SERVICE DISCOVER MECHANISM FOR THE DIAGNOSIS OF COVID-19 USING CONVOLUTIONAL NEURAL NETWORK AND HYBRID ML CLASSIFIERS

A Thesis

Submitted

In Partial Fulfillment of the Requirements

for the Degree of

Master of Technology

In

Computer Science & Engineering

Submitted by

FATIMA ALVI

Enroll No. 1600100725

Roll No. 2001621004

Under the Supervision of

Dr. FAIYAZ AHMAD

(Assistant Professor)

Department of Computer Science & Engineering

Faculty of Engineering

INTEGRAL UNIVERSITY, LUCKNOW, INDIA

June, 2022

CERTIFICATE

This is to certify that **Miss. Fatima Alvi** (Roll. No.-2001621004) has carried out the research work presented in the thesis titled **"Service Discover Mechanism for the Diagnosis of COVID-19 using Convolutional Neural Network and Hybrid ML Classifier"** submitted for partial fulfillment for the award of the Degree of Master of Technology in Computer Science & Engineering from **Integral University, Lucknow** under my supervision.

It is also certified that:

(i) This thesis embodies the original work of the candidate and has not been earlier submitted elsewhere for the award of any degree/diploma/certificate.

(ii) The candidate has worked under my supervision for the prescribed period.

(iii) The thesis fulfills the requirements of the norms and standards prescribed by the University Grants Commission and Integral University, Lucknow,India.

(iv) No published work (figure, data, table etc) has been reproduced in the thesis without express permission of the copyright owner(s).

Therefore, I deem this work fit and recommend for submission for the award of the afore-said degree.

Signature of Supervisor

Dr. Faiyaz Ahmad
(Assistant Professor)
Department of CSE,
Integral University, Lucknow

Date:

Place:

DECLARATION

I hereby declare that the thesis titled **"Service Discover Mechanism for the Diagnosis of COVID-19 using Convolutional Neural Network and Hybrid ML Classifier"** submitted to Computer Science and Engineering Department, Integral University, Lucknow in partial fulfillment of the requirements for the award of the Master of Technology degree, is an authentic record of the research work carried out by me under the supervision of **Dr. Faiyaz Ahmad**, Department of Computer Science & Engineering, Integral University, Lucknow. No part of this thesis has been presented elsewhere for any other degree or earlier.

I declare that I have faithfully acknowledged and referred to the works of other researchers' diploma wherever their published works have been cited in the thesis. I further certify that I have not willfully taken other's work, para, text, data, results, tables, figures etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites without their permission, and have not included those in this MTech. thesis citing as my own work.

In case, this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

Date: Signature__________________

Name: **Fatima Alvi**

Roll. No**. 2001621004**

COPYRIGHT TRANSFER CERTIFICATE

Title of the Dissertation: "**Service Discover Mechanism for the Diagnosis of COVID-19 using Convolutional Neural Network and Hybrid ML Classifier"**

Candidate Name: **Fatima Alvi**

The undersigned hereby assigns to Integral University all rights under copyright that may exist in and for the above dissertation, authored by the undersigned and submitted to the University for the Award of the M.Tech degree.

The Candidate may reproduce or authorize others to reproduce material extracted verbatim from the dissertation or derivative of the dissertation for personal and/or publication purpose(s) provided that the source and the University's copyright notices are indicated.

FATIMA ALVI

RECOMMENDATION

On the basis of the declaration submitted by "**Fatima Alvi**", a student of M.Tech CSE (FT), successful completion of Pre presentation on 24/06/2022 and the certificate issued by the supervisor **Dr. Faiyaz Ahmad**, Assistant Professor, Computer Science and Engineering Department, Integral University, the work entitled "**Service Discover Mechanism for the Diagnosis of COVID-19 using Convolutional Neural Network and Hybrid ML Classifier**" , submitted to the department of CSE, in partial fulfillment of the requirement for award of the degree of Master of Technology in Computer Science & Engineering, is recommended for examination.

Program Coordinator Signature

Dr. Faiyaz Ahmad

Dept. of Computer Science & Engineering

Date:

HOD Signature

Mrs. Kavita Agarwal

Dept. of Computer Science & Engineering

Date:

ACKNOWLEDGEMENT

On the submission of my thesis report of "**Service Discover Mechanism for the Diagnosis of COVID-19 using Convolutional Neural Network and Hybrid ML Classifier"** I would like to express my sincere gratitude & thanks to my supervisor **Dr. Faiyaz Ahmad**, Assistant Professor, Computer Science Engineering, for his constant motivation and support during the course of my work. I truly appreciate and value his esteemed guidance and encouragement from the beginning to the end of this thesis. I am indebted to him for having helped me shape the problem and providing insights towards the solution.

I am greatly thankful to my co-supervisor **Dr. Faiyaz Ahamad**, Computer Science & Engineering for his invaluable suggestions, cooperation and constant encouragement throughout the thesis work. I express my gratitude to **Mrs. Kavita Agarwal**, Head of Department of Computer Science & Engineering for her invaluable suggestions and constant encouragement all through the thesis work.

