



REPUBLIC OF TURKEY
ALTINBAŞ UNIVERSITY
Institute of Graduate Studies
Information Technologies

**DATA MINING TECHNIQUES FOR
EXTRACTION AND ANALYSIS OF COVID-19
DATA**

Mohammed Ghanim Fathi AL-OBAIDI

Master's Thesis

Supervisor

Asst. Prof. Ayça KURNAZ TÜRKBEN

Istanbul, 2022

**DATA MINING TECHNIQUES FOR EXTRACTION AND ANALYSIS
OF COVID-19 DATA**

Mohammed Ghanim Fathi AL-OBAIDI

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled “DATA MINING TECHNIQUES FOR EXTRACTION AND ANALYSIS OF COVID-19 DATA” prepared BY MOHAMMED GHANIM FATHI AL-OBAIDI and submitted on 10/08/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Asst. Prof. Dr. Ayça KURNAZ TÜRK BEN

The Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Ayça KURNAZ
TÜRK BEN

Faculty of Engineering and
Architecture,

Altınbaş University

Asst. Prof. Dr. Oğuz KARAN

Faculty of Engineering and
Architecture,

Altınbaş University

Asst. Prof. Dr. Serdar KARGIN

Faculty of Biomedical
Engineering,

Beykent University

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

Submission date of the thesis to Institute of Graduate Studies. ___/___/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Mohammed Ghanim Fathi AL-OBAIDI

Signature

DEDICATION

I sincerely thank my great supervisor Professor Dr. Ayça KURNAZ TÜRKBEN for helping me conduct the research and prepare the thesis. This thesis is wholeheartedly dedicated to my beloved family, who has been my source of inspiration.



ABSTRACT

DATA MINING TECHNIQUES FOR EXTRACTION AND ANALYSIS OF COVID-19 DATA

AL-OBAIDI , Mohammed Ghanim Fathi

M.Sc., Information Technologies ,Altınbaş University,

Supervisor: Asst. Prof. Dr.Ayça KURNAZ TÜRK BEN

Date: 08.2022

Page Number: 57

In this research, data mining techniques that included the deep learning with different scenario is presented for extraction and analysis of covid-19 data. Energy of the features are implemented and calculated from the CT scan images. Different patients with different problems are investigated and analysed. Also the results are compared with other studies. The results of proposed method shows that the proposed method has high accuracy than other methods.

Keywords: Data Mining Techniques, Extraction and Analysis, Covid-19 Data

TABLE OF CONTENTS

	<u>Pages</u>
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
1. INTRODUCTION	1
1.1 BACKGROUND.....	1
1.2 USEAGE AREAS IS DATA MINING.....	1
1.3 STEPS OF DATA MINING	2
1.4 DATA MINING TECHNIQUES AND METHODS.....	2
1.4.1 Classification	2
1.4.2 Clustering.....	2
1.4.3 Reinforcement Learning.....	2
1.5 STEPS IN THE DATA MINING PROCESS	3
1.5.1 Data Cleaning	3
1.5.2 Data Integration	3
1.5.3 Data Selection.....	3
1.5.4 Data Transformation.....	4
1.5.5 Data Mining.....	4
1.5.6 Pattern Evaluation.....	4
1.5.7 Knowledge Representation.....	4
1.6 PROBLEM STATEMENT	4
1.7 THESIS ORGANIZATION	5

2.	LITERATURE REVIEW	6
2.1	BACKGROUND.....	6
3.	MATERIAL AND METHOD	19
3.1	COVID-19 DATASET	19
4.	EXPERIMENTAL RESULTS AND DISCUSSION	28
4.1	BACKGROUND.....	28
4.2	METHOD	28
4.3	RESULTS.....	32
4.4	COVID-19 DETECTION DATASET	32
4.5	COVID-19 DATASET 2.....	36
4.6	COVID-19 DATASET 3.....	40
5.	CONCLUSION AND FUTURE WORK.....	45
5.1	CONCLUSION	45
5.2	FUTURE WORK	46
	REFERENCES.....	47

LIST OF TABLES

	<u>Pages</u>
Table 4. 1: Covid-19 dataset 1 results compressions.....	36
Table 4. 2: COVID-19 dataset 2 results comparisons.....	39
Table 4. 3: COVID-19 dataset 3 results comparisons.....	44



LIST OF FIGURES

	<u>Pages</u>
Figure 3. 1. Three Classes Of COVID-19 Dataset; (1) COVID-19, (2) Pneumonia, And (3) Normal	19
Figure 3. 2. U-Net Structure (Sample 32×32 Pixels With Lower Resolution) [34]	21
Figure 3. 3. Teaching Methods [39].....	22
Figure 4. 1: Proposed Framework.....	31
Figure 4. 2: COVID-19 Dataset 1 Using Our Method.....	35
Figure 4. 3: COVID-19 Dataset 2 Using Our Method.....	39
Figure 4. 4: COVID-19 Dataset 3 Results	43

ABBREVIATIONS

CRM	:	Customer Relationship Management
WHO	:	World Health Organization
COVID-19	:	Corona Virus Diagnosis-19
CXR	:	Chest X-Rays
PSO	:	Particle Swarm Optimization
GLCM	:	Gray Level Co-Occurrence Matrix
AUC	:	Area Under the Curve

1. INTRODUCTION

1.1 BACKGROUND

Data mining means extracting valuable information from huge volumes of data mines. You can think of data mining as a problem-solving method that analyzes large volumes of data to extract repetitive patterns. It then provides solutions to the challenges by finding connections between these patterns. In fact, Data Mining turns useful and useless information into usable information by obtaining useful and valuable results [1].

In answer to the question of what data mining is and what it is used for, we can say that data mining means discovering knowledge within data. Discovering knowledge within data is one of the most important and effective concepts in the information age, which is becoming more and more important every day.

1.2 USEAGE AREAS IS DATA MINING

Companies and organizations that use data mining to analyze competitors and markets can easily predict current trends. Therefore, in their future plans, they move in line with the needs of the general public and attract the attention of customers before other competitors [2].

Today, this issue is used in various educational, political, economic and other fields. Here are some different applications of data mining.

Public Health: Activities to spread the culture of public health at the lowest cost in different parts of the world [3].

Education: Activities to improve the quality of the education system and guide students properly [4].

Construction: Activities to facilitate road construction and the application of urban optimization models due to population growth.

Customer Relationship Management (CRM): Activity to improve the organization's relationship with customers and ultimately increase productivity.

Purchasing Market Research: This item seeks to identify items related to the customer's shopping cart to increase their purchasing power.

1.3 STEPS OF DATA MINING

In this section, we want to briefly get acquainted with the general steps in a Data Mining process. These steps are summarized as follows [5]:

Extract, transfer and store data in multidimensional databases

Provide access to business layer data through data mining software

Display the results of data analysis in the form of graphs or graphs

1.4 DATA MINING TECHNIQUES AND METHODS

Using data mining techniques, the speed of calculations and the space required in memory (RAM) is significantly improved. In general, data mining techniques can be grouped into one of the following three categories or a combination of them [6].

1.4.1 Classification

In this type of learning, data is labeled based on defined characteristics and placed in different classes. This algorithm is able to learn the labeling model and, using the intelligent learning system, label new samples and differentiate them. This separation is a kind of learning and the algorithm after this learning, can apply its model to new data.

1.4.2 Clustering

In this case, the algorithm groups the data based on their nature. For example, customers divide an online store into different clusters based on their similarities (age, gender, level of education, etc.).

1.4.3 Reinforcement Learning

In this learning, the algorithm, by exchanging information and operations with the environment, continuously discovers information and learns. Consider, for example, an algorithm that, by

interacting with the environment and simulating it intelligently, designs different types of shopping cart forms to create the best design for customers and ultimately increase sales and profits.

1.5 STEPS IN THE DATA MINING PROCESS

So far we know what data mining is and what its steps and techniques are. It is now good to know that data mining is also known as Knowledge Discovery From Data, which means the process of extracting knowledge and information from data in the database.

Data mining involves several steps. This process begins with raw data and continues until new knowledge is formed. In the following, we will examine these steps in the form of data mining training.

1.5.1 Data Cleaning

Data wiping refers to the process of detecting, deleting, and correcting inaccurate data from a set of tables, records, or databases, as well as identifying incomplete and inaccurate pieces of data and then correcting and replacing them. The purpose of data cleansing is to extract accurate and correct information, because incorrect information can lead to wrong conclusions and cause problems for your business.

1.5.2 Data Integration

Information integration creates a relatively new insight into customers, products, marketing channels, etc., and provides a good platform for a comprehensive and complete view of the core elements of a business in an organization. Without data integration, you will not have much to say in today's competitive marketplace [7].

1.5.3 Data Selection

In the selection section, data related to data analysis must be selected and retrieved from the dataset to perform the analysis. A principled and correct choice can improve inductive learning in many ways, including learning speed and generalized capacity.

1.5.4 Data Transformation

Sometimes, in order to increase the accuracy of the analysis, we have to make changes to the raw data available to us for analysis, one of which is the data conversion process. Data conversion is a mathematical method used for variables that do not follow the statistical indicators of normality, linearity, uniform scattering, and so on [8].

Data conversion is also a method of data consolidation. In this phase, the selected data is converted to another form. This helps simplicity, accuracy and precision of data mining.

1.5.5 Data Mining

This section uses clever methods to extract important and effective patterns from the data. These methods include the following:

Swarm intelligence approach using artificial bee colony to solve optimization problems

Apriori algorithm with implementation code in Python

Apriori algorithm and exploring repetitive patterns in data mining

1.5.6 Pattern Evaluation

In this section, the patterns obtained in the previous section are examined and evaluated before various aspects such as accuracy, correctness, generalizability and..

1.5.7 Knowledge Representation

Data mining ultimately leads to the presentation of knowledge. The knowledge obtained in this section is presented to the user in a clear and understandable way. Of course, for more effectiveness, visualization methods are also used, which despite these methods, users will be more successful in understanding and interpreting data mining results.

1.6 PROBLEM STATEMENT

The computer-aided analysis of MRI and X-ray analysis leads to gain in processing such as time, cost, and quality. Exactly in last year, the Covid-19 pandemic leads to an increase in the number

of X-ray images that are requested to detect the Covid-19 in all healthcare institutions. This lead that the number of doctors, physicians, and expertise becomes less to deal with this case. Furthermore, in most hospitals, the expertise that can deal with these cases is not found. This concludes to find new alternatives speedy can deal with this problem. Then, researchers developed many machine learning, data mining, and deep learning applications to deal with this problem. One of the technical problems faced by researchers is the high computation time in convolutional neural networks. The extracted features from pre-trained models lead to a fixed number of features which a number of these features unaffected most of the time.

1.7 THESIS ORGANIZATION

We have organized this thesis as Chapter 2 includes a review on various methodologies for optimization methods that used in the thesis. Chapter 3 describes the depicts the implementation of the selected methodology and illustrates the methods and analysis on the model. In Chapter 4, the proposed algorithm has also been presented for a proposed system and analyzed under various conditions, and then Chapter 5 discusses the conclusions and future scope for research.

2. LITERATURE REVIEW

2.1 BACKGROUND

Medicine is one of the fields where computer science advancement is making significant progress. The usage of medical computers improves precision and accelerates data processing and diagnosis. There are currently a variety of computer-assisted diagnostic systems, with deep-learning algorithms playing an important role. Systems that are more precise and faster are required. Computer-aided diagnostic systems and various sets of medical data analysis software utilize classification as a common deep-learning application. Medical datasets contain a large number of features, many of which are related to one another, making it imperative to decrease the number of features. To achieve improved classification accuracy, it is important to select functional sets and adequate features for the classification model. The computer-aided analysis of Magnetic Resonance Imaging (MRI) and X-ray analysis leads to gain in processing such as time, cost, and quality. Exactly in the last years, the World Health Organization (WHO) declared the current epidemic to be a pandemic [9], the Corona Virus Diagnosis-19 (COVID-19) pandemic leads to an increase in the number of X-ray images that are requested to detect the COVID-19 in all healthcare institutions. This lead that the number of doctors, physicians, and expertise becomes less to deal with this case. Furthermore, in most hospitals, the expertise that can deal with these cases is not found. This concludes to find new alternatives speedy can deal with this problem.

The other important medical issue is the covid-19 pandemic, which afflicted the entire planet, began in China in the final months of 2019. The World Health Organization (WHO) classified the virus as SARS-CoV-2 as its official designation (severe acute respiratory syndrome-Coronavirus-2) [10]. Because of the virus's rapid spread and the deaths of many individuals, WHO has designated it a "pandemic." It is well established that air and physical contact are the most important factors in the virus's propagation. According to reports, this virus causes severe pneumonia by directly attacking the lungs [11]. It is challenging and time-consuming to identify because it is an RNA virus. To minimize the impact of the covid-19 virus, early detection is critical [12][13]. As a result, those infected with the virus are more likely to survive the life-threatening scenario and receive prompt treatment. For all diagnoses of covid-19 disease, the WHO recommends the "reverse transcription-polymerase chain reaction (RT-PCR)" test. This procedure,

however, is inconvenient, time-consuming, and costly. Furthermore, the real detection sensitivity is weak [14][15]. This test is insufficient on its own, and medical image processing procedures like chest x-rays (CXR) should be used to reinforce it.

The motivation of this study is that the expanding number of patients is a significant challenge in diagnosing clinical big data. In this study, an automated method based on effective features is employed to improve diagnosis quality and reduce prediction error rates. The PSO algorithm is used in the proposed method to select the effective and important features based on the natural swarm behavior of birds and fish for obtaining food, and the proposed method's second phase uses various machine learning and deep learning methods to classify covid-19 and brain tumors by analyzing MRI and CXR images. Deep learning approaches will be used for feature extraction combined with the PSO algorithm that will select the best features and improve the performance of the models. The PSO method was applied to exclude irrelevant and unnecessary characteristics from high-dimensional datasets. Several evaluation parameters will calculate to prove results outcomes such as Accuracy, sensitivity, and specificity.

The main contributions of this study:

- a. Extraction of features from the medical image-dataset using auto-encoder and well-known pre-trained CNNs, Alexnet, Googlenet, Resnet50, and Densenet201.
- b. Selection of the most significant features using the PSO algorithm to enhance the accuracy of detection.
- c. A generic framework is proposed that can work on multiple datasets, such as MRIs and X-rays.
- d. Evaluation of the proposed classification model against the current baseline diagnostic models.

