance of machine learning algorithms are the value of hyperparameters and selected features. The issue of determining the features has been mentioned in section 2.5, and the determination of hyperparameters will be discussed in this section. In machine learning algorithms, the hyper parameter is a value that must be decided by the person who created the relevant model, and its value may vary depending on the problem to be solved and the data set being used. For example, the k value in the k-nearest neighbor algorithm is a hyperparameter. The success rate of the mostly created models may also depend on the hyperparameters. Therefore, it is necessary to select the most appropriate hyperparameters in order to increase the success rate and ultimately make the model fit for purpose [110]. The selection of hyperparameters may generally depend on the

intuition and experience of the people creating the model and the influence of previous applications. However, recently, various techniques such as heuristic search, grid search and Bayesian method have been used to determine the optimum value of hyperparameters. In the heuristic search method, the model is created with the hyper-parameters determined by using previously owned information and the results are observed. Afterwards, new hyper-parameter estimates, which are considered to increase the success rate of the model, are made and these processes are continued until the expected result is obtained. The grid search method, on the other hand, is a brute force application based on the determination of the best performers from a set of predefined parameters (Adamsson, 2017). That is, to determine the parameters, all possible variations of the model are evaluated within the framework of predefined parameters and the most successful variation is selected as the final model. In the application of the said method, it is important to determine the range for the values that hyper parameters can take, especially considering that some hyper-parameters can take infinite values, using the prior knowledge, and the main points are selected from these determined intervals. As a result, models are tested with all combinations created using the values of these main points, and the combinations that achieve the best result are decided. In Bayesian method, the results of previous experience are used in the selection of the next hyperparameter and hyperparameters are tried to be determined by probability calculation over the known Bayes theorem [111].

On the other hand, in addition to the methods mentioned above, it is possible to use smarter techniques such as deep learning and genetic algorithm, which are still under development in this field, to determine hyperparameters. However, which method to use may vary depending on the complexity of the model to be developed and the hyperparameters to be determined, and whether the method to be used will be sufficient, and the grid search method has been applied in this thesis. While designing the deep learning model, there are some parameters that the designer must decide on in determining the algorithm used in the model. These parameters to be determined play an important role in the success rate of model training. Generally, it is not clear exactly which hyper-parameters will be used initially; problem, dataset etc. varies depending on factors. There may be more than one hyper parameter group where the model provides high performance. However, the selection of appropriate hyperparameters may vary according to the designer's intuition and experience on the subject, and the application of the problem in different areas, at this stage, the selection of hyperparameters is one of the most important problems to be dealt with in model

design. The most commonly used hyperparameters; data set size, batch size, epoch number, learning rate, determination of weight starting values, optimization algorithm selection, activation function selection. It is frequently observed that the size and diversity of the data set have a high effect on the success rate of deep learning models. The larger the data set, the more the learning will increase, but not only the size but also the diversity in the data set is a feature that increases the performance of the model. As the dataset size increases, there is no such thing as a constant increase in performance. After a while, the increase rates will decrease and the limit point will be determined by considering these rates while the model is being trained. Batch size, each piece size that needs to be determined in order to process multiple entries in the data set in pieces is called batch. Processing all the data in the dataset at the same time, especially in very large datasets, is a very costly process in terms of time and memory. Backward gradient computation on the network with back propagation in each iteration of the deep learning model and the weight values are updated in this way. The more data in this calculation process, the more time the calculation will take. To solve the computation time and memory cost problem; The data set is divided into small groups and the learning process is performed on these selected small groups. Since the specified batch size must fit in the GPU memory, the batch size should be determined in the form of 2,4 ,8, 16, 32.. in multiples of two. If it is not determined in this way, there may be sudden drops in performance. Epoch is a term for iterating a model using the entire training set to update the model weights as part of the training. While the model is being trained, it is not trained as a whole, but in parts that we mentioned as batches. During the first training, the performance of the model is tested and the weights are updated with back propagation. It is retrained with the new part and the weights are updated again. This iteration is repeated at each training step to calculate the optimum weight values for the model in this way. Each of these training steps is called an epoch. Since the most appropriate weight values are calculated in each step during the training, the success rate will be low in the first epochs, but the success rate will increase as the number of epochs increases. The size of the epoch number varies depending on the problem. Since the performance will start to increase very little after a certain number of epochs, the training is terminated at these low increase points. Learning rate is an important parameter used to measure the learning performance of the model in deep learning. Updating parameters in deep learning Batch size, each piece size that needs to be determined in order to process multiple entries in the data set in pieces is called batch. Processing all the data in the dataset at the same time, especially in very large datasets, is a very

costly process in terms of time and memory. Backward gradient computation on the network with back propagation in each iteration of the deep learning model and the weight values are updated in this way. The more data in this calculation process, the more time the calculation will take [120]. To solve the computation time and memory cost problem; The data set is divided into small groups and the learning process is performed on these selected small groups. Since the specified batch size must fit in the GPU memory, the batch size should be determined in the form of 2,4 ,8, 16, 32.. in multiples of two. If it is not determined in this way, there may be sudden drops in performance. Epoch is a term for iterating a model using the entire training set to update the model weights as part of the training. While the model is being trained, it is not trained as a whole, but in parts that we mentioned as batches. During the first training, the performance of the model is tested and the weights are updated with back propagation. It is retrained with the new part and the weights are updated again. This iteration is repeated at each training step to calculate the optimum weight values for the model in this way. Each of these training steps is called an epoch. Since the most appropriate weight values are calculated in each step during the training, the success rate will be low in the first epochs, but the success rate will increase as the number of epochs increases. The size of the epoch number varies depending on the problem. Since the performance will start to increase very little after a certain number of epochs, the training is terminated at these low increase points. Learning rate is an important parameter used to measure the learning performance of the model in deep learning. Updating of parameters in deep learning is provided by back propagation process. This update process is done by finding the difference by using backward derivatives and calculating the new weight value by subtracting the result from the weight values of the result obtained by multiplying the difference value with the learning rate parameter. The learning rate parameter used during this process can also be determined as a fixed value or an incremental value. Generally, large values are chosen for the learning rate at the beginning, but smaller values towards the end. Choosing large learning rates in the beginning will speed up the training steps, while decreasing the learning rate as the target is approached will increase the classification success. Determining the initial weight values of the model affects the learning and speed of the model. Although there are different weight determination methods, definitions can be made with a standard deviation such as 0.5 or a uniform distribution between 0.5 and 0.9. If the weight values of the model are initially initialized to zero

Since the matrix multiplication is a sum in the calculation, the weights should not be given as zero at the beginning, since the inputs will be the same as the output. Optimization methods in deep learning models are important to find the optimum value in solving nonlinear problems. Commonly used optimization algorithms; stochastic gradient descent (SDG), adam, adamax, RMS-Prop, adagrad. Although SGD is the optimization algorithm used by default in deep learning models, it can be said that it works slower than other methods. At the same time, it can give very bad results in some problems, especially image recognition problems. Adaptive algorithms learn the learning rate itself. If we talk about a few optimization algorithms; AdaGrad makes large updates for sparse parameters and smaller updates for frequent parameters, and therefore it can be said that it is more suitable for sparse data such as NLP and image recognition. There is a learning rate for each parameter in Adagrad, and the learning coefficient gets extremely small as the learning rate continues to grow during the update process. Rmsprop, on the other hand, was developed as a solution to exactly this extreme shrinkage problem in Adagrad. Instead of using all the values obtained from the squares of all past slopes in the Adagrad algorithm, it has limited the amount of value to a certain frame size. Another optimization algorithm, Adam, caches the momentum changes as well as the learning rates of each of the parameters. Activation functions, linear in multilayer neural networks are functions used to transform operations into nonlinear operations. In deep learning, a linear structure is formed as a result of matrix operations in hidden layers. Deep learning methods are used to solve nonlinear problems [112].

Since it is more effective in solving the model, the model is transformed into a nonlinear structure with activation functions. Some of the activation functions used can be exemplified as Sigmoid, Tangent, Relu and Leaky Relu. Although sigmoid is the most frequently used activation function, the most useful activation function can be shown as Relu, since the parameters are learned more quickly. If negative values are important in our problem, Leaky Relu will be more useful because it catches negative values that Relu misses. Again, the selection of the activation function varies depending on the problem.

4. CONVOLUTIONAL NEURAL NETWORK BASED ON SUPPORT VECTOR MACHINE AND FACTOR ANALYSIS

In this chapter the proposed method of our research named CNN based on SVM and GA was presented. The proposed method consist from three stages, feature extraction using CNN, genetic algorithm applied as feature selector to reduce the size of output features of the CNN. Then, SVM applied for for classification of the extracted features.

In the first stage, AlexNet is a CNN channel hosted by Alex Krizhevsky. AlexNet was involved in ImageNet's core work on image recognition in 2012. The network was ranked in the top five errors with a 15.3% rate, down 10.8% from its completion rate.