Lastly, I would also like to express my whole hearted gratitude to my parents for their moral support and motivation which had helped me in the completion of this work

FATIMA ALVI

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

There are several uses for machine learning in prediction which is a novel method. The COVID-19 pandemic necessitates the use of this approach to detect high-risk individuals, as well as their mortality rates and other irregularities. Utilizing it is one way to have a better knowledgeof the structure of the virus and to anticipate potential issues in the future. Machine Learning along with Convolutional Neural Network are the major techniques that can shed light on well-known data sets that are used to train these networks. Imaging methods such as computed to- mography have made it possible to accurately diagnose infected people at an earlier stage in the evolution of the illness. In this research we have collected datasets in the form of CT-scans and chest X-rays of Covid-19 positive patients and healthy people from Kaggle repository. The images that are collected are resized and data normalization was performed on the dataset for better learning of the system. Data pre-processing procedures include noise removal, scaling, and augmentation. The data partitioning approach divides the data into three sets for the exper- iment: training, validation, and testing. Feature extraction and classification are the most im- portant steps. In this experiment, we have used five different models that are XGBoost, Deci- sion Tree, Adaboost, Extra Trees and Support Vector Classifier and performed hybridization method in our HN-model. Finally, criteria such as accuracy, precision, F1-score, recall and others are used to evaluate the built system.

1.1 INTRODUCTION

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, emerged in 2019 as a serious infectious disease that can cause severe respiratory distress. Since its discovery in Wuhan, China, the disease has claimed the lives of millions of people all over the world [1,2]. COVID-19 has become widespread, since the propagation of the virus through human-to-human transmission. Viruses can manifest themselves long before symptoms appear. At the time of writing, over 130 billion people all around the world had been infected [3], posing a massive healthcare burden. Coughing, sneezing, speaking, singing, and breathing can spread the virus from an infected person's mouth or nose in microscopic liquid particles. Larger respiratory droplets to smaller aerosols are among these particles.

If you are near someone who has COVID-19, you can become infected by breathing in the virus or touching a contaminated surface and then contacting your eyes, nose, or mouth. Indoors and in crowded places, the virus is more easily spread. It is very dangerous disease. Many people have lost their lives. So, the best way to prevent or minimize the risk of this disease is to have proper knowledge about it and to do rapid screening of this disease.

COVID-19 is detected using the RT-PCR test [4]. Because the coronavirus is an RNA virus, its genetic material is reverse transcribed to produce complementary DNA (cDNA). This can then be amplified using PCR to make it easier to measure. It is, however, a complicated and time-consuming procedure that takes nearly 2–3 hours and involves the assistance of a professional. Although newer technology can deliver results in 15 minutes, it is also more expensive. Even Nevertheless, investigations have shown that RT-PCR can produce false negative results [5].

Antibody testing is another way of COVID-19 detection [6]. Its goal is to detect the antibody produced by the COVID-19-affected body's immunological response. This method was created for bulk testing of those who had already been infected. It is inexpensive and quick, taking only

15 minutes to complete and can be done in a small laboratory. The problem is that the coronavirus has an average incubation period of 5.2 days [7], and antibodies are rarely produced before a week after infection, and occasionally even later. As a result, early containment is impossible. Due to the presence of modest symptoms, this diagnostic approach is prone to both false positives and false negatives.

However, a shortage of resources and stringent test environment requirements make it difficult to screen questionable cases quickly and effectively. Furthermore, in some circumstances, RT-PCR inspection results in false negatives [8]. Unfortunately, clinical vaccinations and precise drug/therapy methods are currently unavailable.

While the RT-PCR test is the gold standard for identifying COVID-19, it has some limitations that make the disease challenging to diagnose. RTPCR is a labor-intensive, complex, expensive, and time-consuming procedure. One of the method's disadvantages is that it requires a laboratory kit, which many countries find difficult or impossible to get during crises and epidemics. This procedure is not error-free and biased, much like all diagnostic and laboratory methods used in healthcare. The nose and throat mucosa must be sampled by an expert laboratory technician, which is a painful procedure [9–12].

With the advancement of technology in the modern world, many new techniques have been investigated in healthcare sector for the diagnosis of diseases and prevention of the same. The best way to minimize the risk of this disease is to do quick screening of the disease and to have proper knowledge about it.

Different imaging modalities, such as methods [13-15], are used for COVID-19 diagnosis. When accessible, CT screening is favored over X-rays because of its versatility and three-dimensional pulmonary image [16, 17], despite the fact that X-rays are more economical and commonly available. Traditional medical imaging technologies are critical in pandemic control.

Another method for analysing and forecasting the effects of covid-19 on the human body is medical imaging [18]. Healthy persons and Covid-19 infected patients can be investigated in parallel using CT (Computerized Tomography) scans [19, 20] and chest X-ray imaging. We'll gather data from numerous sources, including X-ray images of healthy and Covid-19-infected patients, and apply three different models (InceptionV3, Xception, and ResNeXt) to contribute to a Covid-19 study. The data is analysed using CNN, a machine learning tool. This study looks at how CNN models can be used to categorise chest X-ray pictures in coronavirus-infected individuals.