The Particle Swarm Optimization (PSO) method was invented by Kennedy and Eberhart in the mid-1990s. In the PSO method, a set of randomly selected solutions (primary group) is published in the design space in order to achieve the optimal solution among a number of iterations (movements) based on a large amount of information about the design space that is combined and All members of the group benefit from it. The PSO method is inspired by the ability of flocks of birds, fish, and flocks of animals to adapt to the environment, find rich food sources and stay away

from predators (fishermen) by implementing the method of information sharing and has an evolutionary advantage. References [16] and [17] describe the complete history of the development of the PSO algorithm and act as a motion simulator for the innovative optimization method.

An important challenge in diagnosing clinical big data is the growing number of patients. In this study, an automated method based on effective features is used to increase the speed of diagnosis and reduce the error of prediction rate. In the proposed method, the modified ant colony optimization algorithm is used for selecting the effective and important features based on the intelligent behavior of ant life for getting the food, and in the second phase of the proposed method, the different machine learning and deep learning methods used to classify the COVID-19 by analyzing MRI images. Deep learning techniques enhanced by combining with ant colony optimization algorithm will apply as a feature extractor to select the best features and enhance the accuracy of the model. The ant colony algorithm was used to remove trivial and redundant characteristics from high-dimensional datasets. Several evaluation parameters will calculate to prove results outcomes such as Accuracy, sensitivity, and specificity. In the last years, several studies based on deep learning techniques presented either for COVID-19 classification.

The proposed method outperforms other classification methods in terms of accuracy. The use of feature selection allows deep learning to focus on the most essential features, decreasing the error rate in detecting infected from healthy people.

Jainy et. al. [18], principal component analysis (PCA) is used for reduction of dimensionality of the feature space and Artificial Neural Network (ANN) is used for classification.

In [19], they proposed a strategy to upgrade the characterization execution. To start with, the increased tumor locale by means of picture expansion is utilized as the ROI rather than the first tumor district since tumor encompassing tissues can likewise offer significant hints for tumor types. Second, the expanded tumor area is parted into progressively fine ring-structure sub-regions. The adequacy of this technique was assessed on an enormous dataset with three component extraction strategies, specifically, force histogram, Gray Level Co-Occurrence Matrix (GLCM), and Bag of-words (BOW) model.

The GLCM that contains the contrast and energy image and also the input image is used for the feature extraction. The accuracy obtained was 96.5%. Navid Ghassemi et. al. [20] introduced an unsupervised pretraining technique to defeat overfitting utilizing generative adversarial networks. The unsupervised pretraining permits utilizing comparable unlabeled datasets. This technique can be utilized for preparing complex organizations on little datasets. Simulation results are acquired on a brain tumor classified dataset.

A number of studies have been presented to diagnose COVID 19 disease in the world, most of which use CT scan image analysis and try to diagnose coronary heart disease in individuals with image processing methods. In this method, lung images of individuals are considered as educational input and a machine learning method is taught to diagnose the affected area [21], including methods for analyzing lung images to diagnose COVID disease. Used include artificial neural network [22], support vector machine [23], deep learning [24], random forest [25], Stacked-autoencoder-based model [26], Deep Convolutional Autoencoder [27], CNN with Autoencoder [28].

Swati et. al. [29], used the transfer learning and fine-tuning for MRI classification. The used the pretrained CNN model and the new fine tuning scenario that it was depended on the transfer learning. They achived 94.82% from 5 fold cross validation number.

An important challenge in the methods of diagnosing coronary heart disease using a CT scan of the lungs is that this method is not always available and on the other hand, due to the use of ionizing radiation, it cannot be used in large numbers. The use of X-rays to image the lungs is dangerous for people with cancer and pregnant women [30]. Machine learning algorithms have been widely used in medical applications for decades, such as diagnosing disease, proposing treatment, and prescribing medication. Today, various neural networks, such as multilayer neural networks, are a subset of machine learning algorithms with a unique structure of artificial neural networks that are used to classify images or analyze patient records. Despite the development of a large number of studies in the field of lung image processing for the diagnosis of COVID disease, a small number have been developed based on the information of patients and their records for the diagnosis of coronary heart disease. To diagnose COVID 19 disease, various methods have been developed so far, one of which can be the analysis of information collected by medical centers with machine learning methods. So far, a large number of data sets related to coronary heart disease have been

presented in valid databases. Each data set has a number of features for diagnosing the disease in individuals, but not all the features that exist in coronary heart disease are necessarily important. Others are also important because of their attack is the age and gender of the people. An important way to determine the extent to which a person has a coronary artery is to use a learning method such as a neural network, provided that only important features of education are considered. Here, a feature selection method based on the Ant Colony Optimization algorithm [31] is used to use only the important features of coronary artery disease for learning in artificial neural network and to consider features that are less important. Our contribution to this research is to use the learning of living things to diagnose emerging coronary heart disease, to provide a binary version of the Ant colony learning algorithm, to combine ant learning and artificial neural network to diagnose the COVID 19 dataset, and to present an automated method for diagnosing COVID 19 disease. In [32]], they introduced the COVID-Net, which is the deep convolutional neural network designed for the detection of COVID-19 from the X-ray imaging system. Also, they introduced the new open-access benchmark that has included the 13,975 CXR images across 13,870 patient cases dataset. The three-player knowledge transfer and concentration framework counting a pre-trained appearing network that abstracts the CXR imaging features from a large scale of lung disease CXR images is presented in [33]. In their work the data transfer from the AP network to the RF network. Parnian Afshar et. al. [34] presented a method based on the capsule network for COVID-19 cases recognition from X-ray images. They approved that their method is efficient than the CNN method and they obtained the 95.7%, 90%, 95.8%, and 0.97 for Accuracy, Sensitivity, Specificity, and Area Under the Curve (AUC) respectively. The deep learning is based on the DarkNet model used in [35] for the classification of the binary and multi-classes cases of COVID-19. In this model, the classification model is used for the YOLO real-time object detection system. The classification accuracy rate was 98.08% and 87.02% for binary classes and multi-class cases respectively. A rapid and valid method is used in [14] for the diagnosis of COVID-19 with an artificial neural network. They used and tested ten convolutional neural networks to distinguish infection of COVID-19 cases from non-COVID-19. From ResNet-101 the result was 0.994, 100%, 99.02%, 99.51% for AUC, sensitivity, specificity and accuracy respectively. For the Xception model, the results were 0.994, 98.04%, 100%, and 99.02% for AUC, sensitivity, specificity, and accuracy respectively.

Swati et. al. [29], used transfer learning and fine-tuning for MRI classification. They used the pre-trained CNN model and the new fine-tuning scenario that was dependent on transfer learning. They achieved 94.82% from 5-fold cross-validation number.

One of the interesting behaviors in nature for finding food, which has a high intelligence nature, is the behavior of ants. Ants have a clever way of finding the food that can be used to find optimal and to reduce their output error in finding the shortest way. In this thesis, the behavior of ants is first introduced and a modified ant colony optimization for feature selection method is introduced based on it. In the second phase, the deep learning and autoencoder-based method are used for classifying the COVID-19 disease and brain tumors data [31].

The ant colony optimization algorithm was proposed in 1990 by Marco Dorigo as an innovative method for solving hybrid optimization problems [31]. This algorithm is derived from the actual behavior of ants to find food in the shortest path [36]. Each ant leaves a chemical material that called pheromone in the path of finding food, which other ants choose the shortest path using the previous pheromones. This algorithm is very useful for solving non-deterministic polynomial (NP) problems and is used in problems such as itinerant vendor, scheduling problem, vehicle routing problem, etc. [37].

To solve any NP problem using the ant colony algorithm, the following must be specified:

- a. First, you have to turn the problem into a graph, including nodes and edges.
- b. See distance nodes (η) are raised and specified.
- c. A possible solution is created according to the problem.
- d. The pheromone update rule is used to determine the effective edges in achieving the best answer.
- e. The probabilistic transition rule is used to find the next node [38].

Several different implementations of ant colony algorithms such as Ant System, Max-Min Ant System, and Ant Colony System have been proposed, the main difference of these methods is in the pheromone update formula [38].

The associated COVID-19 disease dataset, as well as DM approaches like text mining, are investigated in [21]. The use of DM approaches to analyze the data of scientific researches on mortality regarding the first corona-virus wave was recommended in the work [22]. They use DM to attempt to uncover hidden COVID-19 patterns with the use of immunity, mortality, and vaccination data. The impact of several medicines on COVID-19 was examined with the use of DM approaches in the work [23]. The findings of their studies reveal that the drug combination of Azithromycin and Hydroxychloroquine, while effective in treating COVID-19 disease, can also cause digestive infections and thin illness in people. According to their findings, such medication combinations should be used with caution in individuals with immune system diseases, CVD, or acquired fat problems. A demonstrated method for the prediction of the crown pestilences using a half breed model of ANN has been offered in the study [24]. The COVID disease that had spread to over two-hundred countries and was declared a pandemic via the WHO, could be a challenging task. With limited information, predicting COVID-19 infection could be complicated. To predict the COVID-19 pestilence, researchers hope to propose a cross breed model that includes test mode assessment and ANN. Time-arrangement data between Jan. 22nd, 2020, and May 18th, 2020, was used in this investigation. Time was organized using exploratory mode parsing to construct sub-flags and misshape genuine data, and the consolidated information was prepared using ANN. The results of their studies suggest that the suggested approach works better than traditional measurable examination.

A DL technique for the modelling and forecasting of COVID-19 in 5 Indian states has been presented in the paper [25]. DL is employed in this work to develop comparative analytical model regarding COVID-19 outbreak using incremental accuracy of ANNs. In the case when new datasets have been obtained from ever-changing education data, the suggested model intelligently adjusts to new real-time facts. They evaluated and examined DL methods to predict coronavirus infection in [26]. As a result of the rising numbers of the patients on a daily basis, which is a high rate in various nations, lab analyses are time-consuming, and findings and treatment restrictions emerge. A clinical choice model with prescient computations is required due to the limitations of present indicative frameworks. By differentiating illnesses, prescient computations may be able to reduce the strain on medical service systems. They evaluate clinical forecast models using DL and lab data in this investigation. Their suggested approach was validated by 10 approval approaches and 18 lab discoveries from a total of 600 patients. The diagnoses of corona-virus

patients has a sensitivity of 99.42% and accuracy of 88.66%, according to experimental studies. A feature selection approach combined with classification algorithms was provided in [27] for diagnoses of COVID-19 with utilizing computed tomography scans. A CNN has been used to determine the image's main features in the first step. Second, the suggested approach employs the whale optimization algorithm for performing feature selection, and lastly, for diagnosing coronary COVID-19 disease, it employs a suggested classification depending on SVM voting, NNs, nearest neighbors, and decision trees. Experimentations have shown that the suggested method is at least as effective in property selection as particle optimization algorithm. In [28], a DL-based recommendation system for COVID-19 disease was suggested. Physician diagnostic volume is reduced by a system that advises diagnosis. To identify COVID-19 fast, the suggested technique examines a huge volume of CT-Scan images. Their findings suggest that DL models based on the CNN artificial neural network may be able to detect illness. In 2019, irregular backwoods indicator has been introduced with order clinical data using highlight positions in [29]. The authors of this paper have conducted large-scale testing on 10 benchmark datasets and found promising results. They provided extremely precise indications of ten different infections. According to study [30], a whale advancement computation that has been based upon the theory of choppiness will be available in 2020 for better help the precision of the machine in the clinical applications. The suggested disease detection model has been compared to numerous competitive SVM models depending on other algorithms of optimization, such as BFO, PSO and GA.

Jainy et. al. [18], principal component analysis (PCA) is utilized for the reduction of dimensionality of the feature space and ANN is used for classification.

In [19], they proposed a strategy to upgrade the characterization execution. To start with, the increased tumor locale by means of picture expansion is utilized as the ROI rather than the first tumor district since tumor encompassing tissues can likewise offer significant hints for tumor types. Second, the expanded tumor area is parted into progressively fine ring-structure sub-regions. The adequacy of this technique was assessed on an enormous dataset with three component extraction strategies, specifically, force histogram [39].

Thejaswini P et. al. [40], presented the method for the MRI image for segmentation. They used the Adaptive Regularized Kernel-based Fuzzy C-Means Clustering. For brain tumor segmentation and classification, the SVM and ANN combined and used for feature extraction. Their suggested

approach was tested and evaluated on 94 images and they obtained 98%, 78%, and 91.4% for Sensitivity, Specificity, and Accuracy respectively. A truthful and fully automatic system based on preprocessing that does not take huge time is represented in [41]. For extraction of features from MRI brain images, the deep transfer learning approach has been utilized. For improving the performance, a confirmed classifier model was implemented on the feature matrix. The GoogleNet is used in the proposed model and a 98% accuracy is obtained. The deep convolutional neural network-based multi-grade is provided in [42] for the classification of brain tumor. In this method, in multi-grade tumor classification for sidestepping lack of data problem the data augmentation techniques are used. In [43], for improving the accuracy of classification, they adopted and incorporated the CapsNets. Also for the real set data for magnetic resonance imaging system, the over-fitting problem of CapsNets is investigated. They developed a visualization model for the output of the CapsNet to obtain the best features for improving the accuracy. The deep learning method with pre-trained VGG-16 CNNs combination, for classification of 3 brain tumor types that contain the glioma tumor, meningioma and pituitary is presented in [44]. The GLCM that contains the contrast and energy image and also the input image has been utilized for the feature extraction. The accuracy rate that has been obtained was 96.5%. Navid Ghassemi et. al. [20] introduced an unsupervised pretraining technique to defeat overfitting utilizing generative adversarial networks. The unsupervised pretraining permits utilizing comparable unlabeled data-sets. This technique may be utilized for preparing complex networks on little data-sets. Simulation results have been acquired on brain tumor classified dataset.

Some researches were provided for diagnosing COVID 19 disease in the world, most of which utilize the CT scan image analysis and attempt to diagnose coronary heart disease in individuals with image processing methods. In this method, lung images of individuals are considered as educational input and ML technique is taught to diagnose the affected area [21], including methods for analyzing lung images to diagnose COVID disease. Used include ANN [22], SVM [23], DL [24], RF [25], Stacked-autoencoder-based model [26], Deep Convolutional Autoencoder [27], CNN with Autoencoder [28].

Abdu Gumaie et. al. [45], provided a hybrid feature extraction approach depending on the regularized extreme learning machine. They have utilized the brain MRI images for the evaluating

of their proposed method. For the preprocessing and enhancement of the contrast the Min-Max used. In their study 94.233% is obtained from the experiments.