AlexNet is the name of a convolutional network that has had a great impact on machine learning, especially in deep learning computer vision applications. As you know, we won the 2012 ImageNet LSVRC competition by a wide margin (26.2% (second place) with an error rate of 15.3%). Network - Yang LeCun et al. But LeNet was deeper in a convolutional layer that applies multiple filters layer by layer. Consists of 11x11, 5x5.3x3, line, max connection, interrupt, data increase, ReLU trigger, SGD with pulse. Activating Connected ReLU after fully connected layer with all stacking layers. AlexNet trained with two Nvidia Geforce GTX 580 GPUs simultaneously for six days, splitting the network into two channels.

AlexNet has eight levels. The convolutional layer was the first 5 geotags, some of which had the maximum level of clustering, the fc layer shown in the last three, and the activation function was not applied. Shown improved training accuracy for Thane and Sigmoid. The AlexNet shoen in Figure 3.6.

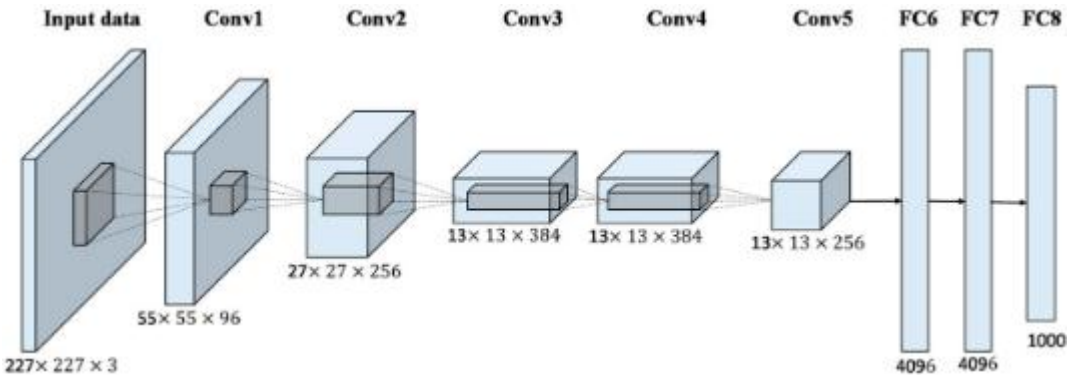


Figure 4.1: AlexNet structure [44]

Then, the Genetic Algorithm, used in the field of Artificial Intelligence, is a kind of algorithm that searches for the best spot. It's about finding a solution to a problem. If we can turn real problems (like containers on dry cargo ships, how to move from point to point, how to create optimal delivery routes, etc.) into research problems, we can solve this problem. Genetic algorithm's theme. A genetic algorithm (GA) is a function that performs permutation-based optimization and looks for a criterion for the likelihood of convergence. It is a research and optimization method that works in the same way as the evolutionary process observed in nature. Genetic algorithms are described in the literature as follows: Genetic algorithms are powerful evolutionary strategies based on the basic principles of biological evolution. The researcher must first correctly identify the type of the variable and the problem with which he is dealing, and then code according to that definition. Then the fitness function is determined, which is one of the inputs to the algorithm, and this is the objective function that needs to be optimized. Genetic operators such as mating and mutation are applied stochastically at many stages of evolution, so it is necessary to determine the probability of their occurrence. Finally, we need to meet the convergence criteria and solve the problem at the lowest cost. GA performs a global search instead of a local one to solve the problem. If the problem is influenced by too many factors, the literature recommends using a genetic algorithm to solve it.. Briefly, the working principle of the Genetic Algorithm is as shown in figure 4.2.

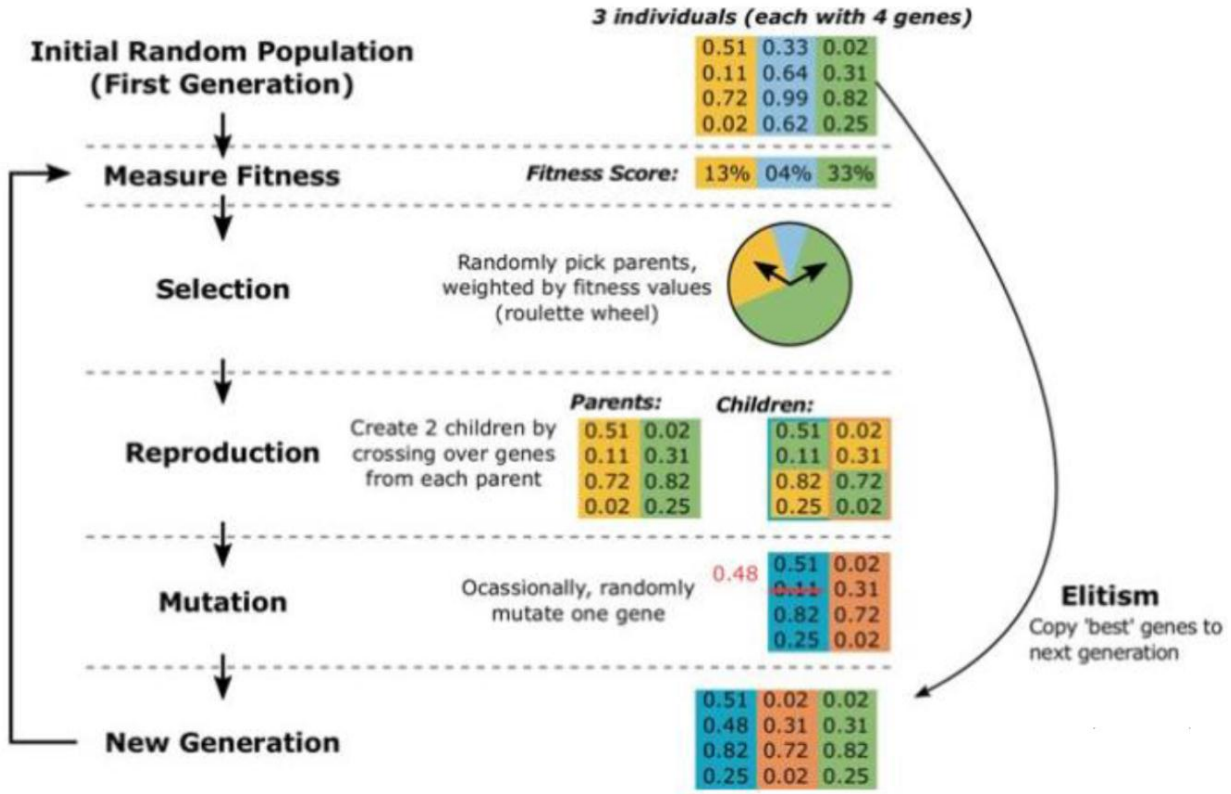


Figure 4.2: GA process [48].

The flowchart of the algorithm presented in the figure 4.3.