1.2 OBJECTIVE

The coronavirus (Covid-19) pandemic has caused destruction all around the world. People have faced a severe crisis due to this pandemic. Researches are being made all over the world to find the best innovation to combat the Coronavirus. The healthcare system is facing lots of challenges; therefore, it is very important to find the best solution for the prediction of Covid-19. To detect Covid-19, imaging techniques like X-rays and Computer Tomography (CT) can be used as it is a lung-related problem. Machine Learning (ML) along with Deep Learning are the major techniques that can shed light on well-known data sets that are used to train these networks. Reverse transcriptase-polymerase chain reaction (RT-PCR)1 testing is commonly used to detect Covid-19 but it does not give instant results. Furthermore, it is a complicated manual process so switching to screening methods such as radiography examination can be a better way to have an early prediction of Covid-19. By using machine learning techniques and including its subset i.e., Deep Learning which can be used to analyze a good amount of datasets of chest x-ray images that can help in predicting and screening Covid-19. The main objectives of this research are as follows:

- To find the best solution for the prediction of Covid-19. To detect Covid-19, imaging techniques like X-rays and Computer Tomography (CT) can be used as it is a lung-related problem.
- To collect data in the form of Computer Tomography (CT) and X-rays from various resources or repository.
- By using machine learning techniques and including Convolutional Neural Network, which can be used to analyze a good amount of data-sets of chest x-ray images that can help in predicting and screening Covid-19.

1.3 RESEARCH QUESTIONS

To achieve the objectives of our thesis, there are some research questions that have been formulated:

1. Which suitable machine learning technique can be used to predict COVID-19?

   Motivation: The motivation of the research question is to conduct a conjunctive literature study and experiment to see what are the appropriate machine learning algorithms that can be best applied to the given data and also to find out which algorithm gives us the best results in predicting COVID-19.

2. What are the features that will influence the predictive result of COVID-19?

   Motivation: The motivation of this research is to conduct an experiment to identify the features that will influence the results of prediction of Corona virus in human beings.

1.4 DEFINING THE SCOPE OF THE THESIS -

This research focuses on development of a machine learning model for predicting COVID-19 in patients. We also work to identify the features from the clinical information of patients that would influence the predictive result of COVID-19. This study does not focus on outer factors such as weather or any environmental factors that might influence results.

1.5 BACKGROUND

1.5.1 MACHINE LEARNING

Machine Learning is a subset of Artificial Intelligence (AI) that originated from pattern recognition and allows data to be organized for user comprehension. Machine Learning has recently been used in a variety of industries, including healthcare, banking, military equipment, and space exploration. Machine Learning is now a fast expanding and quickly developing field. It optimises the performance of computers by programming them with data. It uses training data or previous experiences to understand the parameters for optimising computer programmes. It can also forecast the future using the data. Machine Learning can also assist us in developing a mathematical model based on data statistics. Machine Learning's major goal is to learn from feed data without the intervention of humans, that is, it learns from given data (experience) and produces the desired output by searching for trends/patterns in the data[24].

It is broadly classified into four types:

- **SUPERVISED LEARNING**

Supervised learning approaches are machine learning techniques or algorithms that use labelled data to link prior and current datasets to predict future events [25, 26]. The learning procedure starts with a dataset training procedure and progresses to targeted activity in order to predict output values [27–29]. The strategies can produce results in input data with an adequate training procedure, compare results to actual results and expectations to discover flaws, and alter the model based on the findings [27,30].

Discrete responses are predicted by classification. The algorithm labels each case by selecting two or more classes. It's termed binary classification if it's done between two classes, and it's called multi-class classification if it's done between two or more classes. Handwriting recognition, medical imaging, and other classification applications are only a few examples.

Continuous responses are predicted through regression. The algorithms return a statistical value in this case. For example, a set of data is gathered to show that people are happy when the amount of sleep is taken into account. Both sleep and happiness are factors in this equation. The analysis is now carried out by creating predictions [31]. Linear regression and logical regression are two prevalent regression approaches.

➢ UNSUPERVISED LEARNING

When the training dataset is unclassified or unlabeled, unsupervised learning approaches are applied [32]. From an unlabeled dataset, the learning algorithms deduce a function to extract hidden knowledge or a pattern [34]. The technique does not determine the correct output; rather, it pulls observations from the dataset to uncover latent patterns in the unlabeled dataset [29,33].

The assumptions are set in this case so that clusters can be discovered that are pretty well matched to a classification. This is a data-driven method that works best when enough data is available. For example, movies on Netflix.com are recommended based on the clustering of movies principle, which groups multiple related movies based on a customer's recently seen movie list. It primarily discovers unknown patterns in data, however these approximations are typically weak when compared to supervised learning [35].

➢ SEMI-SUPERVISED LEARNING

Semi-supervised learning approaches consists of both supervised and unsupervised learning techniques, in which labelled and unlabeled datasets are combined in the training process [33]. Learning strategies often take into account a smaller labelled dataset and a larger unlabeled dataset [34]. When a labelled dataset requires competent and sufficient resources for training or learning, the learning techniques can be customized to obtain improved accuracy, and the techniques are preferable [29, 36-39].

Semi-supervised machine learning achieves great accuracy while requiring less annotation. To improve classifiers, semi-supervised machine learning uses largely unlabeled data supplemented with tagged data. Humans have less work to accomplish here since less annotation labor is sufficient to provide acceptable accuracy.