Kaplan et. al. presented a method based upon modified local binary pattern for feature extraction from MRI brain images [46]. In their method at first the filtering has been used in order to normalize and smooth the image, and then use the different modified LBP method that called nLBP and aLBP that the nLBP method was obtained in high accuracy 94.56%. For classification they used the ML such as ANN, RF, K-NN and decision tree.

The liver and brain tumor disease classification based on the CNN, discrete wavelet transform and LSTM is proposed in [47]. In their work the dataset be used from Firat university that contain 56 benign and 56 malignant images. The accuracy that they obtained for the liver tumor and brain tumor was 98.60% and 99.10% respectively.

Swati et. al. [29], used the transfer learning and fine-tuning for MRI classification. The used the pretrained CNN model and the new fine tuning scenario that it was depended on the transfer learning. They achived 94.82% from 5 fold cross validation number.

Brain tumor detection in MRIs using a hybrid technique depending on CNNs and neural autoregressive distribution estimation [47]. A significant issue in the use of lungs' CT scan to diagnose coronary heart disease is that this approach is not often available and, at the same time, it cannot be utilized in large numbers because of the use of ionizing radiation. Using X-rays to examine lungs is risky for pregnant women and cancer patients [14]. For decades, ML algorithms were widely employed in medical applications like therapy recommendations, disease diagnosis, and medication prescription. Multilayer neural networks, for example, are a subset of ML algorithms with a distinctive structure of ANNs which are utilized for classifying images or analyzing patient information. In spite of the fact that many researches in lung image processing were established for diagnosing COVID disease, only a few were created for diagnosing coronary heart disease depending on patient records and information. Many approaches were created for diagnosing COVID 19, one of them is the analysis of data provided by medical centers using ML approaches. A huge number of data sets about coronary heart disease were given in reliable databases so far. Each one of the datasets has some traits that can be used to diagnose the condition in individuals, however not all of the variables found in coronary heart disease are necessary. The gender and age of individuals are also crucial due to their attack. A significant approach for determining the extent to which an individual has has a coronary artery is to use learning approach

like NN, provided that just significant features of education are taken into account. Here, a feature selection technique depending on the Ant Colony Optimization algorithm [48] is utilized to learn in an ANN just the significant elements of coronary artery disease and ignore the less relevant features. Our role in this study is using living things' learning for the diagnosis of the emerging coronary heart disease, provide a binary version of Ant colony learning algorithm, combine ant learning and ANNs for diagnosing COVID 19 dataset, and present an automated approach for the diagnosis of COVID 19. They introduced the COVID-Net, a DNN designed for COVID-19 detection with the use of an X-ray imaging system, in [49]. In addition, they unveiled a new open-access benchmark which contained 13975 CXR images from 13870 patient cases. [50] describes a three-player knowledge transfer and concentration system that includes a pre-trained appearing network extracting the CXR imaging features from a large set of the lung disease CXR images. The data transfer from AP to RF networks in their study. Parnian Afshar et al. [51] provided an approach for COVID-19 case detection from X-rays depending on capsule network. They concluded that their approach is more effective compared with CNN, with sensitivity, accuracy, area under the curve (AUC), and specificity scores of 90%, 95.7%, 0.97%, and 95.8%. The DL is based on the DarkNet model, which was utilized in [52] to classify COVID-19 binary and multi-class cases. The classification model is utilized in this model for the YOLO real-time object detection system. For binary classes and multi-class cases, the classification accuracy rate was 98.08% and 87.02%, respectively. [53] uses an ANN to diagnose COVID-19 in a quick and accurate approach. For the identification of COVID-19 cases from non-COVID-19 cases, they utilized and evaluated 10 CNNs. The sensitivity, AUC, accuracy and specificity of ResNet-101 were 100%, 0.994%, 99.51%, and 99.02%, respectively. The sensitivity, AUC, accuracy, and specificity for the Xception model have been 98.04%, 0.994%, and 99.02%, and 100%, respectively. Abdu Gumaei et. al. [45], had presented hybrid feature extraction approach that is based upon the regularized extreme learning machine. They used brain MRI images for the evaluating of their proposed method. For the preprocessing and enhancement of the contrast, the Min-Max was used. In their study, 94.233% is obtained from the experiments.

Kaplan et. al. presented a method based upon modified LBP for feature extraction from MRI brain images [46]. In their method, at first, the filtering is used to normalize and smooth the image, and then use the different modified LBP methods called nLBP and α LBP that the nLBP

method was obtained in high accuracy 94.56%. For classification, they used ML such as ANN, RF, K-NN, and decision tree.

The liver and brain tumor disease classification based on the CNN, discrete wavelet transform, and LSTM is proposed in [47]. In their work, the dataset is used from Firat university that contains 56 benign and 56 malignant images. The accuracy that they obtained for the liver tumor and brain tumor was 98.60% and 99.10% respectively.

Swati et. al. [29], used transfer learning and fine-tuning for MRI classification. They have utilized the pre-trained CNN model and the new fine-tuning scenario that was dependent on transfer learning. They achieved 94.82% from 5-fold cross-validation number.

The hybrid method based on the CNN and neural autoregressive distribution estimation used to brain tumor detection in MRI images [48].

One of the interesting behaviors in nature for finding food, which has a high intelligence nature, is the behavior of ants. Ants have a clever way of finding the food that can be used to find optimal and to reduce their output error in finding the shortest way. In this thesis, the behavior of ants is first introduced and a modified ant colony optimization for feature selection approach is introduced based upon it. In the second phase, the deep learning and autoencoder-based method are used for classifying the COVID-19 disease and brain tumors data [31].

The ant colony optimization algorithm has been suggested in 1990 by Marco Dorigo as an innovative method for solving hybrid optimization problems [31]. This algorithm is derived from actual behavior of ants to find food in the shortest path [36]. Every one of the ants leaves a chemical material that called pheromone in the path of finding food, which other ants choose the shortest path using the previous pheromones. This algorithm is very useful for solving non-deterministic polynomial (NP) problems and is used in problems such as itinerant vendor, scheduling problem, vehicle routing problem, etc. [37].

To solve any NP problem using the ant colony algorithm, the following must be specified:

- a. First, you have to turn the problem into a graph, including nodes and edges.
- b. See distance nodes (η) are raised and specified.
- c. A possible solution is created according to the problem.

- d. The pheromone update rule is used to determine the effective edges in achieving the best answer.
- e. The probabilistic transition rule is used to find the next node [38].

Several different implementations of ant colony algorithms such as Ant System, Max-Min Ant System, and Ant Colony System have been proposed, main difference of these methods is in the formula of pheromone update [38].



3. MATERIAL AND METHOD

3.1 COVID-19 DATASET

COVID-19 (coronavirus disease 2019) is a coronavirus strain that causes the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The initial instances were discovered in late December 2019 in Wuhan, China, before spreading worldwide [49]. On March 11, 2020, the WHO declared the current epidemic to be a pandemic. The dataset contains 6432 x-ray images in total [50]; image sizes vary and are not fixed; all images have been modified. This dataset is organized into 80% for the train of total images and the remainder for the test dataset, and it consists of three classes: COVID-19, pneumonia, and normal.

These images are divided as representing (460 COVID-19), (3418 pneumonia), and (1266 normal) for training the model. Also, (116 COVID-19), (855 pneumonia), and (317 normal) to test the model. Figure 3.1 shows a sample from each class for the COVID-19 dataset.



Figure 3. 1. Three classes of COVID-19 dataset; (1) COVID-19, (2) Pneumonia, and (3) Normal

The main reason for implementing these two datasets was that they included a large number of images and enough sizes to train every deep learning technique. Furthermore, a significant number of related papers have used these datasets to create an accurate comparison with different classification techniques. In addition, COVID-19 prompted us to evaluate the necessity for a quality detection system for MRI and X-ray images in order to limit the virus's risk.

Preprocessing of image datasets is an important step that improves the outcome of the image analysis and the quality of data feature extraction. To enhance the quality and reduce the time consuming for training, we use the created training image data-store and convert the images from RGB scale to grayscale in the range of [0, 255]. Furthermore, to solve the problem of the unstable size of our dataset sizes, we reduce the size of the two medical original datasets and became

unified, [227 227] the size of images for pre-trained (Alexnet) convolutional neural network and [224 224] for the else of the used pre-trained CNN, as the input for these networks.

As for the auto-encoder, it has been reduced to be [64 64] for both of these two medical datasets to address issues such as training time, training size dataset instability, and the out-of-memory problem that occurred. After changing the color scale and size of images, for the Auto-encoder model, we convert the input training images from the matrix to a single array (vector) without losing any features and enhancing the quality of the training model and learnable classifiers.

Neural network structure:

There are several convoluted neural network structures today that are successful in challenging topics such as the large-scale visual recognition challenge [30] and have therefore become very popular in connection with deep learning.

Among popular structures such as AlexNet [31] and GoogleNet [32] specifically, U-Net charging is a semantic separation task while being selected for the VGG-16 category.

U-Net:

U-Net [33] has a ConvNet structure designed by Ranberger et al. From the University of Freiburg to solve biomedical image resolution problems.

The network structure is interpreted in the figure 3.2. The network integrates a convolutional network structure (left conventional path) with a deconvolution structure (wide right path) to achieve semantic separation. The convolution network consists of an iterative pattern of two 3 کان 3 convolution operators. It is followed by the ReLU layer and the down-sampling process (creating a smaller image than the original image) is done through the 2×2 Max Polling operation with Stride 2. On the other hand, the deconvolution structure includes an up-sampling operation that is obtained from the map features along the contract path. It is followed by a 2 2 deconvolution that divides the map features into two parts. A combination of feature map results and what is obtained from the conventional path is required, followed by 3 3 convolution and the ReLU layer. The whole grid has 23 layers of deep convolution, the last layer of which shows the map of each combination of attribute vectors and is used as a class number.

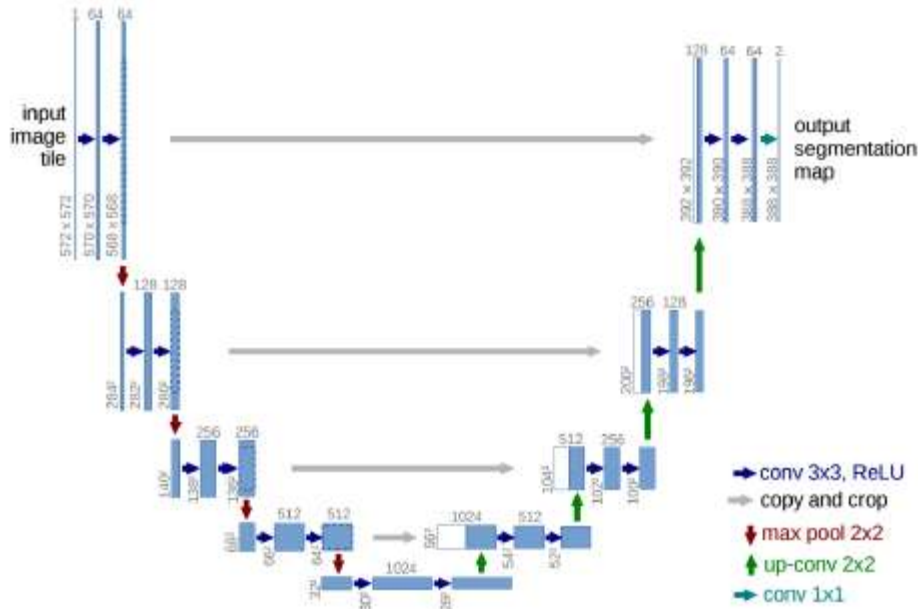


Figure 3. 2. U-Net structure (sample 32×32 pixels with lower resolution) [34]

The structure accepts RGB input images and equivalent binary masks. The training operation ends with the training of weights by the random gradient in Keras [35].

VGG-16:

The use of VGGNet structure in convolutional neural networks is very valid and common [36]. The reason for the popularity of this ConvNet is the excellent performance on the ImageNet dataset [37]. This structure has different variables including 16 and 19 weighting layers. In this study, the structure of VGG-16 has been selected because it is suitable for other datasets.

The input layer of the network is RGB image with 224×224 pixels. The input image passes through 5 convolutional blocks. Small convolution filters with 3×3 receiver field are used. Each convolutional block has a two-dimensional convolution layer (the number of filters varies between blocks). All hidden layers, such as the activation function layer (nonlinear operator), are equipped with ReLU and use space pooling using a Max Pooling layer. The network results in a classification block with three fully connected (FC) layers.

VGG16 should meet our needs in the following order:

The final fully connected output layer performs a binary classification (malignant or benign), not 1000 different classes.

The activation function in the modified layer changes from Softmax to Sigmoidal.

Preprocessing:

This project improves ConvNet features due to input preprocessing: Several processing techniques are required. Some preprocessing forms are done:

Medium Subtraction: In order to center the RGB values of the input data close to zero image dimensions, an average subtraction is used according to the image characteristics.

Normalize the image: By dividing each RGB dimension of the input images by standard segmentation, normalize from the original pixel values (0 to 255) to the normalized values of 1 and 0.

Image resizing and cropping: Input images are pre-processed by the structure to be acceptable. So the image is cropped to the required size and the original image is resized to 64×80 pixels for U-Net and to 224×224 pixels for VGG-16.

Teaching methods:

To address the two tasks of this project, similar teaching methods were proposed. The only difference between classification tasks and separability is used in the model. The U-Net structure is a good structure for resolution, while the VGG-16 is for the classification model. According to the diagram, the training data is trained through defined learning algorithms with each model that uses random gradients. Once the model learns the weights, the forecasting algorithm categorizes the different data according to the instruction. By comparing the predicted values with the actual data, the final model is evaluated. The Teaching methods is shown in figure 3.3.