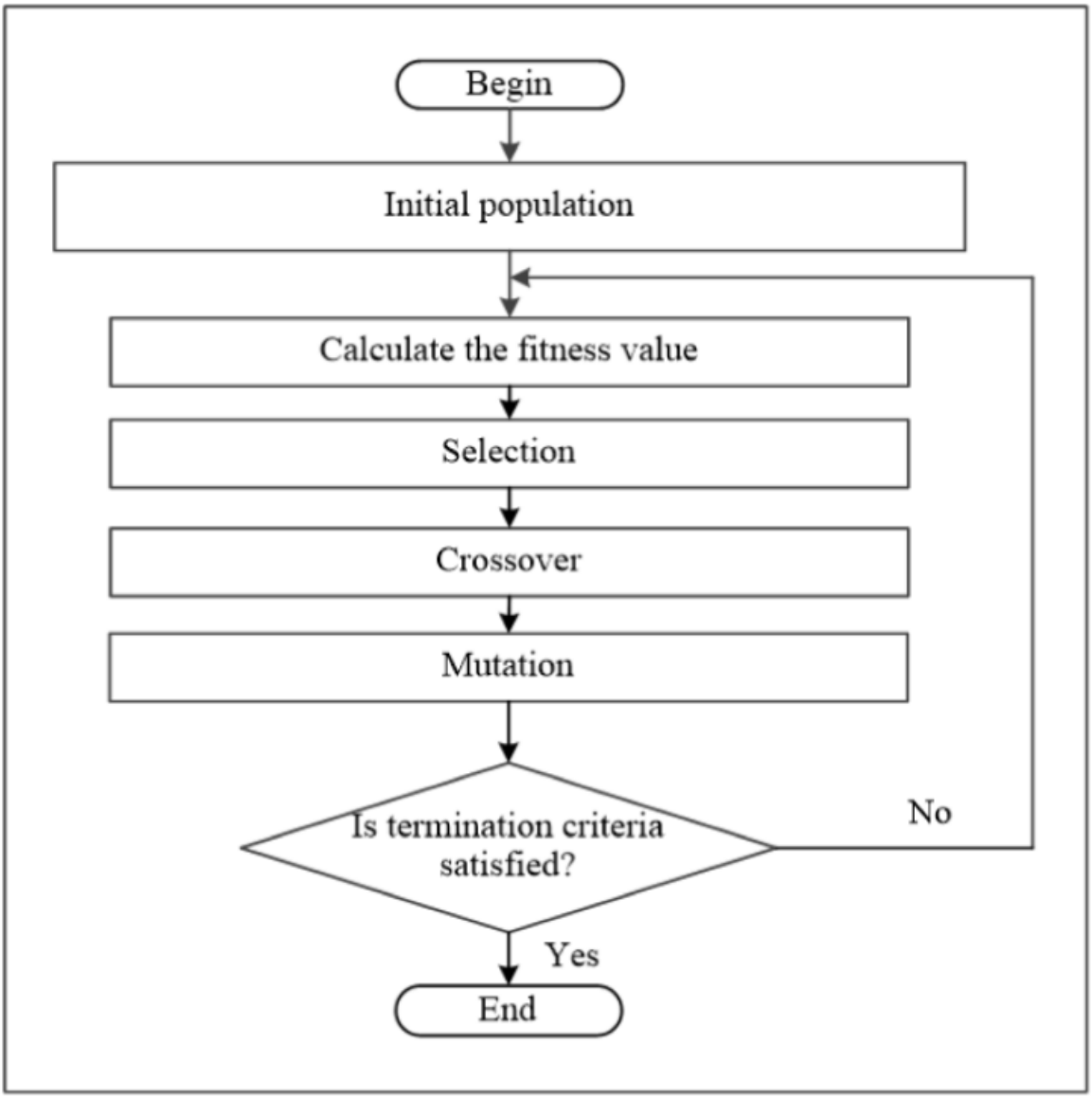


Figure 4.3: GA flowchart [113].

How people understand a problem depends on the problem. The most important factor that determines the success of the genetic algorithm in solving a problem is the representation of the person who offers the solution to the problem. Everyone in the population has a fitness function to decide whether there is a solution to the problem. Individuals who score higher according to the evaluation obtained by the fitness function are capable of producing offspring with other individuals in the population. At the end of the mating process, these humans give birth to new humans called babies. The child has the characteristics of the parents (mother, father) who created it. Individuals with low fitness scores are less likely to be selected when creating new individuals,

so they will be removed from the population after a while. A new population is created by increasing the number of healthy people in the old population. At the same time, this population contains most of the characteristics of well-educated individuals from the previous population. Therefore, good traits are passed down from generation to generation and combined with other good traits as a result of a genetic process. The more individuals with a high fitness value come together and create new individuals, the better a working area is obtained in the search space. In order to find the best solution to the problem; The concepts used in the genetic algorithm are used in a similar sense to the theory of evolution in biology. In natural life, populations are formed by the coexistence of individuals. The population created for the GA algorithm is formed by the gathering of many individuals, in other words, by the gathering of many possible solution candidates. Candidate solutions are kept in sequences coded according to the problem. Each element that makes up this array is called an individual, and each individual represents a specific region in the search space.

In the genetic algorithm, the first starting individuals are usually randomly generated, but this is not a requirement. Especially in very constrained optimization problems, better candidates can be created by paying attention to some of the defined constraints to create the starting individuals. As a result of exposing individuals to the fitness function process, the fitness value, which evaluates how close the solution is to the optimal solution, is determined. The genetic algorithm with the initial population created works with three evolution operators. These; selection, crossover and mutation operators. In general, each of these operators is applied to each individual of the population that will form in the next generation.

Selection is the process of selecting parent individuals to create new individuals based on their fitness values in the population. The crossover operator is applied after the selection process and expresses the mutual replacement of certain parts of the chromosomes of the parent individuals and thus the formation of individuals with new characteristics. The mutation process is the process of changing a gene in any of the chromosomes of the newly formed individual depending on the probability of mutation.

There are various methods to terminate the genetic algorithm process. These methods are; When the desired solution is found during the operation of the algorithm, when the total number of iterations defined at the beginning of the GA is reached, or when the fitness value remains constant,

the solution represented by the best individual found is presented as the most suitable solution found for the problem.

In the last stage, the SVM applied to classify the selected features by the genetic algorithm and the flowchart of the method presented in the Figure 4.4.

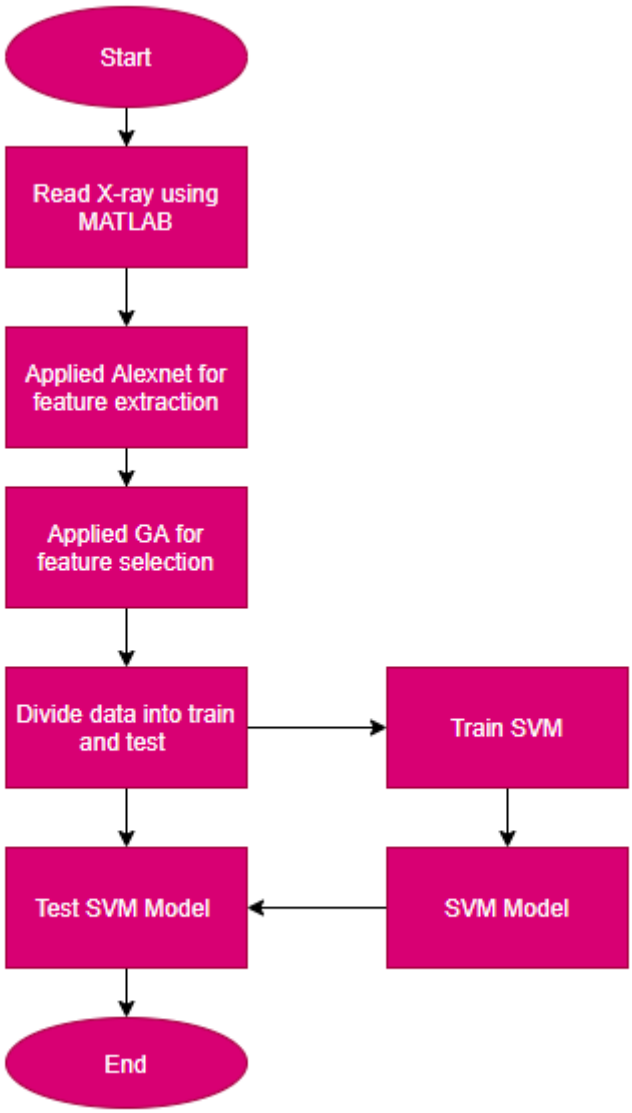


Figure 4.4: Proposed Method flowchart.

5. EXPERIMENTAL RESULTS

5.1 EVOLUTION METRICS

Different performance criteria are used in order to examine the success rate of the models created using classification algorithms, that is, to decide which model can achieve more accurate results. The basis for the calculation of some of these criteria is the creation of an error matrix, and this section includes information about the error matrix and the mentioned performance criteria.

5.1.1 Confusion Matrix

In machine learning, the error matrix, also known as the complexity matrix, is used to visualize how successful the classification algorithms are. In this matrix, there are real situations and situations that reflect the predictions of classification algorithms. Therefore, the error matrix contains the correct and incorrectly predicted values that allow the performance of classification algorithms to be evaluated, and the rows of the said matrix contain the actual class values and the columns the predicted class values. It is also possible to apply the opposite of the values represented by rows and columns. As an example, an error matrix created for the binary classifier is given in table 5.1 below. The said binary error matrix consists of the following values