➢ REINFORCEMENT LEARNING

Actions to discover faults interact with the learning environment in reinforcement learning systems [7]. Some of the common elements of reinforcement learning approaches are delayed rewards and trial and error searches, and the techniques are used to discover the optimum behavior in a certain scenario to improve the model's effectiveness [29,34, 40].

There is a display of the input or output data in Reinforcement Learning. Instead, when the agent selects the desired action, the agent is immediately informed of the reward, and the next state does not take long-term actions into account. The agent must actively know about states, rewards, transitions, and actions in order to function optimally. The model is formalized as [41]:

- ❖ a discrete set of environment states, S;
- ❖ a discrete set of agent actions, A;

- ❖ a set of scalar reinforcement signals; typically {0;1} or the real numbers.

By comparing general deep learning-based feature extraction frameworks to create the greater accurate feature, which is a key module of learning, the technique of automatic classification of COVID-19 may be used. Among a group of deep convolutional neural networks CNN, MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, and NASNet were chosen. The classification was completed by feeding the extracted features into some machine-learning classifiers, which identified them as COVID-19 or other diseases cases [42].Progressive machine-learning algorithms can integrate and evaluate a large amount of data relating to COVID-19 patients in order to provide the best understanding of the viral spread pattern, improve diagnostic accuracy, develop new and effective treatment options, and even identify individuals who are at risk of the disease based on genetic and physiological characteristics (43).

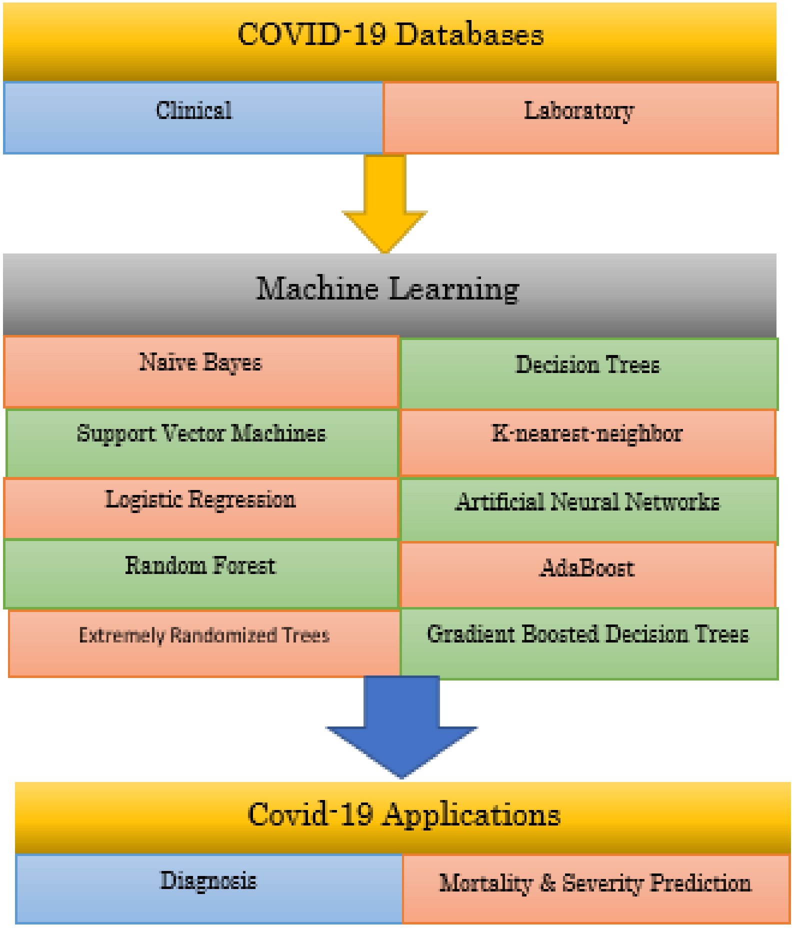

Fig.1.1. Classification of Machine Learning in Covid-19

A machine learning-based approach, similar to those used for other diseases such as breast cancer and pneumonia, could be a potential answer for evaluating individuals more quickly. [44].

As previously noted, achieving excellent performance with a CNN based classifier would necessitate a significant amount of data, which is not available given the disease's novelty. Although small datasets have been gathered, they are insufficient to achieve the requisite level of generalization [44] [45].

In addition, to avoid over-fitting, augmentation is used in this collected dataset. Rotation, zooming, and sharing of photographs were among the enhancements. The data was then jumbled to make it broader.

2.1 RELATED WORK

COVID-19 is identified utilizing chest X-rays in studies with binary or multiple categories. Raw data is used in certain research, whereas feature extraction is used in others. The quantity of data points employed in investigations varies as well. Convolutional neural networks have been found to be the most popular strategy in investigations (CNN). COVID-19-induced pneumonia was employed by. The approach known as transfer learning has been used in particular. The diagnosis of numerous anomalies in tiny medical image datasets can be accomplished with transfer learning, and the results are generally impressive [46].