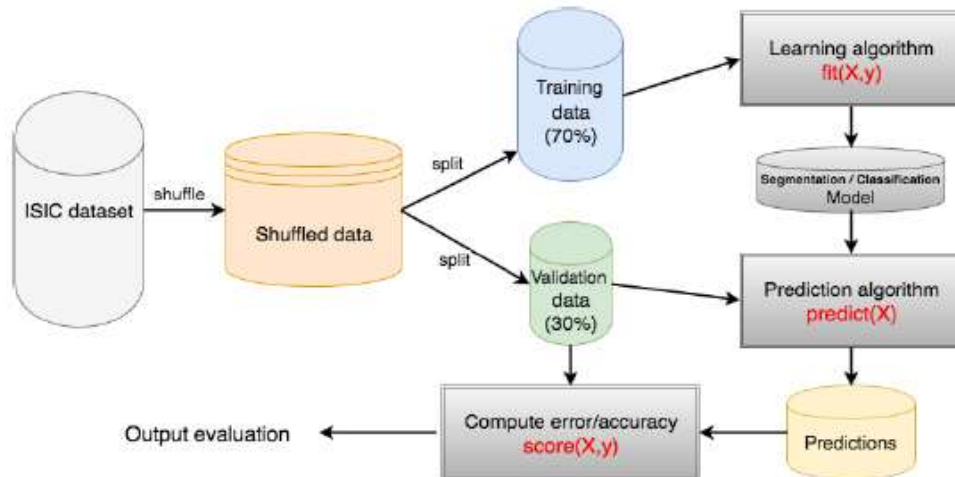


Figure 3. 3. Teaching methods [39]

Define hyper parameters:

During training, in order to create the best performance in the network and solve problems, some parameters should be considered as alternatives. Common ConvNet parameters are as follows:

Batch Size: Batch size is an attribute created in a number of images taught in the forward or backward direction. It is worth noting that the larger batch size requires more memory.

Terms: The number of terms is the number of forward or backward paths: Each path uses a number of image categories.

Period: The number of courses expresses the number of times each image was viewed during the training (for example, a period means that each image is viewed only once). The course can also be defined as a one-way route and a return route for all samples. This value is obtained numerically through the following equation:

$$epochs = \frac{batch\ size * iterations}{training\ images}$$

Loss function: The loss function, also called the cost function, evaluates the difference between the predicted values and the actual values in each category.

Learning rate: The learning rate parameter defines the step size for each weight of a model according to a random gradient.

Corruption: Corruption weight is an additional weighting update parameter that converts corrupt weights to zero at the end of the upgrade.

Optimizer: In order to find the most optimal set of hyper parameters for each model, the Keras framework provides an optimizer [33]. Some examples of optimizers are SGD, RMSprop and Adam.

Resolution:

The general structure of the separability task is described in the figure. Training data is processed from a separable dataset with a Python script to load masks and images and convert them to NumPy binaries (.npy). In this case, faster loading occurs during prediction and learning algorithms. The data is pre-organized based on training, training masks and test folders.

The proposed model of the U-Net network is taught. This means that weights are optimized by random propagation. The network output is a Numpy array that contains 379 experimental binary masks that can be converted to a 64 x 80 pixel JPG image, which means that previous processing techniques are required to enlarge the image.

Because of the reduced size of the COVID-19 dataset, the proposed approach starts the model with weights from the VGG-16 that are trained on a larger dataset (such as ImageNet). This process is known as transfer learning. The hypothesis behind transfer learning is that the pre-trained model has trained features that are useful for categorization.

Another common transfer learning technique involves not only retraining categories at the top of the network with a new data set, but also fine-tuning the grid layers to fine-tune the network and continue publishing. In this project, it is proposed to keep the lower level layers of the network constant because these layers have more general features of the data set. The desire to train the upper layers of the network stems from the ability to extract more specific features. For this reason, four layers of basic convolution are set up in the original VGG-16 structure with weights from the ImageNet dataset. The fifth and final convolution block is set up with the weights stored during feature extraction.

Due to the inherent reduction of the COVID-19 data set, it is observed that VGG is overfitting due to the small number of images. To avoid this, the VGG-16 structure has been chosen over the more up-to-date VGG-19 structure. This choice is due to reduced network capacity. It should be noted that the parameters to be calculated are reduced from 144 M (in VGG-19) to 138 M (in VGG-16) [34]. Overfitting can also be reduced by adding dropout. Dropout is a network regulation technique that reduces coordination between neurons by increasing repetition for convergence.

Because the COVID-19 dataset is inherently unbalanced (approximately 80% of responses are benign and the rest are malignant), it must be ensured that the network teaches the same number of images for each class. Instead of manually balancing the data set through Down Sampling, which further reduces the small size of the data set, a weighted loss function [40] was proposed to perform the small-class balancing operation.

$$H(p, q) = \sum_x \alpha(x) p(x) \log(q(x)), \quad \alpha(x) = \begin{cases} \rho & x = \text{major class} \\ 1 & \text{otherwise} \end{cases}$$

In this regard, p and q are the probability distributions of real values, and ρ is the weight loss value that protects the main class samples during training. ρ is defined as an apparent frequency (partial class).

The processes described in this section are repeated for each classification task in this project: (1) unchanged classification datasets, (2) fully segregated classification datasets, and (3) segregated classification datasets Automatically

suggested method

In general, this research was conducted in 4 stages as follows:

- a. The required data was extracted through the data set
- b. Using Excel features, the editing of the data obtained from the previous step was modified and the variables that had unreal values were removed from the list.
- c. Then, using the CORREL function in Excel software, the dependence between input and output variables was investigated so that the output signals that have the least dependence on the input signals were identified and removed from the total data.
- d. At this stage, the obtained information was categorized as follows for use in MATLAB software
- e. 70% of the data prepared in the previous steps were used as training data. This information was saved in the form of two Excel files called TrainX.xls and TrainY.xls and was ready for use in the next steps.
- f. 30% of the data were taken into consideration for use as validation data and were ready to be used in the next steps in the form of two files called TestX.xls and TestY.xls.
- g. Now it is time to enter the information created in the previous step into MATLAB software, so first the input data must be entered into the software in matrix format. By selecting the Import Data option from the HOME menu, the information in the Excel files

o TrainX.xls

TrainY.xls

o TestX.xls

o TestY.xls

Entered MATLAB software. It is emphasized here that the above names have been chosen arbitrarily

The second step is to design the system. This stage itself is divided into four parts

Part 1: Determining the parameters required to set up the initial FIS system and create it, the list of these parameters is as follows:

- a. Determining the number of membership functions, of which 5 were determined in this project
- b. Select the type of membership function that gaussmf type was used in this project
- c. Select the type of output function that was used in this project of linear type
- d. It is also necessary to select the type of FIS function, in which the sugeno function was used

The second part is determining how to train in the system. To do this, we need to define four parameters in the program as follows

- I. The maximum number of fault tolerances, which was considered 100 in this project
- II. Determining the amount of error that if we reach it, the training process will stop and in this project 0.001 was considered
- III. Determine the value of the initial step size, the value of which was considered 0.01
- IV. Determining the coefficient of decrease or increase of step size, which was considered 0.9 and 1.1 in this project, respectively

At the end of this section, by calling the Trainoption () function, the values of the above parameters were entered into the program as a structure.

Part 3: In this part, settings related to the screen are made while running the program. In this project, the four parameters related to the screen settings are as follows

DisplayInfo: Displays the system specifications on the screen.

DisplayError: Displays the amount of error on the screen at each step of the program

DisplayStepSize: Displays the step size at each step of the program on the screen

DisplayFinalResult: Displays the output value of the system or the final value

By calling the Displayoptions () function, the above values are entered into the software.

Part 4: Determining the optimization method or the same method used for training. In MATLAB software, can have two options for the type of training, which are the error retransmission method and hybrid, which in this project, the error retransmission type was used.

The following text contains the instructions that have been written for this project and includes all the settings mentioned above.



4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 BACKGROUND

The computer-aided analysis of MRI and X-ray analysis leads to gain in processing such as time, cost, and quality. Exactly in last year, the Covid-19 pandemic leads to an increase in the number of X-ray images that are requested to detect the Covid-19 in all healthcare institutions. This lead that the number of doctors, physicians, and expertise becomes less to deal with this case. Furthermore, in most hospitals, the expertise that can deal with these cases is not found. This concludes to find new alternatives speedy can deal with this problem. Then, researchers developed many machine learning, data mining, and deep learning applications to deal with this problem. One of the technical problems faced by researchers is the high computation time in convolutional neural networks. The extracted features from pre-trained models lead to a fixed number of features which a number of these features unaffected most of the time. The computer-based recognition systems also presented high advantages in other medical image recognition problems such as brain tumor classification, Pneumonia recognition etc....

4.2 METHOD

In this study, a combination of pre-trained convolutional neural network with Energy spectral density and some types of classifiers for image classification. In the first stage, the pre-trained model applied for extracting important features from the image. Several pre-trained techniques are applied, one of these techniques is the pre-trained AlexNet is a common model applied in several image classification studies. The aim of this stage is to extract high level and sensitive features from input images. This model realizes human vision technology. AlexNet consists of several layers which are a convolution layer, pooling layer, and a fully connected layer. This structure contains the main functions of a convolutional neural network model. Nowadays the use of CNN is widely in several computer vision problems because it implements the shared weight technique to reduce the computation rather than the fully connected technique. The extracted features differ from one model to another:

- a. AlexNet (4096) features.
- b. GoogLeNet (1000) features.

- c. Squeezenet (1000) features.
- d. Vgg19 (1000) features.

The convolutional layer is generally the first layer in CNN. A convolution layer has the main advantage which is weight sharing. Weight sharing happens across the receptive field of the neurons (filters) in a particular layer. Weights are the numbers within each filter. This step mathematically can be represented as shown in Eq. (1):

$$X_m^L = f(\sum_n x_m^{L-1} * W_{mn}^L) + b_n^L \quad (1)$$

M and n are the dimensions of the convolution filter, L represented the layer, x represented the features, W represent the weight, and b represented the basis.

where * denotes the convolutional procedure and f represents the activation function which can be Hyperbolic Tangent (tanh), Sigmoid, Rectified Linear Unit (ReLU). The activation functions can be changed according to the data type.

Moreover, the extracted features by the convolution layer wired to the next layer (pooling layer). The pooling layers generally medium layers between convolution and fully connected layers. The pooling layer decreases the number of training parameters and computation cost and control overfitting. The output of pooling layer can be represented mathematically as shown in Eq. (2):

$$x_n^L = \text{down}(x_m^{L-1}) \quad (2)$$

where down (·) is a kind of pooling operation and can be Max pooling, Average pooling, etc.

After this operation completes, the extracted features from several convolution and pooling layers wired to the fully connected layers. Which is simply feed-forward neural networks. Also, it is the last few layers in the network structure of CNN. The output of fully connected layer can be represented as shown in Eq. (3):

$$(y = 1|x; w) = \frac{1}{1 + \exp(-w^T x)} \quad (3)$$

where y represents the labels of the data, $x \in R^{(K+1) \times 1}$ indicates feature vector in K- dimension, $w \in R^{(K+1) \times 1}$ indicates the parameters of the weight vector.

The main contribution of this work is to apply ESD on the model working as a feature extraction technique, the aim of this step is to reduce the size of the features that are extracted by using a Pre-

trained model (previous stage) by using a single and simple function. This will lead to solve the problem that occurs in most pre-trained models that sometime the extracted features become higher size and most of these features not effective. This step represented as shown in Eq. (4):

$$ESD \int_{-\infty}^{\infty} x(f)^2 df \quad (4)$$

ESD refers to the total energy of features $x(f)$, $-\infty$ and ∞ limits the upper and lower index of each period features in the features vector. The period values may become an integer value such as 2,3,4,5...etc. The period values are estimated by employing a trial-and-error approach which facilitates obtaining the best features to produce the highest accuracy. Then, the ESD calculated for each period is collected in a new vector and represented the extracted features by ESD function and wired to the classifiers.

In the last stage, applying some types of classifiers to classify the extracted features and evaluate the combination model depending on the labels of input type.

The dataset divided into train and test sets, the ratios of train and test sections also differ from one application to another. The classifiers trained in a supervised fashion to learn the weights of each label and testing to calculate the learning rate. The flowchart of the proposed method presented in Figure 4.1.

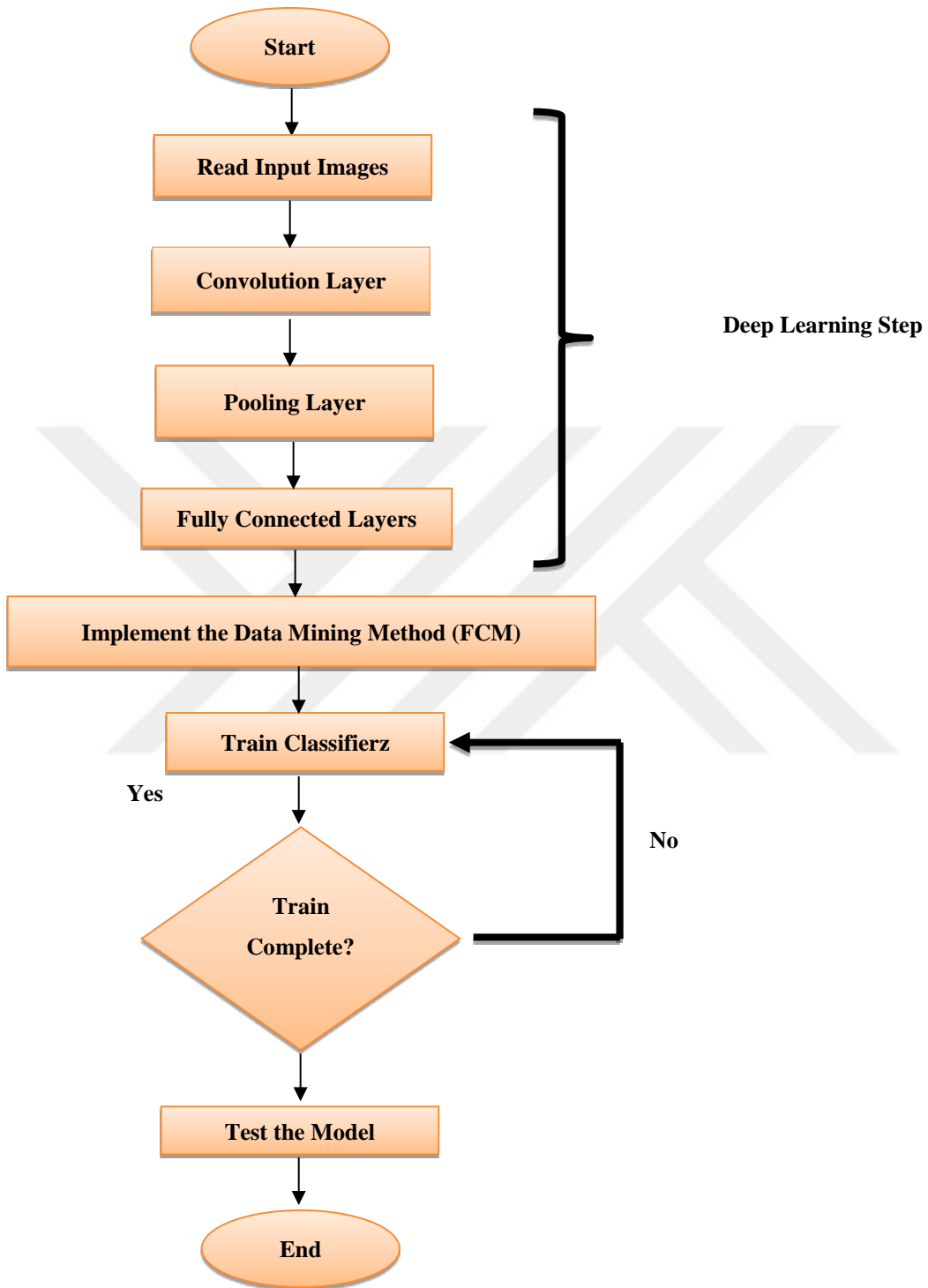


Figure 4. 1: Proposed Framework

4.3 RESULTS

In this study, MATLAB2020b used to execute the proposed method on Intel(R) Core (TM) i7-6500U CPU @ 2.50GHz 2.60 GHz. Evaluating the performance of this study, four pre-trained convolutional neural networks were proposed. Moreover, four types of classifiers applied to three different medical datasets. Furthermore, the confusion matrix used to evaluate the presented method. The results can be seen as follows:

4.4 COVID-19 DETECTION DATASET

As shown in figure 4.2, the experimental results show that the PROPOSED METHOD presented remarkable results in both datasets [51]. The CNN extracted high level features with 4096 features and these features wired to the ESD. The ESD applied as feature extractor with 2 periods which mean 2048 features extracted this lead to reduce the computation time. The extracted features classified using (DT, SVM, KNN, and ensemble learning). The proposed method with SVM classifier presented best results compared with other classifiers and higher than these state-of-art studies.