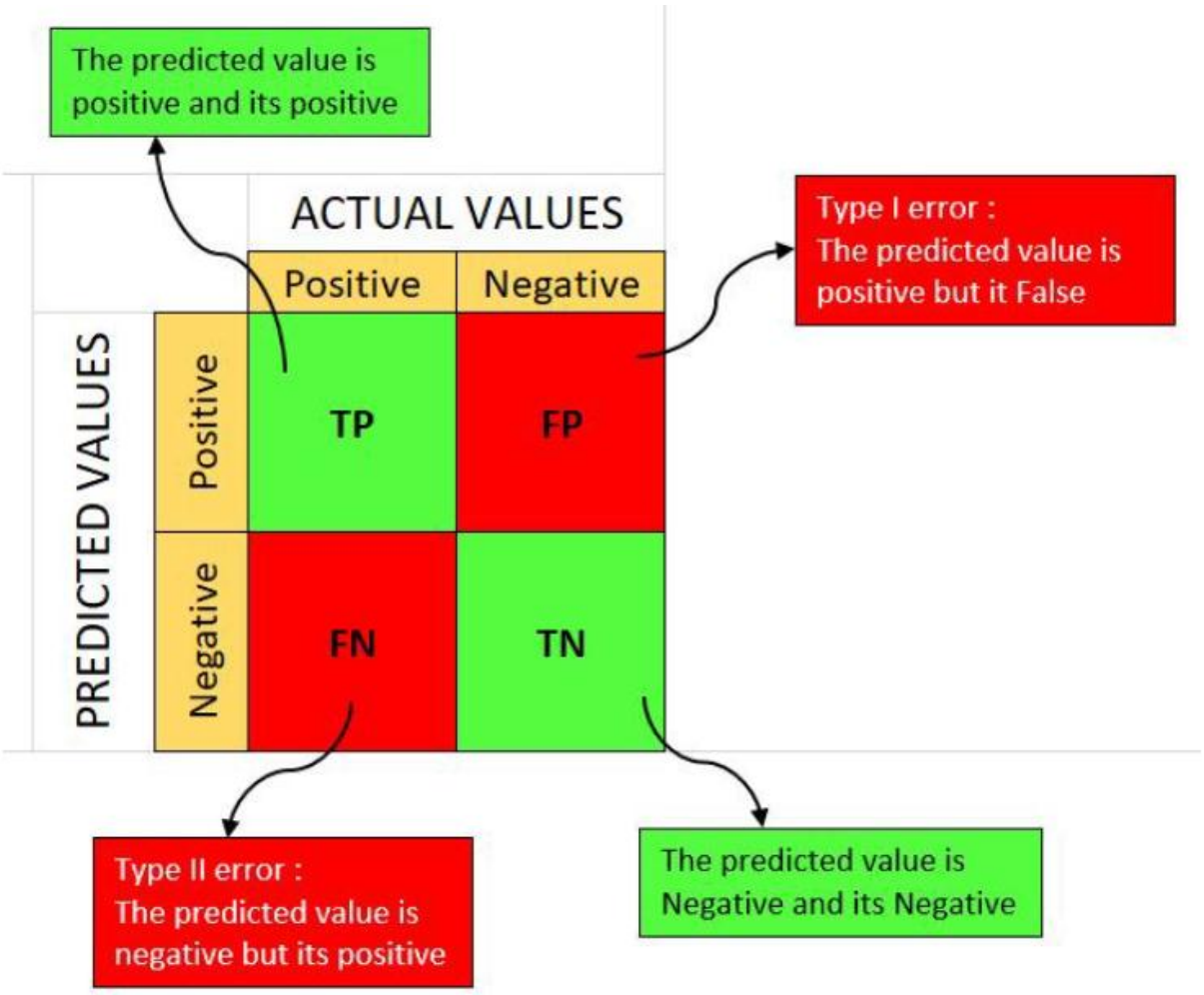


Figure 5.1: Confusion matrix [114]

A Confusion matrix is an N x N matrix used to evaluate the performance of a classification model; where N is the number of target classes. The matrix compares the actual target values with the values predicted by the machine learning model. This gives researchers a holistic view of how well the classification model is performing and what kinds of mistakes it has made. More than one parameter in the Confusion matrix and these together form the Confusion matrix. Below are descriptions of these parameters:

True Positive (TP): The predicted value matches the actual value. The true value was positive and the model predicted a positive value.

True Negative (TN): The predicted value matches the true value. The true value was negative and the model predicted a negative value.

False Positive (FP): The predicted value was incorrectly estimated. The true value was negative but the model predicted a positive value.

Using the above parameters, more than one criterion is calculated and these criteria show the performance of the model. The first of these is the criterion of accuracy, it is calculated by adding the number of correctly classified values and dividing by the total number of values. Accuracy is one of the most important criteria used in classification problems. Therefore, the accuracy value is constantly compared during the comparison.

In the chapter, the results of the proposed method presented and explained using confusion matrix to evaluate the results. The parameters that are calculated in this chapter lead to diagnosis the power and weaknesses of the proposed method.

iii. Accuracy

The accuracy rate is obtained by dividing the correctly classified samples by the total number of samples. While calculating the accuracy rate, all error types are taken into account and all samples are equally important. Therefore, the accuracy rate, whose formula is given in figure 2.9 below, is used to evaluate the overall performance of the algorithm used in the classification process. However, it is effective in cases where the data set is unevenly distributed. it is impossible to say. Namely, in the model developed to predict whether a person is sick or not, if the number of sick people in the sample data set is only a small part of the data set, such as 5%, it will be possible to predict with 90% success that the person is not sick even if a random estimation is made. In such cases, high accuracy will not mean high performance.

Accuracy (ACC) calculated in equation (4.1):

$$\text{Accuracy (Acc)} = \frac{(\text{TP} + \text{TN})}{(\text{P} + \text{N})} \tag{4.1}$$

Also, Sensitivity (TPR) calculates the percentage of correctly recognized positives. Measured Precision is as shown in Equation (4.2):

$$\text{Sensitivity (TPR)} = \frac{TP}{(TP + FN)} \tag{4.2}$$

In addition, Specificity (SPC) calculates the percentage of correctly recognized negatives. Measured Specificity is as shown in Equation (4.3):

$$\text{Specificity (SPC)} = \frac{TN}{(FP + TN)} \tag{4.3}$$

TP indicates true positive, TN indicates true negative, P indicates condition positive, N indicates condition negative, and FPR indicates false positive.

5.2 CROSS VALIDATION

As mentioned in the previous sections, in the application of machine learning algorithms, the data set is divided into two groups as training and testing. Machine learning models are created using the training data set, and the performance of the models is evaluated using the test data set. However, if the performance is evaluated through only one training and test data in this way, it is not possible to determine whether the training and test data sets coincide with good results by chance. In such a case, the implementation of the model may result in unexpected results. At this point, various methods are applied in order to evaluate the performances of machine learning models more effectively, and in this thesis study, k-fold cross validation method was applied for the aforementioned purpose. The Figure 5.1 shown the cross validation method.

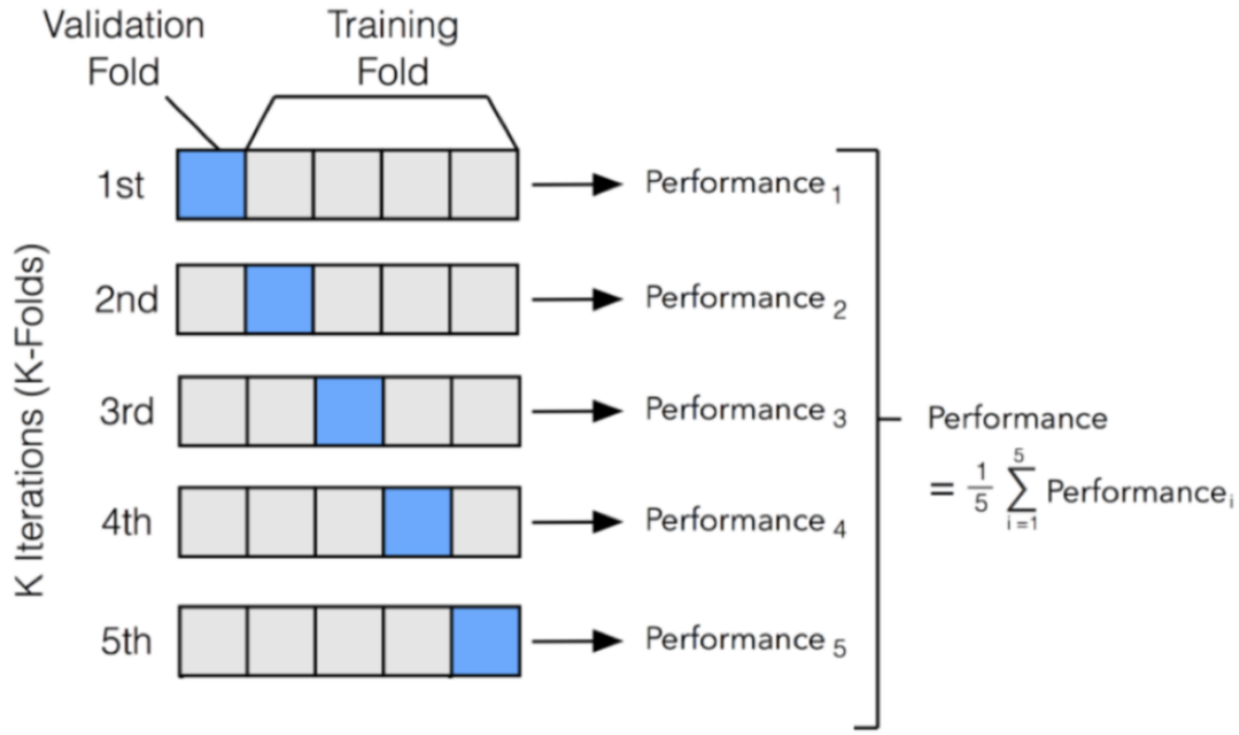


Figure 5.2: Confusion matrix [115]

In the k-fold cross validation method, first the data set is randomly mixed and divided into k equal groups. Then one group is selected as the test dataset and all other groups are used without training the model. The performance of the model created with the training data set in question is measured and stored over the relevant test data set. These processes are repeated for each group and the final performance of the model is determined by averaging the performance criteria obtained in each step. Thus, with the mentioned method, the problem of machine learning models failing in practice while giving successful results in the development environment can be avoided and the created models become more stable. Figure 5.3 below shows the 5-fold cross-validation process as an example [116].