Based on chest X-ray images, this research aims to develop a deep learning-based model that can detect COVID-19 with high sensitivity, providing fast and reliable scanning [47]. Deep learning approaches can learn from simple representations to understand complicated situations. The ability to learn accurate representations and the property of learning the data in a deep manner where multiple layers are utilized sequentially [48],[49] are the major qualities that have made deep learning methods popular. Deep learning algorithms are frequently employed in biomedicine [50], smart healthcare [51], drug development [52], medical image analysis [53], and other medical systems.

COVID-19 can be diagnosed using radiography pictures, such as chest X-ray and computed tomography (CT), because the disease primarily affects the human respiratory system [54]. A typical pneumonia [55] and organizing pneumonia [56] are the most common abnormalities on chest X-rays. Ground Glass Opacity (GGO), which refers to a region of enhanced attenuation in the lung, is the most prevalent finding in chest radiography pictures. A chest X-ray image exhibits some hazy grey tint instead of black with fine white blood vessels. CT scans, on the other hand, reveal GGO and, in extreme cases, consolidation. The appearance of GGO with a superimposed interlobular and intralobular septal thickening is sometimes referred to as "crazy paving" on chest imaging. These results can be seen on their own or in combination. They can affect the peripheral portion of the lungs and occur in numerous lobes.

It's worth noting that chest CT is thought to be more sensitive [57] than chest Xray for early COVID-19 diagnosis since chest Xray can be normal for 4–5 days following the onset of symptoms, whereas CT scan shows a typical pattern of GGO and consolidation. CT scans can also reveal the severity of the disease [58]. Recent research of 51 COVID-19 patients found that CT had a sensitivity of 98 percent for COVID-19 infection, compared to 71 percent for RT-PCR [59].

The key issue is that similar findings are reported in pneumonia cases as well as COVID-19 cases. Many moderate COVID-19 cases have symptoms that are comparable to those of a regular cold, and in some cases, the lungs appear to be normal. Despite the fact that research published in [60] found that the radiographic image of COVID-19-affected lungs varies from that of bacterial pneumonia-affected lungs.The main concern is that similar findings have been reported in both pneumonia and COVID-19 instances. Many moderate COVID-19 patients exhibit symptoms that are similar to a common cold, and the lungs in some cases appear to be normal. Despite the fact that a study published in found that the radiographic appearance of COVID-19-damaged lungs differs from that of lungs affected by bacterial pneumonia.

Many researchers established several distinct architectures. Each one was created for the classification of photos and was created with a distinct aim in mind. Many new image classification architectures include an inception module as a crucial component.

In medical images inception module is one of the great innovations which helps in identifying the features more accurately. Each layer in a standard CNN model sends its output to one of the other layers as input. This is repeated until the output layer is reached, in a pipeline-like approach.

Khabir Uddin Ahameda et al, utilised a preprocessing approach on the picture dataset to allow our deep learning model to analyse it accurately and quickly. They developed a ResNet50V2-based deep learning model that can accurately detect and diagnose COVID-19 patients. In their study, they demonstrated that our algorithm can accurately recognise COVID-19 patients using chest CT-scan and X-ray image datasets. To increase the model's accuracy and durability, advanced image processing techniques like hybrid filtering must be used. They started with ResNet50V2 and added layers. The suggested design grows deeper and more sophisticated by adding layers to the existing levels. Deeper models extract more characteristics, but they require a long time to train on huge datasets. As a consequence, they want to create a deep learning model that is simple, practical, and resilient.

Rachna Jain et al, employed the PA view of chest x-ray images for both covid-19 patients and healthy people. They compared deep learning-based CNN models' performance after picture cleaning and data augmentation. They compared Inception V3, Xception, and ResNeXt. Kagglerepository contributed 6432 chest x-ray scan samples for training and validation. The Xception model accurately identifies Chest X-rays (97.97%). This study examines various covid-19 categorization algorithms and makes no medical claims.

Lin Li et al, In 2019, used deep learning to identify coronavirus in chest CT images. Using independent testing data, they found that this model had high sensitivity (90%) and specificity (96%). In COVID-19 and CAP, AUC was 0.96. (95 percent CI: 0.94, 0.99). COVNet can extract two-dimensional local features. The system's backbone is RestNet50, which takes in CT slices and develops features for them. The recovered features are max-pooled across all slices. The resulting feature map is put into a fully linked layer with a softmax activation function (COVID-19, CAP, and non-pneumonia). They preprocess a 3D CT image and extract the lung area as the region of interest using U-net (17). The picture is then transmitted to our

COVNet for prediction. The results also reveal that a convolutional network model can identify COVID-19 from CAP.

Himadri Mukherjee et al, compared to deep CNN designs, this study proposes a shallow CNN architecture with only four layers. The main objective was to build a light architecture with few parameters (weights) to save calculation time. Researchers studied Covid-19 positive, pneumonia positive, and healthy CXRs. 321 COVID-19 positive occurrences were used in 5-fold cross validation on both balanced and unbalanced datasets to assess its durability. A popular deep learning tool like MobileNet, InceptionV3, and ResNet50 was compared to the suggested shallow CNN-tailored architecture, as was Covid-19 identification using CXRs. To screen for Covid-19 positive patients in chest X-rays, the proposed shallow CNN customised architecture might be applied.