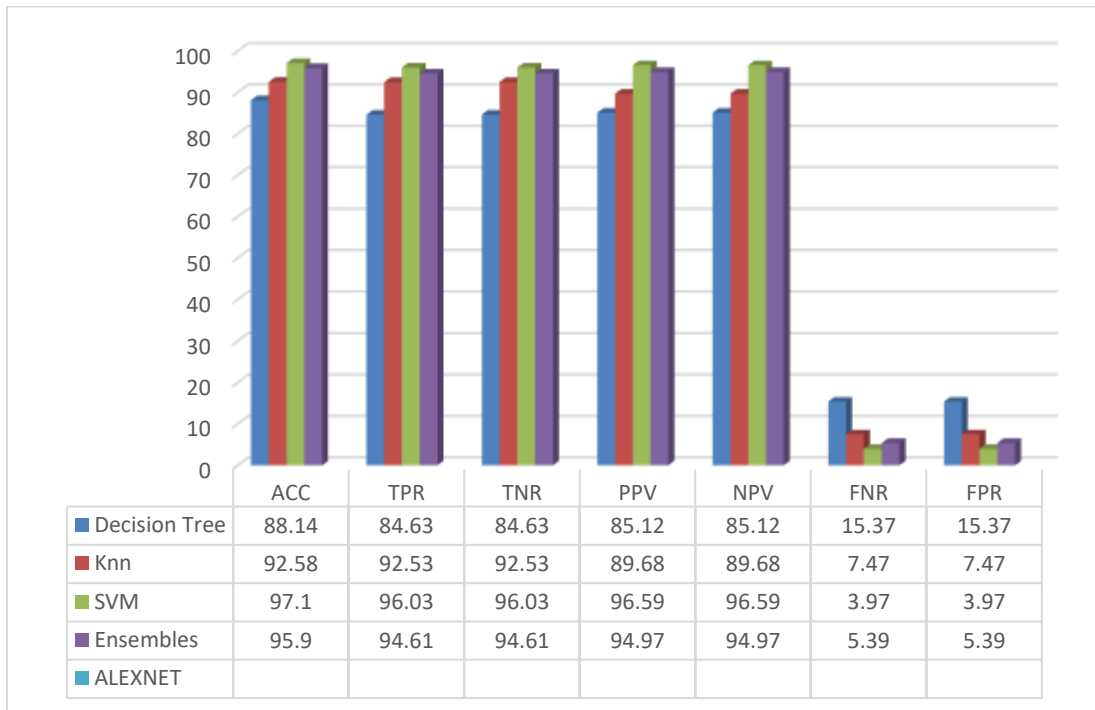


Figure 4. 2: a) COVID-19 dataset 1 using our method

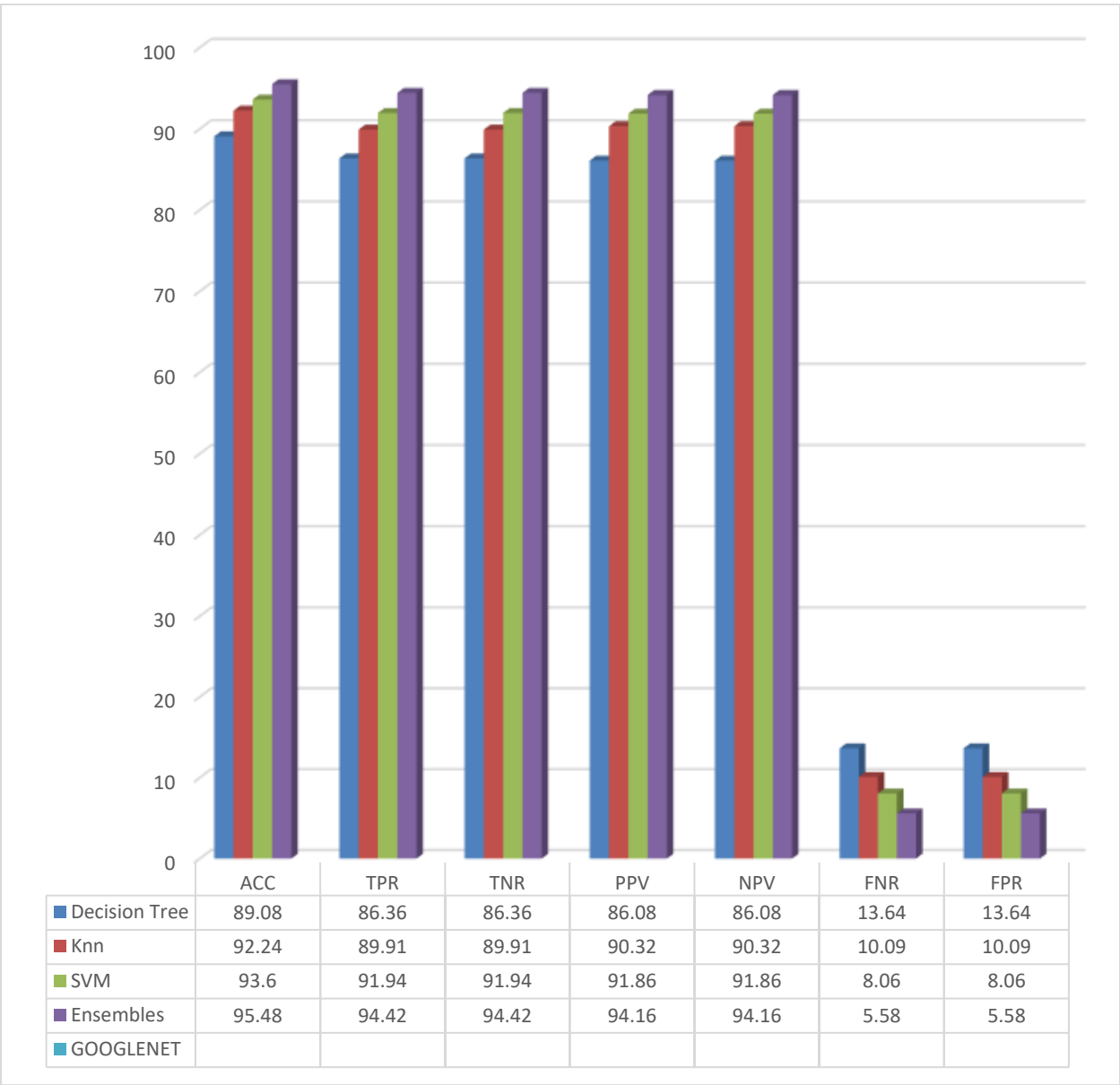


Figure 4. 3: b) COVID-19 dataset 1 using our method

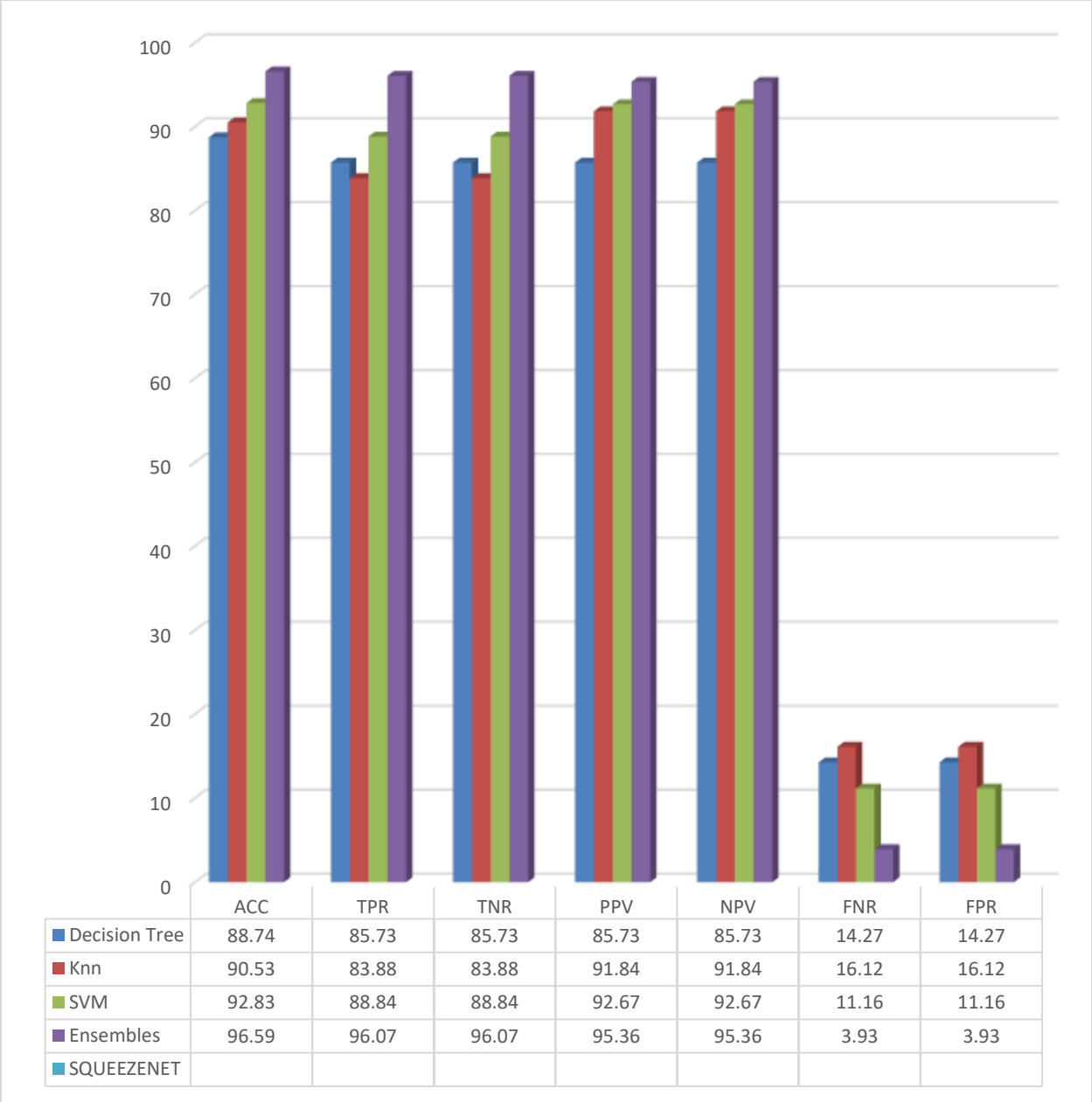


Figure 4. 4: COVID-19 c) dataset 1 using our method

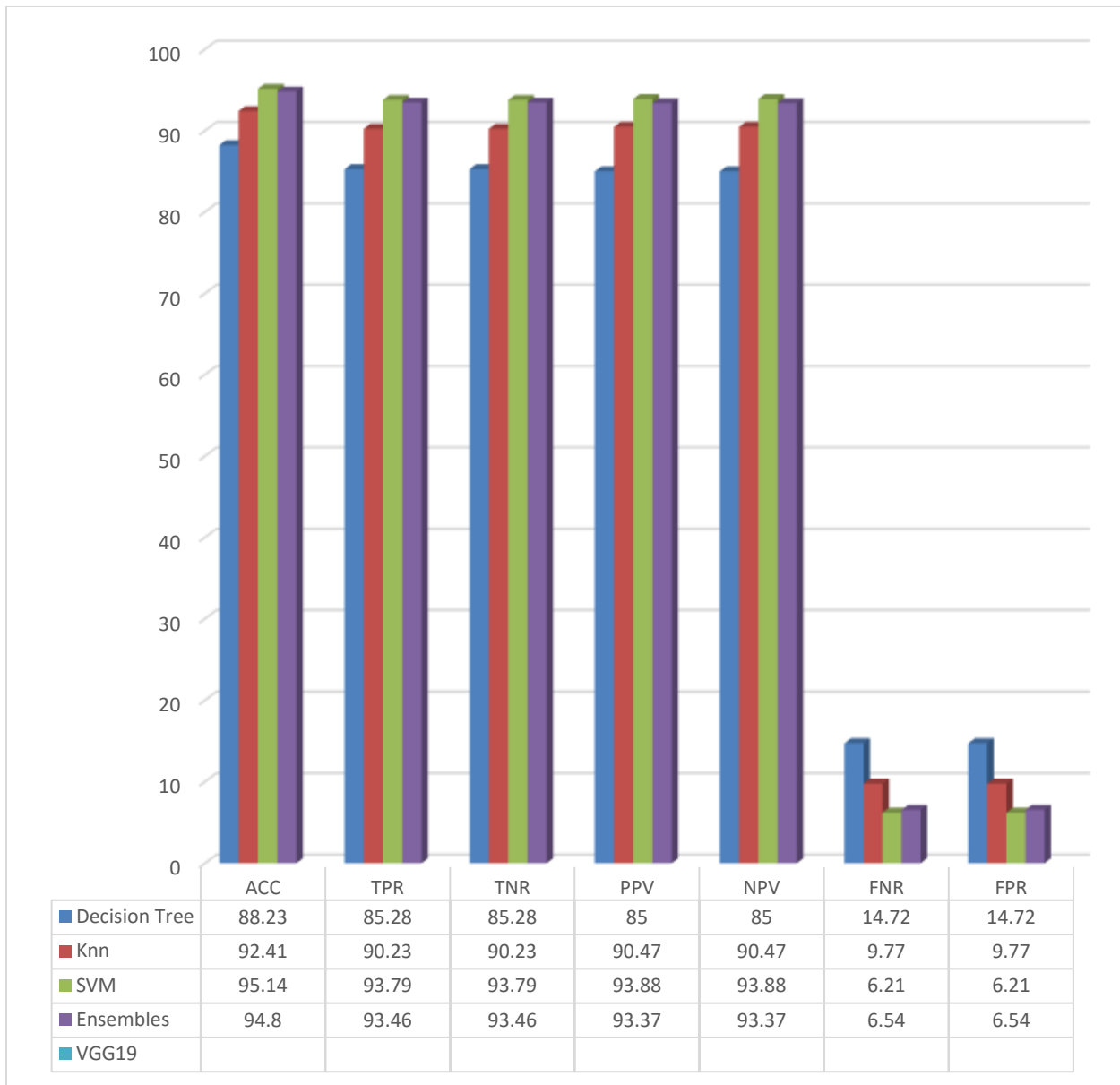


Figure 4. 5: COVID-19 dataset 1 using our method

Furthermore, the proposed method compared with several state of art studies presented to detect covid-19 by using machine learning and deep learning techniques as shown in Table 4.1.