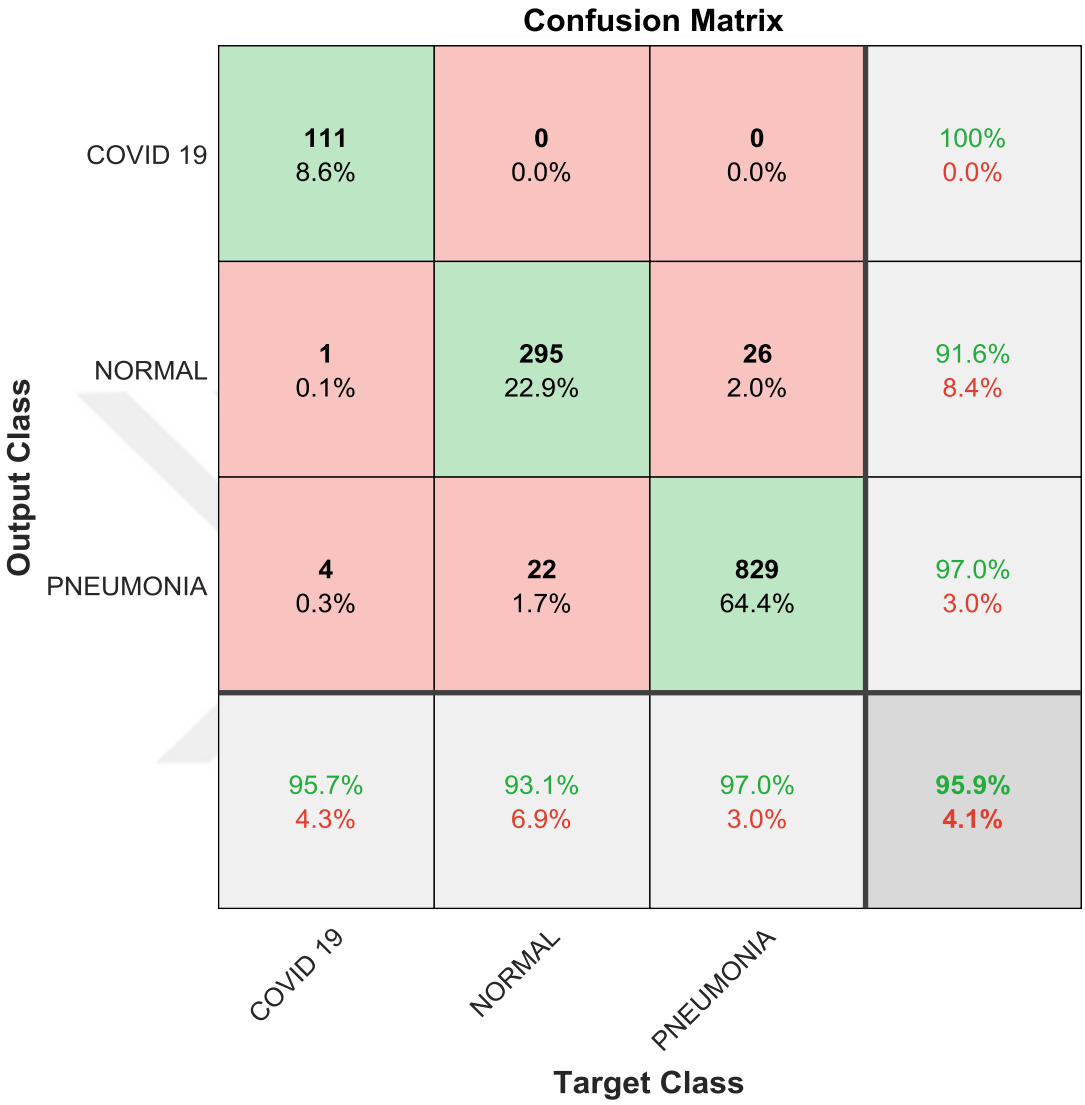
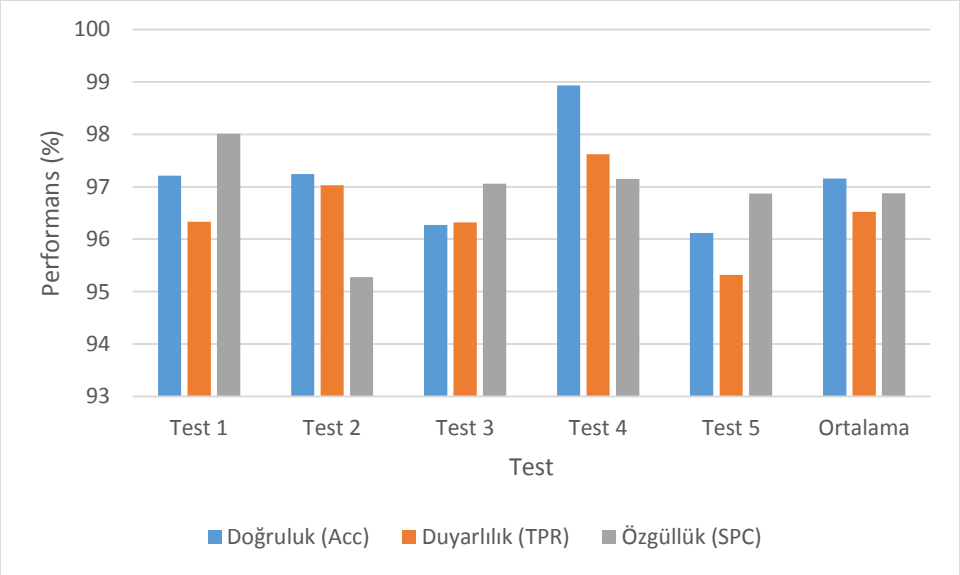


Figure 5.3: Confusion matrix of our method.

The cross validation applied to enhance proposed method performance by calculating the average of 5 experiments. This assist the method to avoid the overfitting problem. The results using cross validation presented in the Figure 5.4.



The presented framework is also compared with previous studies the proposed in related researches and is shown in Table 5.1.

Table 5.1: Comparison with litreature

Ref	Acc (%)
[54]	92.3
[55]	88.90
[56]	94.7
[57]	96.2
[58]	87.02
[59]	93.6
[60]	92.67
Our Method	**95.90**

By analysing of the Table 5.1, the results show that the proposed method produces better results than other major studies. The methods presented use classic methods such as CNNs and genetic algorithms and can show surprising results on a small number of datasets. Deep learning methods, on the other hand, require large amounts of data to perform better than other methods. Also, the complexity of the data model makes training very expensive. In [54] Xin Li et al. dealing with

With the growing demand for millions of COVID-19 tests, computed tomography (CT) -based tests have emerged as a promising alternative to the gold standard for RT-PCR testing. However, it is mostly administered in emergency rooms and hospitals due to the need for expensive equipment and trained radiologists. Adequate, accurate, rapid but inexpensive testing is urgently needed to screen the population in mobile, emergency and primary care settings for COVID-19. Here, we have developed a deep convolutional neural network (CNN) that extracts features from chest x-rays (XCRs) of pneumonia and large area normal mass cases and amplifies them with a small number of COVID-19 cases. Automatically distinguishes COVID-19 cases from pneumonia cases and / or normal XCR images. We demonstrate the great potential of our COVID-Xpert and XCR-based population screening approach to detect COVID-19 cases with impressive experimental results. In [55] Parnian Afshar et al. presents an alternative capsule network-based modeling environment called COVID-CAPS that can handle small datasets of great importance due to the sudden and rapid onset of COVID-19. Our results, based on the X-ray imaging dataset, show that COVID-CAPS has an advantage over previous CNN-based models. COVID-CAPS achieved 95.7% sensitivity, 90% sensitivity, 95.8% specificity and 0.97 area under the curve (AUC) with far fewer parameters than its trainable counterparts. To further enhance the diagnostic capabilities of COVID-CAPS, pre-training will be built on a new dataset from an external X-ray imaging dataset. Pre-training has been further increased with the nature similarity dataset, so the sensitivity is at 98.3%. and the specificity is 98.6%.

6. CONCLUSION

In this section, we have developed a new COVID-19 detection framework that uses an SVM classifier based on a CNN model and a genetic algorithm. The main question in machine learning methods is how to work with large functions in some examples. This results in over-fit of the model or degraded performance during testing. Our goal was to avoid overfitting the feature selection and extraction methods. An important contribution to this research is the combination of genetic algorithms for the extraction and selection of CNN features. This combination consists of extracting the active function from the input data associated with the classifier. Furthermore, the SVM classifier proposes a conservative method to separate two discrete finite groups A and B in an n-dimensional space and concludes that it provides fast and accurate results with less training data. The combination of these methods results in a robust model that provides remarkable results compared to other known studies.

How successful transfer learning is compared to the standard CNN model in a small amount of data has also been observed in this thesis study. In each developed model, solutions to existing problems have been produced, so success rates have been increased. In this thesis study, it was observed that the success rate of model training with transfer learning was high, and results were obtained with AlexNet models in terms of both success rate and time. However, AlexNet was observed as the most successful transfer learning model on this data set.

The feature selection technique used in this study by applying genetic algorithm. The genetic algorithm are number of advantages that are assist this model such as: Parallelism.

- Global optimization.
- A larger set of solution space.
- Requires less information.
- Provides multiple optimal solutions.
- Probabilistic in nature.
- Genetic representations using chromosomes.