Tianyang Li et al, proposed discriminative cost-sensitive learning (DCSL) as a method for COVID-19 aided screening from chest X-rays. DCSL combines fine-grained categorization with cost-sensitive learning. Initial DCSL conditional centre loss DCSL uses cost-sensitive learning at the score level to lower the cost of misclassifying COVID-19 data. DCSL may be used in any DNN. 239 from verified COVID-19 cases, 1,000 from bacterial or viral pneumonia cases, and 1,000 from healthy people. Tests on three-class classification demonstrate their strategy outperforms existing methods. It is 97.01 percent accurate in terms of precision, sensitivity, and F1 score. These results confirm their large-scale COVID-19 screening approach.

Cheng Jin et al, proposed an AI based system where patients with influenza A/B, non-viral CAP, and non-pneumonia were included in the study. Multi-way classification accuracy is 97.81 percent, with AUCs of 92.99 percent and 93.25 percent on two publicly available datasets, the CC-CCII and the MosMed Data.On reader research with five radiologists, the artificial intelligence system outperformed them all by two orders of magnitude in progressively demanding tasks. When comparing the diagnostic performance of CXR and CT, there are several differences.

As an alternate diagnosis method for COVID-19, Ali Narin et al. developed an automatic detection system. "Three distinct convolutional neural network-based models (ResNet50, InceptionV3 and Inception-ResNetV2) have been proposed for the detection of corona virus pneumonia infected patients utilising chest X-ray radio graphs [61]," according to this research. The author also compares and contrasts the three CNN models' classification performance accuracy.

A retrospective, single-center analysis of diverse patient data from Wuhan, China's Jinyintan Hospital was conducted by Nanshan Chen et al. They described epidemiological data, signs and symptoms, laboratory results, CT findings, and clinical outcomes in this study[62]. Although this study does not directly address COVID-19 prediction, it does provide insight into clinical outcomes.

The radiographic abnormalities in CT scans of COVID-19 patients in China have been identified by Shuai Wang et al. He developed COVID-19 as an alternate diagnostic approach by using deep learning technologies to extract graphic information from CT scan pictures. They gathered CT pictures of COVID-19 patients who had been confirmed as well as those who had been diagnosed with pneumonia. Their findings show that AI may be used to predict COVID-19 with high accuracy [63]. This study employs CT scan pictures instead of clinical characteristics and laboratory results for prediction, which is different from ours.

The authors suggested a three-index-based approach to predict mortality risk in [64]. To forecast the mortality risk in patients, they developed a prognostic prediction model based on the XGBoost machine learning algorithm. They devised a clinical path that is straightforward to assess and check for death risk. The study focuses on mortality risk, which differs from ours, in which the prediction is entirely based on the clinical results of COVID-19 patients.

The authors of the article [65] offered a comparison of machine learning algorithms for predicting COVID-19 outbreaks in different nations. Their research and analysis show that machine learning models can be used to predict COVID-19. The entire text was based on the breakout of instances in different nations. We use clinical information to forecast disease in our work.

The epidemiological, demographic, clinical, laboratory, radiological, and therapeutic data from Zhongnan Hospital in Wuhan, China, were described in this study by Dawei Wang et al. The data was analysed and documented so that the infections could be tracked [66]. The author provides more information regarding the radiological and therapeutic data that could be employed in our model to predict COVID-19.

Halgurd S. Maghdid and colleagues have presented a new framework for detecting corona virus sickness using smartphone sensors. The developed AI framework gathers data from a variety of sensors to forecast the severity of pneumonia as well as the disease's infection [67]. To forecast COVID-19, the proposed system uses uploaded CT Scan pictures as the primary technique. This framework is based on multigradings from multiple sensors that are linked to COVID-19 symptoms.

2.2 CONVOLUTED NEURAL NETWORKS (CNN)

Although deep learning was used to abstract feature extraction, the architecture of the neural network has an impact on how features are extracted automatically.

Convoluted neural networks are one such architecture which are used in image classification problems. CNNs are used to increase the recognition of patterns inside images, such as edges or forms, regardless of their location. This is known as translation invariance. To do this, the formula 1 is used to generate a convoluted picture of the input image using a convolution process.

Equation 1 can be written as the dot product of the smoothing kernel and the image's m n matrix. In order to generate feature maps, there is a sliding window of m×n that passes across the image. Additional layers for subsampling the features are added after the convolutional layers in the neural network to lessen susceptibility to input distortions, such as translations and rotations.

$$S(i, j) = \sum m \sum n \; I(m, n)K(i - m, j - n) \text{----------------} Eq.(1)$$

I = A two-dimensional image

K = A two-dimensional smoothing kernel

Equation 1 can be written as the dot product of the smoothing kernel and the image's m×n matrix. In order to generate feature maps, there is a sliding window of m×n that passes across the image. Additional layers for subsampling the features are added after the convolutional layers in the neural network to lessen susceptibility to input distortions, such as translations and rotations.

2.3 DEEP LEARNING

Machine learning is the idea of learning a function that maps data from an input space to a desired output using a data model. This necessitates a computer-readable data representation that can be explained using mathematical models. The data representation is done by selection and extraction phase, either by statistical analysis or intelligent guesses based on domain expertise. However, there are several drawbacks with data representation construction. First, it is task specific; one representation cannot be used for multiple tasks. Second, because of the enormous dimensionality of the input data, feature engineering is painstaking, time demanding, and even impossible for some applications, such as picture classification.