Table 4. 1: Covid-19 dataset 1 results compressions

References	Method	Acc (%)
[32]	Tailored CNN	92.3
[33]	DenseNet	88.90
[34]	Capsule Networks	95.7
[35]	DarkNet-19 based CNN	87.02
Proposed method		95.88

4.5 COVID-19 DATASET 2

Pneumonia dataset [52] used to validate the proposed method. In this experiment, also proposed method presented best results than other classifiers. The presented method results shown in figure 4.3.

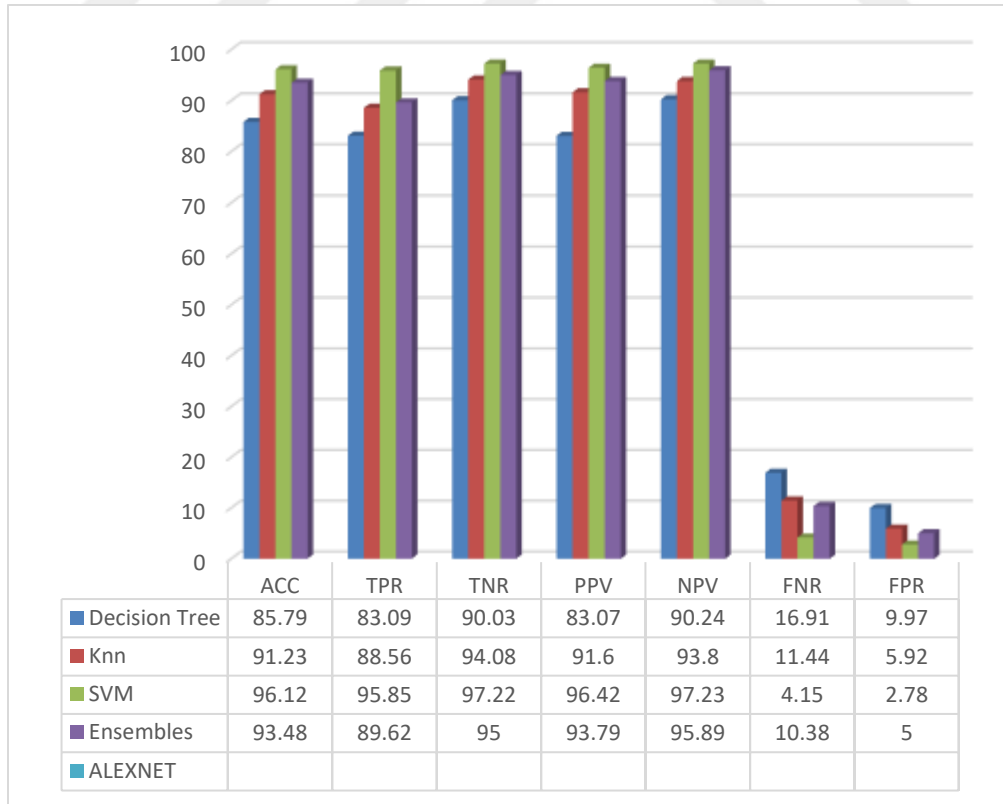


Figure 4. 6: a) COVID-19 dataset 2 using our method

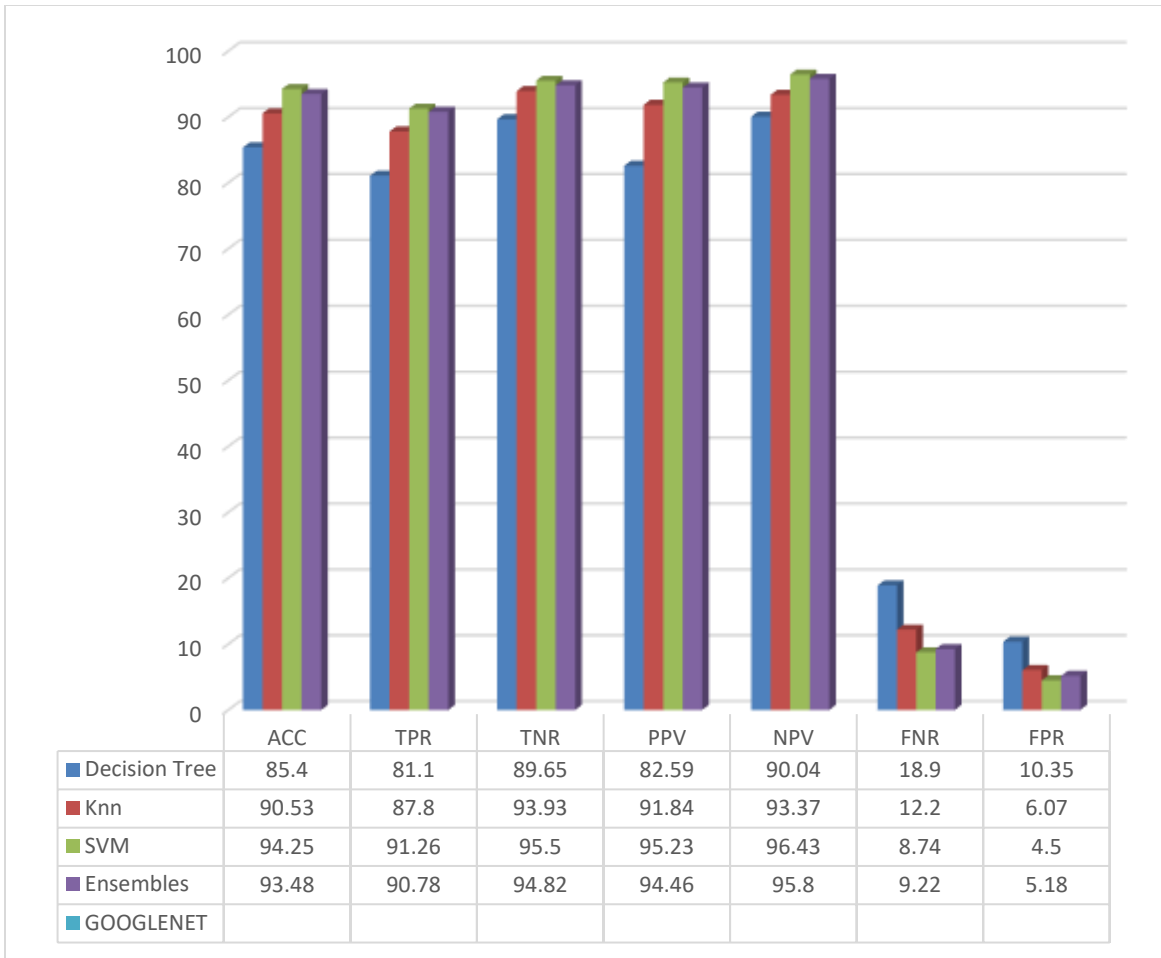


Figure 4. 7: b) COVID-19 dataset 2 using our method

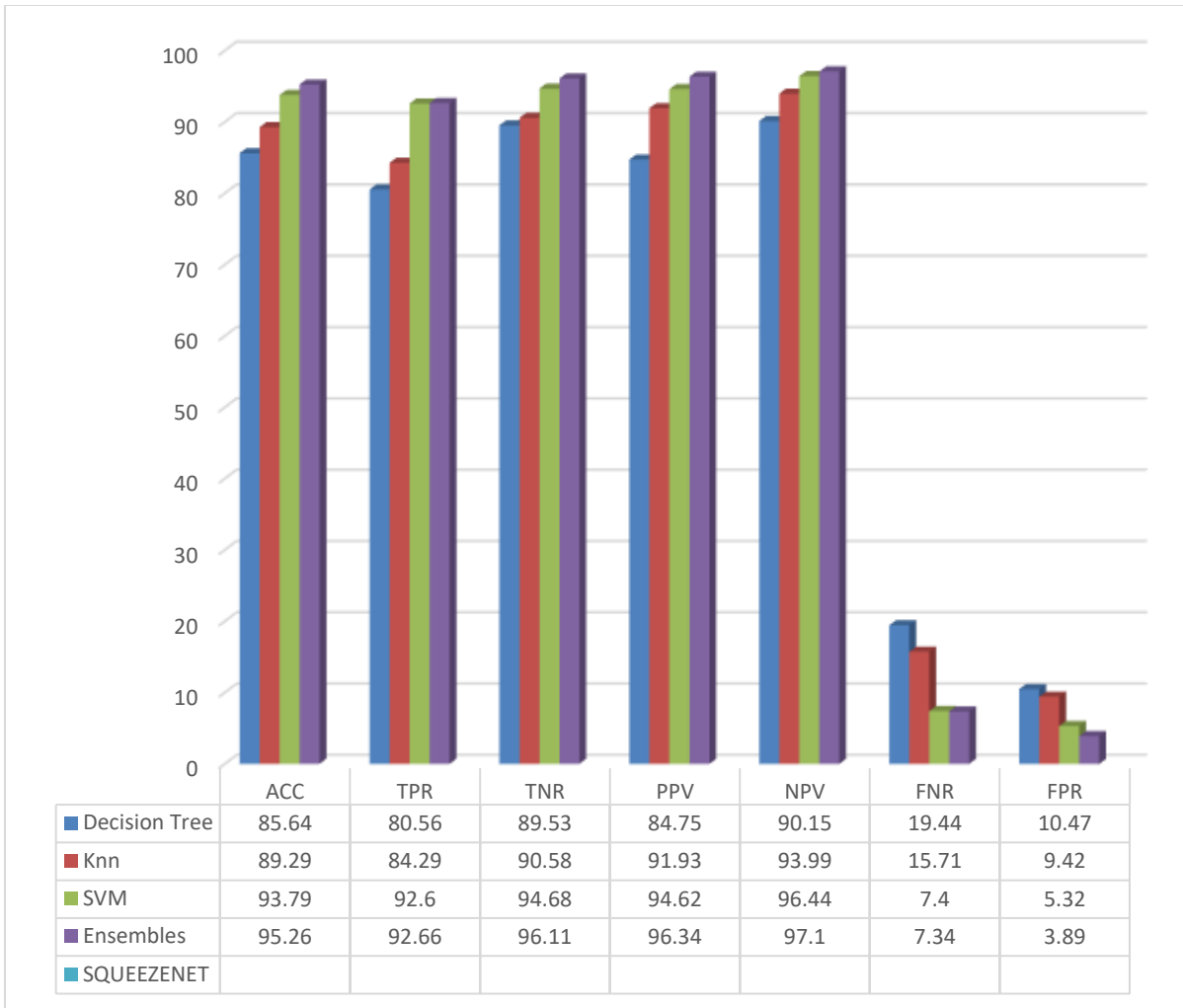


Figure 4. 8: c) COVID-19 dataset 2 using our method

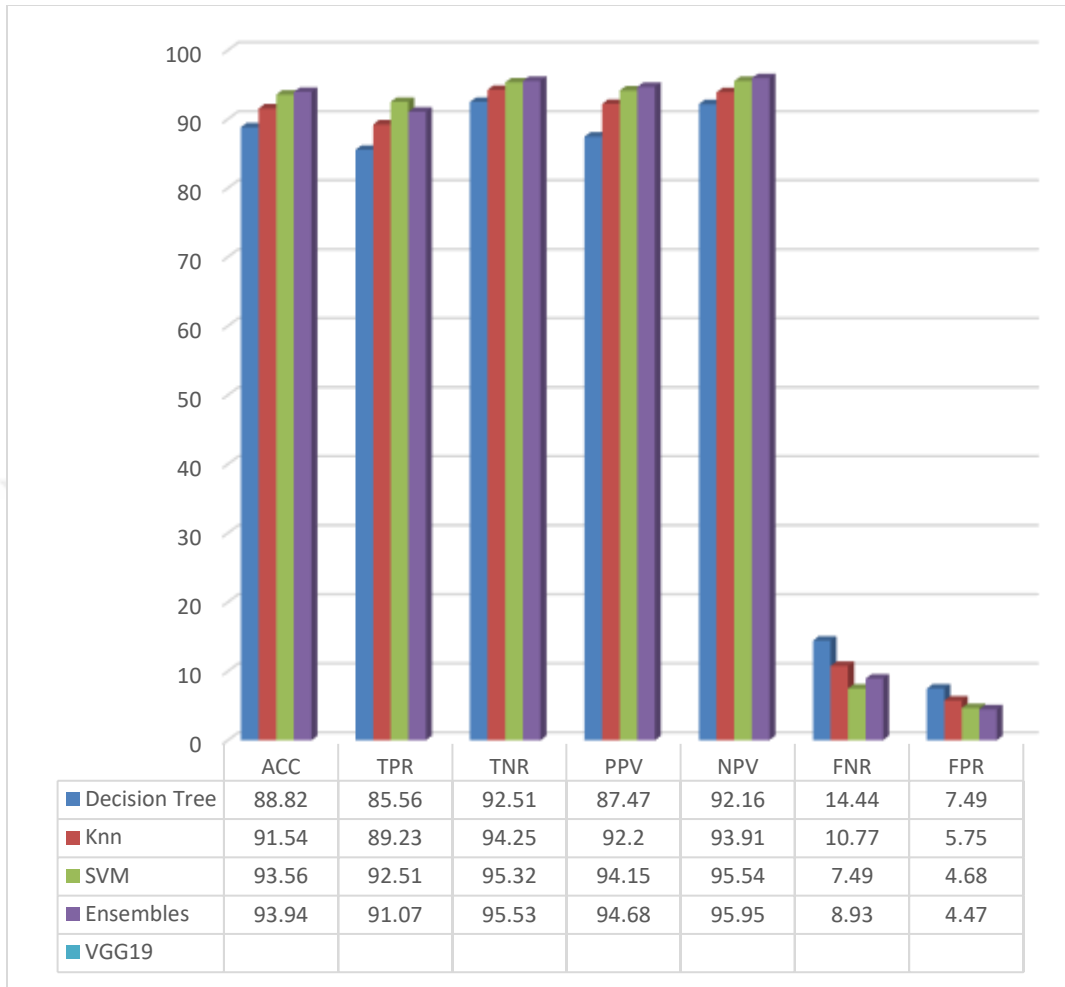


Figure 4. 9: COVID-19 dataset 2 using our method

Then, the proposed method compared with well-known research presented in this field and show that the proposed method is best. The COVID-19 dataset 2 results comparisons is shown in table 4.2.