These features assist the model to select best features that are extracted by AlexNet which the 4096 features reduced to the 1643 which represented best features that presented optimum accuracy in detection of Covid-19.

In the future works, the researcher advises to apply other optimization algorithms that published in the last years instead of genetic algorithm to obtain new results which can be effective with Covid-19 detection.

REFERENCES

[1] World Health Organization (WHO) - https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed at Feb 24, 2020).

[2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han et al., "Epidemiological and Clinical Characteristics of 99 cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study". *Lancet - 2020*, 10.1016/S0140-6736(20)30211-7.

[3] A Y. Yin, R.G. Wunderink, "MERS, SARS and Other Coronaviruses as Causes of Pneumonia", *Respirology*, 23 (2018), pp. 130-137).

[4] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang et al., "Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan", China. *Jama - 2020*.

[5] Diprose J, MacDonald B, Hosking J, Plimmer B. Designing an API at an appropriate abstraction level for programming social robot applications. J. Vis. Lang. Comput., 2017;39,22–40.

[6] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong et al., "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". *The New England Journal of Medicine*. 2020.

[7] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu et al., "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan", China. *Lancet - 2020*.

[8] General Office of National Health Committee. "of a Program for the Diagnosis and Treatment Notice on the Issuance of Novel Coronavirus (2019-nCoV) Infected Pneumonia (trial sixth edition) (2020-02-18)".

[9] M. Chung, A. Bernheim, X. Mei et al., "CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV).", *Radiology 2020*. DOI: 10.1148/radiol.2020200230.

[10] P. Huang, T. Liu, L. Huang et al., "Use of Chest CT in Combination with Negative RT-PCR Assay for the 2019 Novel Coronavirus but High Clinical Suspicion." *Radiology 2020*.

DOI: 10.1148/radiol.2020200330.

[11] D. Li, D. Wang, J. Dong, N. Wang, H. Huang, H. Xu, C. Xia, "False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases." *Korean J Radiol*. 2020 Apr;21(4):505-508.

[12] National Health Commission of the People's Republic of China. "Diagnosis and Treatment Protocols of Pneumonia Caused by a Novel Coronavirus (trial version 5)." Beijing: National Health Commission of the People's Republic of China; 2020.

[13] H.J. Koo, S: Lim, J. Choe, S.H. Choi, H. Sung, K.H. Do, "Radiographic and CT Features of Viral Pneumonia". *Radiographics*. 2018; 38(3): 719-39.

[14] F. Ucar, D. Korkmaz, "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnosis of the Coronavirus Disease 2019 (COVID-19) from X-ray Images", *Medical Hypotheses*, Volume 140, 2020, 109761, ISSN 0306-9877, https://doi.org/10.1016/j.mehy.2020.109761.

[15] L. Li, L. Qin, J. Xia et al, (2020) "Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT". *Radiology*; Published online 19 March ahead of print. https://doi.org/10.1148/radiol.2020200905.

[16] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: Cnn Architectures, Dataset Characteristics and Transfer Learning." *IEEE Trans. Med. Imaging 2016*.

[17] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz et al., "Deep Learning for Chest Radiograph Diagnosis: A Retrospective Comparison of the CheXNeXt Algorithm to Practicing Radiologists." *PLoS Med. 2018*, 15, e1002686.

[18] Makris, Antonios & Kontopoulos, Ioannis & Tserpes, Konstantinos. (2020). COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks. 10.1101/2020.05.22.20110817.

1. [19] Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., & Singh, V.

(2020). Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos, solitons, and fractals, 138, 109944. https://doi.org/10.1016/j.chaos.2020.109944

[20] Zebin, T., Rezvy, S. COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization. Appl Intell (2020). https://doi.org/10.1007/s10489-020-01867-1.

[21] Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, et al. (2020) New machine learning method for image-based diagnosis of COVID-19. PLOS ONE 15(6): e0235187.

[22] Azemin, M. Z. C., Hassan, R., Tamrin, M. I. M., Ali, M. A. M. (2020). COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings, International Journal of Biomedical Imaging, vol. 2020, Article ID 8828855. https://doi.org/10.1155/2020/8828855.

[23] Rajaraman, S., Siegelman, J., Alderson, P. O., Folio, L. S., Folio, L. R. and Antani, S. K. (2020). Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays, in IEEE Access, vol. 8, pp. 115041-115050, 2020, doi: 10.1109/ACCESS.2020.3003810.

[24] Sahlol, A.T., Yousri, D., Ewees, A.A. et al. COVID-19 image classification using deep features and fractional-order marine predators algorithm. Sci Rep 10, 15364 (2020). https://doi.org/10.1038/s41598-020-71294-2.

[25] Li, Matthew & Arun, Nishaeminth & Gidwani, Mishka & Chang, Ken & Deng, Francis & Little, Brent & Mendoza, Dexter & Lang, Min & Vtc Lee, Optomsl & O'Shea, Aileen & Parakh, Anushri & Singh, Praveer & Kalpathy-Cramer, Jayashree. (2020). Automated Assessment and Tracking of COVID-19 Pulmonary Disease Severity on Chest Radiographs using Convolutional Siamese Neural Networks. Radiology: Artificial Intelligence. 2. e200079. 10.1148/ryai.2020200079.

[26] Lars Larsen, Jonghwa Kim, Path planning of cooperating industrial robots using evolutionary algorithms, Robotics and Computer-Integrated Manufacturing, Volume 67, 2021, 102053, ISSN 0736-5845, https://doi.org/10.1016/j.rcim.2020.102053.

[27] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U. Rajen-dra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792.

[28] A.I. Khan, J.L. Shah, M.M. Bhat, Coronet: A deep neural network fordetection and diagnosis of COVID-19 from chest x-ray images, Comput.Methods Programs Biomed. (2020) 105581.

[29] L. Wang, A. Wong, COVID-Net: A tailored deep convolutional neuralnetwork design for detection of COVID-19 cases from chest radiographyimages, 2020, arXiv preprint arXiv:2003.09871.

[30] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,Phys. Eng. Sci. Med. (2020).

[31] Rousan, L.A., Elobeid, E., Karrar, M. *et al.* Chest x-ray findings and temporal lung changes in patients with COVID-19 pneumonia. *BMC Pulm Med* **20,** 245 (2020). https://doi.org/10.1186/s12890-020-01286-5.

[32] Ahmadzadeh S., Ghanavati M., Branch A., Branch M. Navigation of mobiler robot using the PSO particle swarm. 2012;2(1):32–38.

[33] Islam, M. R., Tajmiruzzaman, M., Muftee, M. M. H., & Hossain, M. S. Autonomous robot path planning using particle swarm optimization in dynamic environment with mobile obstacles & multiple target. In International conference on mechanical, industrial and energy engineering, 2014;1-6.

[34] Deepak, B. Parhi, D.R. 2012. Pso based path planner of an autonomous mobile robot. Cent. Eur. J. Comp. Sci. 2(2):152-168.

[35] Zhang, Y. Gong, D. Zhang, J. 2013. Robot path planning in uncertain environment using multi-objective particle swarm optimization. Neuro computing 103: 172– 185.

[36] Tang, Y. Li, Q. Wang, L. Zhang, C. Yin, Y. 2010. An improved pso for path planning of mobile robots and its parameters discussion, Intelligent Control and Information Processing, Dalian, China.

[37] Wang, L. Liu, Y. Deng, H. Xu, Y. 2006. Obstacle-avoidance path planning for soccer robots using particle swarm optimization, Robotics and Biomimetics, Kunming, China.

[38] Saska, M. Macas, M. Preucil, L. Lhotska, L. 2006. Robot path planning using particle swarm optimization of ferguson splines. Emerging Technologies and Factory Automation.

[39] Nasrollahy, A. Z. Javadi, H. H. S. 2009. Using particle swarm optimization for robot path planning in dynamic environments with moving obstacles and target. Sim European Symposium on Computer Modeling and Simulation.

[40] Solea, R. Cernega, D. 2016. Online path planner for mobile robots using particle swarm optimization, ICSTCC, Sinaia, Romania.

[41] Z. Batık, G. Atalı, D. Karayel, and S. S. Özkan, "Using heuristic distance functions in path planning of mobile robots," Electronics World, pp. 34–36, Oct. 2018.

[42] Z. Batık, D. Karayel, S. S. Özkan, and G. Atalı, "A Comparative Study on Heuristic Distance Functions in Path Planning of Mobile Robots," presented at the ICOME2017, İstanbul, 2017.

[41] Z. Batık, G. Atalı, D. Karayel, and S. S. Özkan, "Using heuristic distance functions in path planning of mobile robots," Electronics World, pp. 34–36, Oct. 2018.

[42] Gao, Z. Zeng, X. Wang, J. Liu, J. 2008. FPGA implementation of adaptive IIR filters with particle swarm optimization algorithm. Communication Systems.

[43] Gigras, Y. Choudhary, K. Gupta, K. Vandana. 2015. A hybrid ACO-PSO technique for path planning. INDIACom, 1616-1621.