Deep learning is a machine learning technique that automates feature engineering. By constructing a hierarchical representation of the data and learning features at many levels of abstraction, i.e., a combination of lower-level characteristics makes up higher level features. This is accomplished by the use of neural networks with multiple layers, each of which acts as a hierarchy level. A deep artificial neural network is created by using numerous hidden layers, hence the term deep learning. [68]

2.4 TRANSFER LEARNING

When training a neural network, it learns to solve tasks within a specific feature space and distribution, and when presented with a new task with a different objective, performance is reduced. A network that correctly detects cats in an image, for example, may or may not be good at identifying dogs. The idea of transfer learning is, as the name suggest, transferring knowledge learnt while solving task Ta to the training of a model that solves task Tb, rather than training a new network that solves Tb from scratch. Concretely, this could mean reusing certain low level feature extractors from the network that solves Ta and applying these on Tb.

Given a domain D, which consists of a feature space χ and a probability distribution P(X), such that: D = {X,P(X)}, where X = {x1,x2,...xn} $\in \chi$, that is X is a particular sample. Then the task that needs to be learned is defined as a tuple T = $\{\gamma, f(\cdot)\}$, where γ is the label space and f is the objective predictive function. Assuming there is a source domain DS and task TS, moreover, there exists a target domain DT accompanied with a task TT which needs to be learnt. Transfer learning is the process of improving the performance of the objective function fT within the task of the target domain, by exploiting knowledge from the source domain DS and task TS. There are three methods for transfer learning: inductive, unsupervised and transductive transfer. Which one to be used depends on the relationship between the targets and source's domain, probability distribution and label space.

In this paper, the source domain DS is defined as an input image from dataset A and the source task TS is the task of classifying said image. Moreover, the target task involves identifying presence of COVID19 pneumonia from X-ray images given by dataset B.

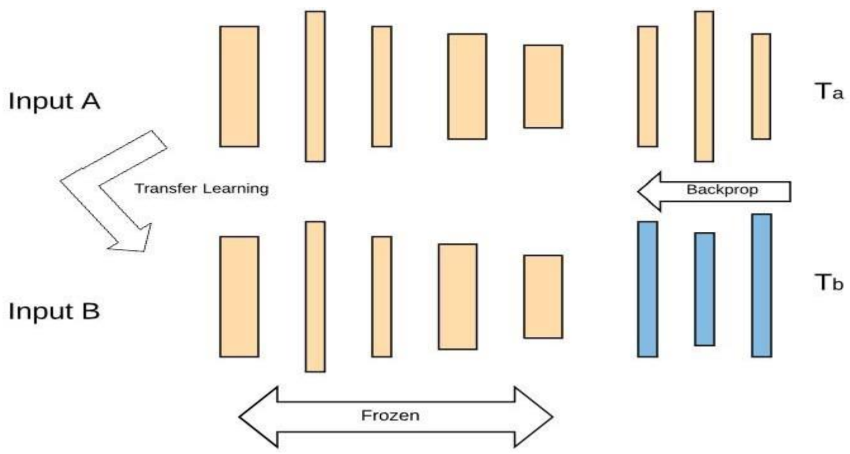

Figure 2.4: Transfer Learning

Inductive transfer will be employed because the tasks for the source and target are different but the domains are the same. The concept of transfer learning can be seen in figure 2.1. [69]

2.5 ARTIFICIAL NEURAL NETWORK

Artificial neural networks, or "neural networks," are computer systems that are modelled after the organic neural networks that make up human brains. Artificial neurons are a collection of interconnected units or nodes in an artificial neural network (ANN) that loosely emulate the neurons in a biological brain. Artificial neural networks are the first type of neural network (ANN). "An Artificial Neural Network (ANN) is a data or signal processing system made up of a large number of simple processing pieces connected by direct links that work together to execute parallel distributed processing to complete a specific computational goal" [70]. Multi-layer perceptrons is another name for an ANN. They're a collection of dense layers that excel at extracting relationships from large datasets. Dense layers, on the other hand, require a lot of parameters and consequently a lot of memory and time. ANNs have a variety of applications, including facial identification, stock market forecasting, and social media [71], despite their computational cost.

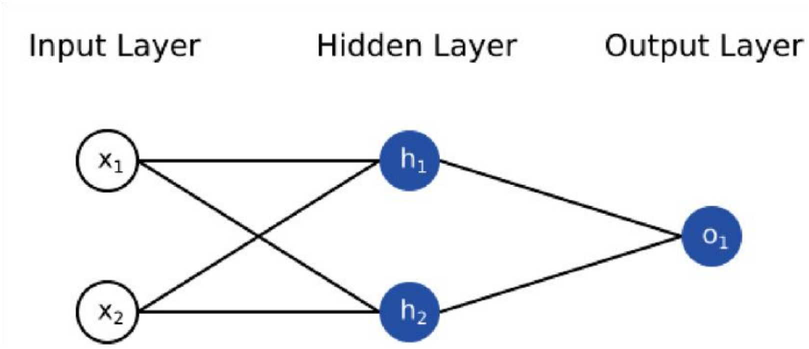

Figure 2.5: Neural Network [72]

2.6 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machines (SVMs) classify data by creating an N-dimensional hyper plane that divides it into two groups [31]. The predictor variable is referred to as an attribute, and the transformed attribute is referred to as a feature in SVM. Feature selection is the process of selecting the most appropriate representative data. A vector is a collection of features that describe a single case.