Table 4. 2: COVID-19 dataset 2 results comparisons

References	Method	Acc (%)
[53]	CNN	80.4
[53]	CNN with Data Augmentation	93.4
[53]	VGG16	85.6
[54]	convolutional neural network model from scratch	93.73
[55]	VGG+ data augmentation + spatial transformer network (STN)	73.0
Proposed method		96.43

4.6 COVID-19 DATASET 3

In this section COVID-19 dataset 3 classification dataset [56] used to validate the proposed method. The COVID-19 dataset 3 results is shown in figure 4.4.

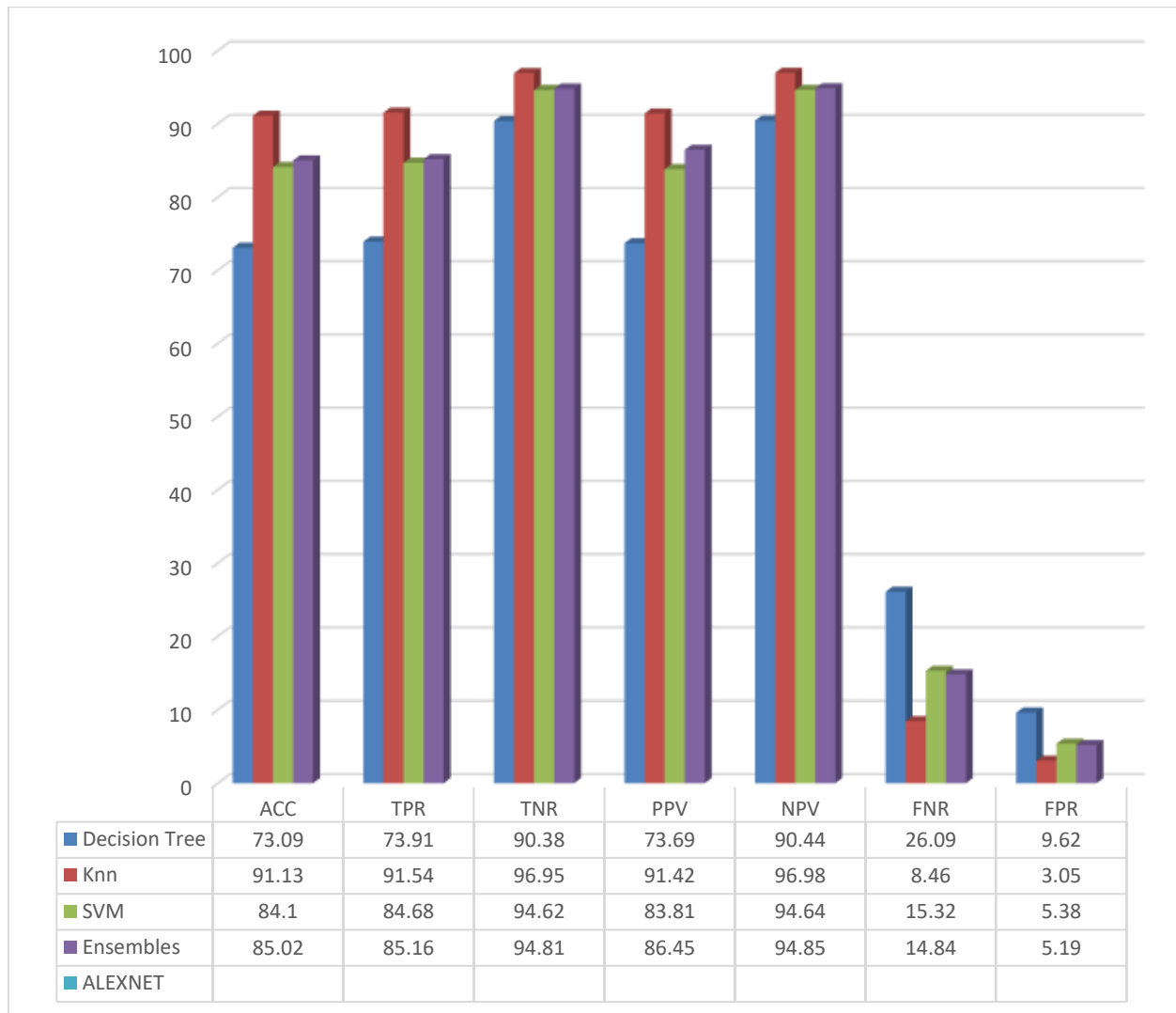


Figure 4. 10: a) COVID-19 dataset 3 results

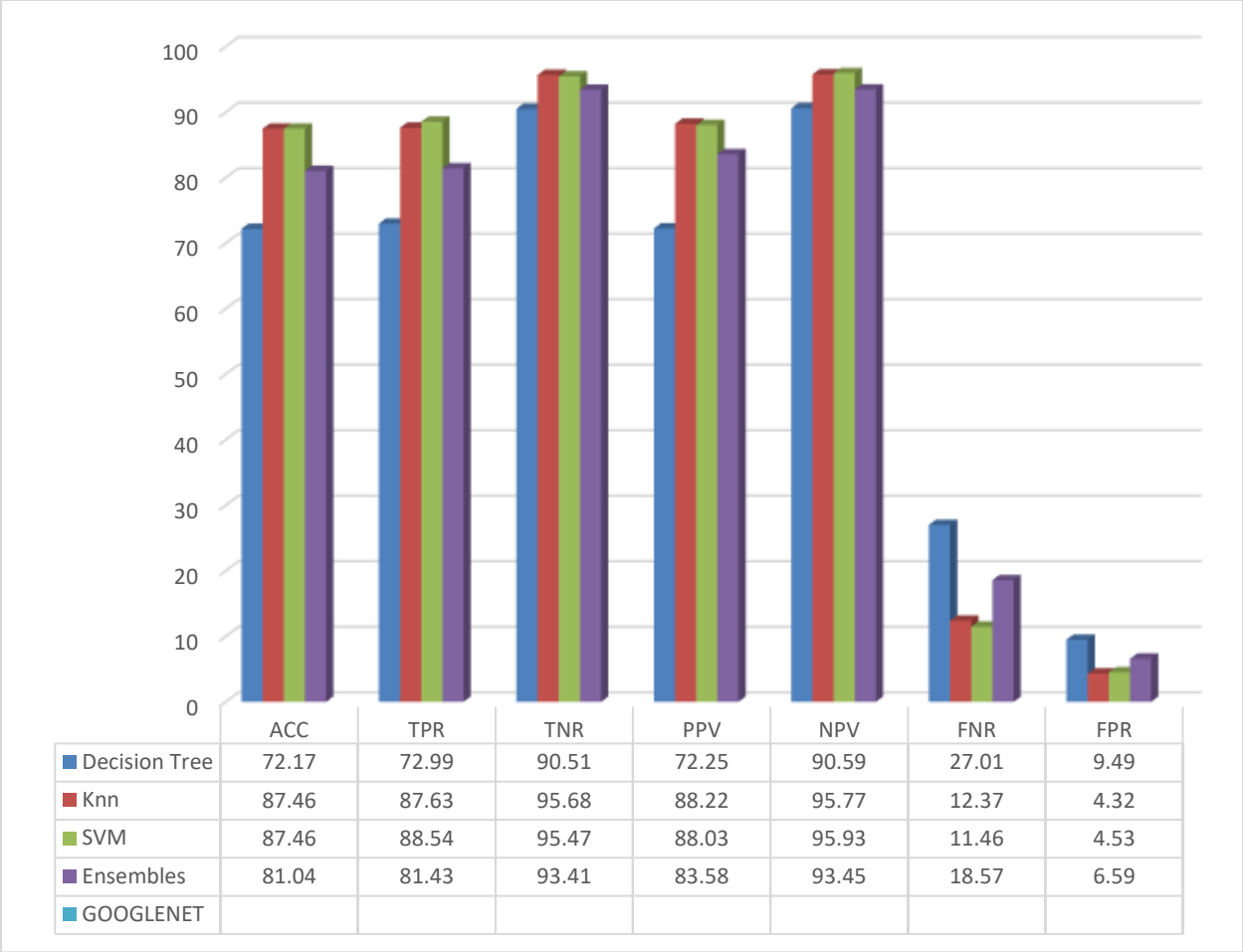


Figure 4. 11: b) COVID-19 dataset 3 results

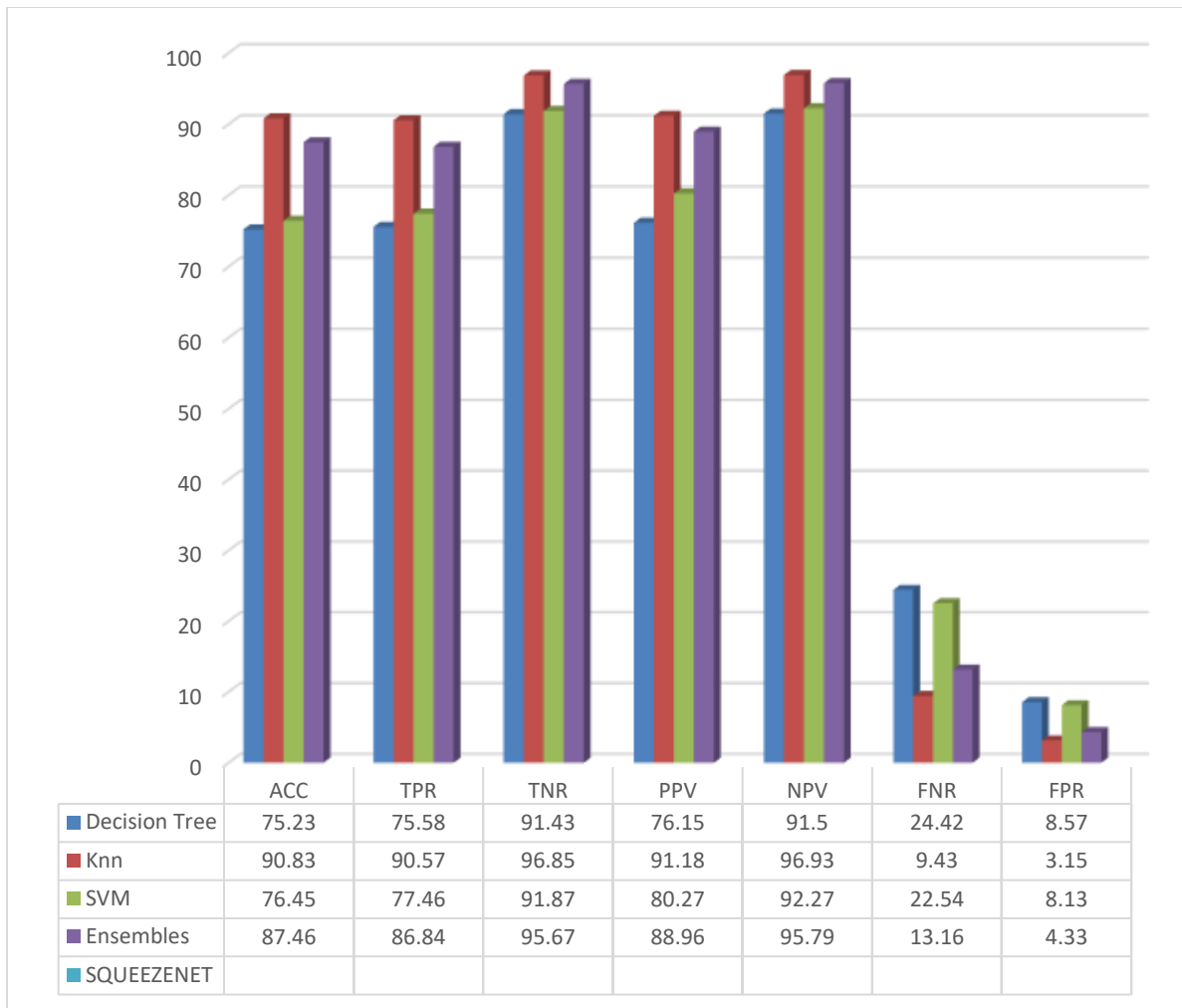


Figure 4. 12: c) COVID-19 dataset 3 results

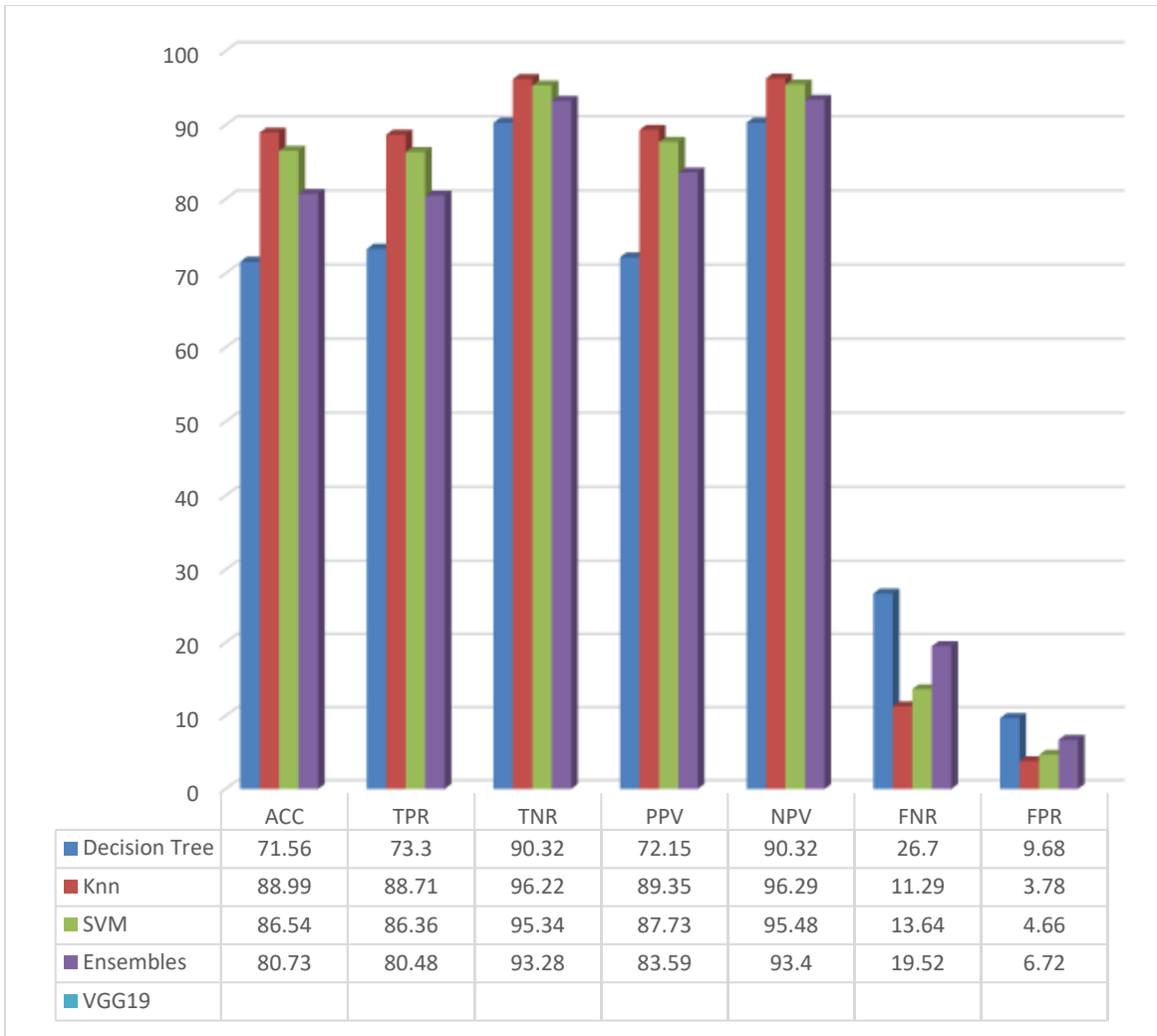


Figure 4. 13: COVID-19 dataset 3 results

In this section, the ESD calculated for each 4 period, then the output of the Alexnet reduced from 4096 to 1024. Then, KNN presented best results than other classifiers and compared with state of arts studies see Table 4.3.

Table 4. 3: COVID-19 dataset 3 results comparisons.

References	Method	Acc (%)
[57]	GLCM+CNN	82.27
[19]	Intensity Histogram+ CNN	87.54
[19]	GLCM + Bow + CNN	89.72
[43]	CNN	86.56
Proposed method		92.24

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

In this thesis, the different networks of deep learning are used to feature extraction of the different datasets from the medical imaging system. Within huge features from the image is not suitable for processing and sometimes the model gives mistake results, for this reason, the Meta-Heuristic method as Ant Colony Optimization method is used to reduce and select the best feature from the feature data. The ANN, SVM, Deep learning, and different types of deep learning networks are tested and evaluated. The best accuracy of 99.61% was obtained from the combination of the CNN, ACO, and SVM methods.

In terms of accuracy, the suggested method outperforms methods such as ANN, CNN, CNN with data fusion, Stacked-autoencoder, Capsule Networks, and DarkNet-19 based CNN for diagnosis of various medical disease datasets. Feature selection allows the training of a deep learning to focus on the most essential features, reducing the error rate in differentiating infected from healthy people. In the proposed method, based on the behavior of ants, a modified meta-heuristic algorithm is introduced and then it presented for finding the best and effective features.

The goal function minimizes the number of the features and find the optimal features. Analyzes show that the proposed method in classifying of different medical datasets have an accuracy index of 99.61% and 99.18% for COVID-19 data respectively. Experiments show that the ant optimization algorithm in finding the optimal features in the evaluation functions of the algorithm, has more accurate than other methods that mentioned in the results.

The analyzes showed that the proposed method was more accurate than the CNN, ANN, Random Forest and other methods. In future researchers, to increase its accuracy, a combination of other group intelligence algorithms and the PSO algorithm for feature selection will be used to diagnose patients.

5.2 FUTURE WORK

For Future work we can use th deep learning to recognize the covid 19 diseas.



REFERENCES

- [1] S.-A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data preprocessing in predictive data mining," *Knowl. Eng. Rev.*, vol. 34, 2019.
- [2] H. He and L. Harris, "The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy," *J. Bus. Res.*, vol. 116, pp. 176–182, 2020.
- [3] J. F. Sallis, D. Adlakha, A. Oyeyemi, and D. Salvo, "An international physical activity and public health research agenda to inform coronavirus disease-2019 policies and practices," *J. Sport Heal. Sci.*, vol. 9, no. 4, p. 328, 2020.
- [4] J. Hou, "Effective Ways for College Students' Mental Health Education Based on Music Therapy," *J. Healthc. Eng.*, vol. 2022, 2022.
- [5] M. Garouani, A. Ahmad, M. Bouneffa, M. Hamlich, G. Bourguin, and A. Lewandowski, "Towards big industrial data mining through explainable automated machine learning," *Int. J. Adv. Manuf. Technol.*, vol. 120, no. 1, pp. 1169–1188, 2022.
- [6] A. Ishaq *et al.*, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE access*, vol. 9, pp. 39707–39716, 2021.
- [7] M. Maiti, D. Vuković, A. Mukherjee, P. D. Paikarao, and J. K. Yadav, "Advanced data integration in banking, financial, and insurance software in the age of COVID-19," *Softw. Pract. Exp.*, vol. 52, no. 4, pp. 887–903, 2022.
- [8] A. Farki, Z. Salekshahrezaee, A. M. Tofigh, R. Ghanavati, B. Arandian, and A. Chapnevis, "COVID-19 Diagnosis Using Capsule Network and Fuzzy-Means and Mayfly Optimization Algorithm," *Biomed Res. Int.*, vol. 2021, 2021.
- [9] W. H. Organization, "Antigen-detection in the diagnosis of SARS-CoV-2 infection using rapid immunoassays," 2020.
- [10] G. Lippi, M. Plebani, and B. M. Henry, "Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis," *Clin. Chim. acta*, vol. 506, pp. 145–148, 2020.

- [11] Y. Wu *et al.*, “Nervous system involvement after infection with COVID-19 and other coronaviruses,” *Brain. Behav. Immun.*, vol. 87, pp. 18–22, 2020.
- [12] G. Xu *et al.*, “Clinical pathway for early diagnosis of COVID-19: updates from experience to evidence-based practice,” *Clin. Rev. Allergy Immunol.*, vol. 59, no. 1, pp. 89–100, 2020.
- [13] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, “Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review,” *Chaos, Solitons & Fractals*, vol. 139, p. 110059, 2020.
- [14] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, “Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks,” *Comput. Biol. Med.*, vol. 121, p. 103795, 2020.
- [15] J. A. Al-Tawfiq and Z. A. Memish, “Diagnosis of SARS-CoV-2 infection based on CT scan vs RT-PCR: reflecting on experience from MERS-CoV,” *J. Hosp. Infect.*, vol. 105, no. 2, pp. 154–155, 2020.
- [16] X.-S. Yang, Z. Cui, R. Xiao, A. H. Gandomi, and M. Karamanoglu, *Swarm intelligence and bio-inspired computation: theory and applications*. Newnes, 2013.
- [17] A. Barolli, S. Sakamoto, L. Barolli, and M. Takizawa, “Performance analysis of simulation system based on particle swarm optimization and distributed genetic algorithm for WMNs considering different distributions of mesh clients,” in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2018, pp. 32–45.