[44] Sung-Il Kim, Ji-Young Lee, Young-Kyoun Kim, Jung-Pyo Hong, Y. Hur and Yeon-Hwan Jung, "Optimization for reduction of torque ripple in interior permanent magnet motor by using the Taguchi method," in *IEEE Transactions on Magnetics*, vol. 41, no. 5, pp. 1796-1799, May 2005, doi: 10.1109/TMAG.2005.846478.

[45] M. Sarshenas and Z. H. Firouzeh, "A new wideband closed-form Green's function for microstrip structures using hybrid Taguchi-gradient optimization method," *7'th*

International Symposium on Telecommunications (IST'2014), 2014, pp. 230-233, doi: 10.1109/ISTEL.2014.7000703.

[46] N. Zhu, D.Y. Zhang, W.L. Wang et al., "A Novel Coronavirus from Patients with Pneumonia in China, 2019." *N Engl Med 2019*; 382:727–733.

[47] R.J. McDonald et al. "The Effects of Changes in Utilization and Technological Advancements of Cross-Sectional Imaging on Radiologist Workload". *Acad. Radiol* 22, 1191–1198 (2015).

[48] R. Paul et al., "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma". *Tomography* 2, 388–395 (2016).

[49] J.Z. Cheng et al. "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans". *Sci. Rep* 6, 24454 (2016).

[50] H. Chen, Y. Zheng, J.H. Park, P.A. Heng and S.K.Zhou, "Medical Image Computing and Computer-Assisted Intervention ", *MICCAI 2016* 487–495 (Athens, Greece, 2016).

[51] B. Liang, Y. Zhai, C. Tong, J. Zhao, J. Li, X. He, Q. Ma. "A Deep Automated Skeletal Bone Age Assessment Model via Region-Based Convolutional Neural Network", Future Generation Computer Systems 2019, 98. 10.1016/j.future.2019.01.057.

[52] H.C. Shin et al. "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning". *IEEE Trans. Med. Imag* 35, 1285–1298 (2016).

[53] L. Wang, A.Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images". *ArXiv* 2200309871 2020.

[54] X. Li, D. Zhu, "COVID-Xpert: An AI Powered Population Screening of COVID-19 Cases Using Chest Radiography Images". *ArXiv*:200403042 2020:1–6.

[55] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K.N. Plataniotis, A. Mohammadi et al., "-Caps: A Capsule Network-Based Framework for Identification of Covid- 19 Cases from X-Ray Images". *ArXiv* Prepr ArXiv200402696 2020:1–4.

[56] M. Farooq, A. Hafeez, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs", *ArXiv*:2003.14395, 2020.

[57] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub et al., "Can AI Help in Screening Viral and COVID-19 Pneumonia?" *ArXiv* 200313145 2020.

[58] Z. Batık, G. Atalı, D. Karayel, and S. S. Özkan, "Using heuristic distance functions in path planning of mobile robots," Electronics World, pp. 34–36, Oct. 2018.

[59] Gao, Z. Zeng, X. Wang, J. Liu, J. 2008. FPGA implementation of adaptive IIR filters with particle swarm optimization algorithm. Communication Systems.

[60] Karim A.M., Mishra A. (2022) Novel COVID-19 Recognition Framework Based on Conic Functions Classifier. In: Garg L., Chakraborty C., Mahmoudi S., Sohmen V.S. (eds) Healthcare Informatics for Fighting COVID-19 and Future Epidemics. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-72752-9_1.

[61] Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xues, Z., Shen, D., Shi, Y. 2020. Lung Infection Quantification of COVID-19 in CT Images with Deep Learning. Arxiv, 2003, 04655.

[62] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., Wang, X. 2020. Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label. IEEE Transactions on Medical Imaging, 1–13.

[63] Cohen, J. P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, F., Shen, B., Mahsa, H. K., Ghassemi, M., Li, H., Duong, T.Q. 2020. Predicting COVID-19 51 Pneumonia Severity on Chest X-ray With Deep Learning Cureus, 12(7), 9448. COVID 19 WHO 16 ağustos 2021 verileri. Web sitesi. https://covid19.who.int/2021. Erişim tarihi 16.08.2021.

[64] Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., Bernheim,

A., Siegel, E. 2020. Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic : Initial Results for Automated Detection Patient Monitoring using Deep Learning CT Image Analysis.Arxiv, 2003, 05037

[65] Javor, D., Kaplan, H., Kaplan, A., Puchner, S. B., Krestan, C., Baltzer, P. 2020. Deep learning analysis provides accurate COVID-19 diagnosis on chest computed tomography. European Journal of Radiology, 133, 109402.

[66] Jain, G., Mittal, D., Thakur, D., Mittal, M. K. 2020. A deep learning approach to detect Covid-19 coronavirus with X-Ray images. Science Direct, 40(4), 1391–1405.

[67] Wang, L., Lin, Z. Q., Wong, A. 2020. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1), 1–12.

[68] Gilanie, G., Bajwa, U. I., Waraich, M. M., Asghar, M., Kousar, R., Kashif, A., Aslam, R. S., Qasim, M. M., Rafique, H. 2021. Coronavirus (COVID-19) Detection from Chest Radiology Images using Convolutional Neural Networks.Biomedical Signal Processing and Control, 66, 02490.

[69] Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., Dong, Y. 2021. The ensemble deep learning model for novel COVID-19 on CT images. Applied Soft Computing, 98, 106885.

[70] Turkoglu, M. 2021. COVID-19 Detection System Using Chest CT Images and Multiple Kernels-Extreme Learning Machine Based on Deep Neural Network.IRBM, 1– 8.

[71] Huang, L., Han, R., Ai, T., Yu, P., Kang, H., Tao, Q., Xia, L. 2020. Serial Quantitative Chest CT Assessment of COVID-19 :A Deep Learning Apporach. Radiology: Cardiothoracic Imaging, 2(2) ,200075.

[72] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.J. 2020. Deep-COVID : Predicting COVID-19 from chest X-ray images using deep transfer learning.Medical Image Analysis, 65, 101797.

[73] Xiyun Yang, Yanfeng Zhang, Wei Lv, Dong Wang, Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier, Renewable Energy, 2020, ISSN 0960-1481,..

[74] Chuan Li, Shaohui Zhang, Yi Qin, Edgar Estupinan, A systematic review of deep transfer learning for machinery fault diagnosis, Neurocomputing, Volume 407, 2020, Pages 121-135, ISSN 0925-2312,..

[75] Yue Wang, Yuting Liu, Wei Chen, Zhi-Ming Ma, Tie-Yan Liu, Target transfer Q-learning and its convergence analysis, Neurocomputing, Volume 392, 2020, Pages 11-22, ISSN 0925-2312,..

[76] Zhengjun Qiu, Shutao Zhao, Xuping Feng, Yong He, Transfer learning method for plastic pollution evaluation in soil using NIR sensor, Science of The Total Environment, Volume 740, 2020, 140118, ISSN 0048-9697,..

[77] Santi Kumari Behera, Amiya Kumar Rath, Prabira Kumar Sethy, Maturity status classification of papaya fruits based on machine learning and transfer learning approach, Information Processing in Agriculture, 2020, ISSN 2214-3173,..

[78] Seunghyeon Kim, Yung-Kyun Noh, Frank C. Park, Efficient neural network compression via transfer learning for machine vision inspection, Neurocomputing, Volume 413, 2020, Pages 294-304, ISSN 0925-2312,..

[79] Xiang Li, Wei Zhang, Hui Ma, Zhong Luo, Xu Li, Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks, Neural Networks, Volume 129, 2020, Pages 313-322, ISSN 0893-6080,..

[80] Piyush Kant, Shahedul Haque Laskar, Jupitara Hazarika, Rupesh Mahamune, CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces, Journal

of Neuroscience Methods, Volume 345, 2020, 108886, ISSN 0165-0270,..

[81] Xinghua Li, Zhongyuan Hu, Mengfan Xu, Yunwei Wang, Jianfeng Ma, Transfer learning based intrusion detection scheme for Internet of vehicles, Information Sciences, Volume 547, 2021, Pages 119-135, ISSN 0020-0255,..

[82] Sheikh Saud, Basharat Jamil, Yogesh Upadhyay, Kashif Irshad, Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach, Sustainable Energy Technologies and Assessments, Volume 40, 2020, 100768, ISSN 2213-1388,..

[83] Jie Wei, Hui Chen, Determining the number of factors in approximate factor models by twice K-fold cross validation, Economics Letters, Volume 191, 2020, 109149, ISSN 0165-1765,..

[83] Jie Wei, Hui Chen, Determining the number of factors in approximate factor models by twice K-fold cross validation, Economics Letters, Volume 191, 2020, 109149, ISSN 0165-1765,..

[84] Tzu-Tsung Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognition, Volume 48, Issue 9, 2015, Pages 2839-2846, ISSN 0031-3203,..

[85] Gaoxia Jiang, Wenjian Wang, Error estimation based on variance analysis of k-fold cross-validation, Pattern Recognition, Volume 69, 2017, Pages 94-106, ISSN 0031-320..