- [18] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, “Segmentation, feature extraction, and multiclass brain tumor classification,” *J. Digit. Imaging*, vol. 26, no. 6, pp. 1141–1150, 2013.
- [19] J. Cheng *et al.*, “Enhanced performance of brain tumor classification via tumor region augmentation and partition,” *PLoS One*, vol. 10, no. 10, p. e0140381, 2015.
- [20] N. Ghassemi, A. Shoeibi, and M. Rouhani, “Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images,”

Biomed. Signal Process. Control, vol. 57, p. 101678, 2020.

- [21] U. Özkaya, Ş. Öztürk, and M. Barstugan, “Coronavirus (covid-19) classification using deep features fusion and ranking technique,” in *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, Springer, 2020, pp. 281–295.
- [22] A. Mollalo, K. M. Rivera, and B. Vahedi, “Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, p. 4204, 2020.
- [23] P. K. Sethy and S. K. Behera, “Detection of coronavirus disease (covid-19) based on deep features,” 2020.
- [24] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, and H. Greenspan, “Coronavirus detection and analysis on chest ct with deep learning,” *arXiv Prepr. arXiv2004.02640*, 2020.
- [25] C. M. Yeşilkanat, “Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm,” *Chaos, Solitons & Fractals*, vol. 140, p. 110210, 2020.
- [26] D. Li, Z. Fu, and J. Xu, “Stacked-autoencoder-based model for COVID-19 diagnosis on CT images,” *Appl. Intell.*, vol. 51, no. 5, pp. 2805–2817, 2021.
- [27] F. Khoshbakhtian, A. B. Ashraf, and S. S. Khan, “Covidomaly: A deep convolutional autoencoder approach for detecting early cases of covid-19,” *arXiv Prepr. arXiv2010.02814*, 2020.
- [28] F. Khozeimeh *et al.*, “CNN AE: Convolution Neural Network combined with Autoencoder approach to detect survival chance of COVID 19 patients,” *arXiv Prepr. arXiv2104.08954*, 2021.
- [29] Z. N. K. Swati *et al.*, “Brain tumor classification for MR images using transfer learning and fine-tuning,” *Comput. Med. Imaging Graph.*, vol. 75, pp. 34–46, 2019.
- [30] D. Dong *et al.*, “The role of imaging in the detection and management of COVID-19: a review,” *IEEE Rev. Biomed. Eng.*, 2020.

- [31] M. Dorigo, M. Birattari, and T. Stutzle, “Ant colony optimization,” *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.
- [32] L. Wang, Z. Q. Lin, and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.
- [33] X. Li and D. Zhu, “Covid-xpert: An ai powered population screening of covid-19 cases using chest radiography images,” *arXiv Prepr. arXiv2004.03042*, 2020.
- [34] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, “Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images,” *Pattern Recognit. Lett.*, vol. 138, pp. 638–643, 2020.
- [35] T. Ozturk, M. Talu, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” *Comput. Biol. Med.*, vol. 121, p. 103792, 2020.
- [36] M. Dorigo and K. Socha, “An introduction to ant colony optimization,” in *Handbook of Approximation Algorithms and Metaheuristics, Second Edition*, Chapman and Hall/CRC, 2018, pp. 395–408.
- [37] M. Dorigo and C. Blum, “Ant colony optimization theory: A survey,” *Theor. Comput. Sci.*, vol. 344, no. 2–3, pp. 243–278, 2005.
- [38] H. R. Kanan, K. Faez, and S. M. Taheri, “Feature selection using ant colony optimization (ACO): a new method and comparative study in the application of face recognition system,” in *Industrial conference on data mining*, 2007, pp. 63–76.
- [39] M. R. Ismael and I. Abdel-Qader, “Brain tumor classification via statistical features and back-propagation neural network,” in *2018 IEEE international conference on electro/information technology (EIT)*, 2018, pp. 252–257.
- [40] P. Thejaswini, M. B. Bhat, and M. K. Prakash, “Detection and classification of tumour in brain MRI,” *Int. J. Eng. Manufact.(IJEM)*, vol. 9, no. 1, pp. 11–20, 2019.

- [41] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Comput. Biol. Med.*, vol. 111, p. 103345, 2019.
- [42] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019.
- [43] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3129–3133.
- [44] O. N. Belaid and M. Loudini, "Classification of Brain Tumor by Combination of Pre-Trained VGG16 CNN," *J. Inf. Technol. Manag.*, vol. 12, no. 2, pp. 13–25, 2020.
- [45] A. Gumaiei, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, "A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification," *IEEE Access*, vol. 7, pp. 36266–36273, 2019.
- [46] K. Kaplan, Y. Kaya, M. Kuncan, and H. M. Ertunç, "Brain tumor classification using modified local binary patterns (LBP) feature extraction methods," *Med. Hypotheses*, vol. 139, p. 109696, 2020.
- [47] H. Kutlu and E. Avcı, "A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks," *Sensors*, vol. 19, no. 9, p. 1992, 2019.
- [48] R. Hashemzahi, S. J. S. Mahdavi, M. Kheirabadi, and S. R. Kamel, "Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1225–1232, 2020.
- [49] H. Zhu, L. Wei, and P. Niu, "The novel coronavirus outbreak in Wuhan, China," *Glob. Heal. Res. policy*, vol. 5, no. 1, pp. 1–3, 2020.
- [50] A. C. Li, D. T. Lee, K. K. Misquitta, K. Uno, and S. Wald, "COVID-19 detection from chest radiographs using machine learning and convolutional neural networks," *medRxiv*, 2020.

- [51] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv Prepr. arXiv2006.11988*, 2020.
- [52] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest X-ray images for classification," *Mendeley data*, vol. 2, no. 2, 2018.
- [53] O. D. Gülgün and E. Hamza, "Classification performance comparisons of deep learning models in pneumonia diagnosis using chest x-ray images," *Turkish J. Eng.*, vol. 4, no. 3, pp. 129–141, 2020.
- [54] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthc. Eng.*, vol. 2019, 2019.
- [55] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Informatics Med. Unlocked*, vol. 20, p. 100391, 2020.
- [56] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain tumor classification (mri)," *Kaggle, doi*, vol. 10, 2020.
- [57] W. Widhiarso, Y. Yohannes, and C. Prakarsah, "Brain tumor classification using gray level co-occurrence matrix and convolutional neural network," *IJEIS (Indonesian J. Electron. Instrum. Syst.)*, vol. 8, no. 2, pp. 179–190, 2018.