[86] Nima Shiri Harzevili, Sasan H. Alizadeh, Analysis and modeling conditional mutual dependency of metrics in software defect prediction using latent variables, Neurocomputing, Volume 460, 2021, Pages 309-330, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.05.043..

[87] Chao Ni, Xiang Chen, Fangfang Wu, Yuxiang Shen, Qing Gu, An empirical study on pareto based multi-objective feature selection for software defect prediction, Journal of Systems and Software, Volume 152, 2019, Pages 215-238, ISSN 0164-1212.

[88] Ruchika Malhotra, Shine Kamal, An empirical study to investigate oversampling methods

for improving software defect prediction using imbalanced data, Neurocomputing, Volume 343, 2019, Pages 120-140, ISSN 0925-2312.

[89] Diana-Lucia Miholca, Gabriela Czibula, Istvan Gergely Czibula, A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks, Information Sciences, Volume 441, 2018, Pages 152-170, ISSN 0020-0255..

[90] Yuanxun Shao, Bin Liu, Shihai Wang, Guoqi Li, A novel software defect prediction based on atomic class-association rule mining, Expert Systems with Applications, Volume 114, 2018, Pages 237-254, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2018.07.042..

[91] Ning Li, Martin Shepperd, Yuchen Guo, A systematic review of unsupervised learning techniques for software defect prediction, Information and Software Technology, Volume 122, 2020, 106287, ISSN 0950-5849..

[92] Zhongbin Sun, Jingqi Zhang, Heli Sun, Xiaoyan Zhu, Collaborative filtering based recommendation of sampling methods for software defect prediction, Applied Soft Computing, Volume 90, 2020, 106163, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106163..

[93] Xingjuan Cai, Shaojin Geng, Di Wu, Jinjun Chen, Unified integration of many-objective optimization algorithm based on temporary offspring for software defects prediction, Swarm and Evolutionary Computation, Volume 63, 2021, 100871, ISSN 2210-6502..

[94] Shuo Feng, Jacky Keung, Peichang Zhang, Yan Xiao, Miao Zhang, The impact of the distance metric and measure on SMOTE-based techniques in software defect prediction, Information and Software Technology, Volume 142, 2022, 106742, ISSN 0950-5849..

[95] Cong Jin, Cross-project software defect prediction based on domain adaptation learning and optimization, Expert Systems with Applications, Volume 171, 2021, 114637, ISSN 0957-4174,

[96] Lei Qiao, Xuesong Li, Qasim Umer, Ping Guo, Deep learning based software defect prediction, Neurocomputing, Volume 385, 2020, Pages 100-110, ISSN 0925-2312.

[97] Meetesh Nevendra, Pradeep Singh, Empirical investigation of hyperparameter optimization for software defect count prediction, Expert Systems with Applications, Volume 191, 2022, 116217, ISSN 0957-4174.

[98] Hua Wei, Changzhen Hu, Shiyou Chen, Yuan Xue, Quanxin Zhang, Establishing a software defect prediction model via effective dimension reduction, Information Sciences, Volume 477, 2019, Pages 399-409, ISSN 0020-0255..

[99] Zhiguo Ding, Liudong Xing, Improved software defect prediction using Pruned Histogram-based isolation forest, Reliability Engineering & System Safety, Volume 204, 2020, 107170, ISSN 0951-8320.

[100] Beyza Eken, Ayse Tosun, Investigating the performance of personalized models for software defect prediction, Journal of Systems and Software, Volume 181, 2021, 111038, ISSN 0164-1212..

[101] Chao Sui, Mohammed Bennamoun, Roberto Togneri, Deep feature learning for dummies: A simple auto-encoder training method using Particle Swarm Optimisation, Pattern Recognition Letters, Volume 94, 2017, Pages 75-80, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2017.03.021.

[102] Hasan Badem, Alper Basturk, Abdullah Caliskan, Mehmet Emin Yuksel, A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited–memory BFGS optimization algorithms, Neurocomputing, Volume 266, 2017, Pages 506-526, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2017.05.061.

[103] Xiao-Han Zhou, Min-Xia Zhang, Zhi-Ge Xu, Ci-Yun Cai, Yu-Jiao Huang, Yu-Jun Zheng, Shallow and deep neural network training by water wave optimization, Swarm and Evolutionary Computation, Volume 50, 2019, 100561, ISSN 2210-6502, https://doi.org/10.1016/j.swevo.2019.100561.

[104] Omid David and Iddo Greental, Genetic algorithms for evolving deep neural networks, GECCO, 2014. DOI:10.1145/2598394.2602287.

[105] V. Nandagopal, S. Geeitha, K. Vinoth Kumar, J. Anbarasi, Feasible analysis of gene expression –a computational based classification for breast cancer, Measurement, Volume 140, 2019, Pages 120-125, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2019.03.015.

[106] Ping He, Baichuan Fan, Xiaohua Xu, Jie Ding, Yali Liang, Yuan Lou, Zhijun Zhang, Xincheng Chang, Group K-SVD for the classification of gene expression data, Computers & Electrical Engineering, Volume 76, 2019, Pages 143-153, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2019.03.009.

[107] Md. Maniruzzaman, Md. Jahanur Rahman, Benojir Ahammed, Md. Menhazul Abedin, Harman S. Suri, Mainak Biswas, Ayman El-Baz, Petros Bangeas, Georgios Tsoulfas, Jasjit S. Suri, Statistical characterization and classification of colon microarray gene expression data using multiple machine learning paradigms, Computer Methods and Programs in Biomedicine, Volume 176, 2019, Pages 173-193, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2019.04.008.

[108] Sarah M. Ayyad, Ahmed I. Saleh, Labib M. Labib, Gene expression cancer classification using modified K-Nearest Neighbors technique, Biosystems, Volume 176, 2019, Pages 41-51, ISSN 0303-2647, https://doi.org/10.1016/j.biosystems.2018.12.009.

[109] Huijuan Lu, Junying Chen, Ke Yan, Qun Jin, Yu Xue, Zhigang Gao, A hybrid feature selection algorithm for gene expression data classification, Neurocomputing, Volume 256, 2017, Pages 56-62, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2016.07.080.

[110] Yuanyu He, Junhai Zhou, Yaping Lin, Tuanfei Zhu, A class imbalance-aware Relief algorithm for the classification of tumors using microarray gene expression data, Computational Biology and Chemistry, Volume 80, 2019, Pages 121-127, ISSN 1476-9271, https://doi.org/10.1016/j.compbiolchem.2019.03.017.

[111] Sai Prasad Potharaju, M. Sreedevi, Distributed feature selection (DFS) strategy for microarray gene expression data to improve the classification performance, Clinical Epidemiology and Global Health, Volume 7, Issue 2, 2019, Pages 171-176, ISSN 2213-3984, https://doi.org/10.1016/j.cegh.2018.04.001.

[112] Lingyun Gao, Mingquan Ye, Xiaojie Lu, Daobin Huang, Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification,

Genomics, Proteomics & Bioinformatics, Volume 15, Issue 6, 2017, Pages 389-395, ISSN 1672-0229, https://doi.org/10.1016/j.gpb.2017.08.002.

[113] Ahmad M. Karim, Mehmet S. Güzel, Mehmet R. Tolun, Hilal Kaya, Fatih V. Çelebi, A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing, Biocybernetics and Biomedical Engineering, Volume 39, Issue 1, 2019, Pages 148-159, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2018.11.004.

[114] Ahmad M. Karim, Mehmet S. Güzel, Mehmet R. Tolun, Hilal Kaya, and Fatih V. Çelebi, A New Generalized Deep Learning Framework Combining Sparse Autoencoder and Taguchi Method for Novel Data Classification and Processing, Mathematical Problems in Engineering, vol. 2018, Article ID 3145947, 13 pages, 2018. https://doi.org/10.1155/2018/3145947.

[115] Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, Grey Wolf Optimizer, Advances in Engineering Software, Volume 69, 2014, Pages 46-61, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2013.12.007.

[116] Siva Shankar G., Manikandan K., Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization, Pattern Recognition Letters, Volume 125, 2019, Pages 432-438, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2019.06.005.

[117] S. Mirjalili, How effective is the grey wolf optimizer in training multi-layer perceptrons, Appl. Intell., vol. 43, Jul. 2015, no. 1, pp. 150-161, https://doi.org/10.1007/s10489-014-0645-7.

[118] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," Science, vol. 286, no. 5439, pp. 531–537, 1999.

[119] Alon, U., Barkai, N., Notterman, D., Gish, K., Ybarra, S., Mack, D. and Levine, A. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences, 96(12), pp.6745-6750.

[120] Singh, D., Febbo, P., Ross, K., Jackson, D., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D'Amico, A., Richie, J., Lander, E., Loda, M., Kantoff, P., Golub, T. and Sellers, W. (2002). Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, 1(2), pp.203-209.

[121] Beer, D. G., Kardia, S. L., Huang, C.-C., et al. (2002), "Gene-expression profiles predict survival of patients with lung adenocarcinoma," Nature medicine, 8(8), 816–824.