The ultimate purpose of SVM modelling is to discover the best hyper plane that separates clusters with a target variable on one side and another category on the other. The support vectors are the vectors that are close to the hyper plane [35]. A simple example of a support vector machine is shown in Figure 2.1.

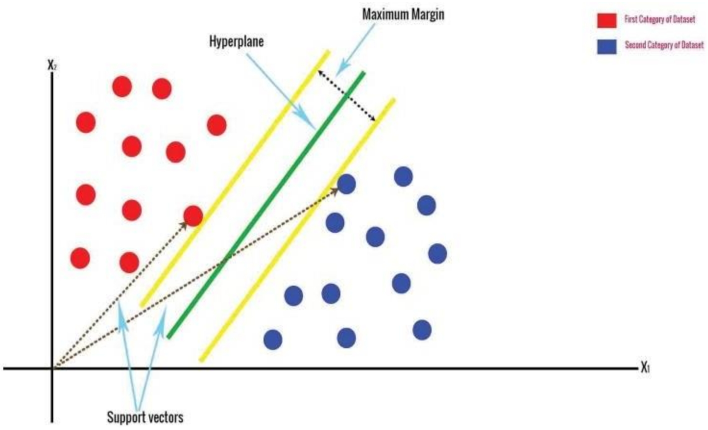

Fig 2.6 Support Vector Machine [73]

2.7 RANDOM FOREST (RF)

RF's random sampling and ensemble procedures allow it to make more accurate predictions and generalizations [74]. The random woodlands have a big amount of trees in them. The accuracy improves as the number of uncorrelated trees increases [75]. Random Forest classifiers can assist in filling in some gaps in data. Figure 2.3 depicts the prediction in Random Forests (RFs).

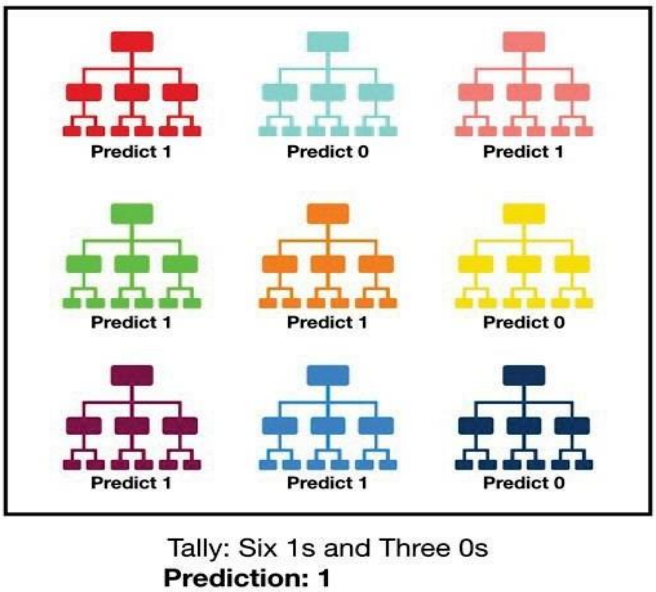

Fig 2.4: Visualization of Random Forest making a prediction [75]

2.8 LOSS FUNCTION

A deep learning neural network is taught to map a set of inputs to a set of outputs using training data. There are too many unknowns to compute the ideal weights for a neural network. Instead, the learning problem is recast as a search or optimization problem. An algorithm is used to navigate the space of possible weight settings that the model could employ to create excellent or adequate predictions.

2.9 GRADIENT DESCENT

A neural network model is typically trained using the stochastic gradient descent optimization process, and the weights are updated using the backpropagation of error algorithm.

The error gradient is the "gradient" in gradient descent. Predictions are made using a model with a set of weights, and the error for those predictions is determined.

2.10 OBJECTIVE FUNCTION

The objective function is the function that is used to evaluate a candidate solution (i.e. a set of weights).

We might try to maximize or decrease the objective function, which means we're looking for a potential solution with the highest or lowest score.

When it comes to neural networks, the goal is usually to reduce the number of mistakes. There are a variety of loss functions to choose from. Binary cross-entropy loss is the loss function employed in this investigation. As a result, the objective function is sometimes referred to as a cost function or a loss function, and the value computed by the loss function as "loss."

The loss during training is determined by the backpropagation technique. The following is the equation for binary cross-entropy loss:

$$-\sum \quad y \cdot \log p(y) \quad + (1 - y) \cdot \log(1 - p(y))$$

where y is the actual value (0 or 1), p(y) denotes the anticipated chance that the image would be categorized as positive (a value of 1), and N denotes the number of photos in the test set. We can use this technique to determine the binary cross-entropy loss for the complete test set by applying the same calculation to each test image.

This is because when y = 1, the calculation is log p(y), or the log of the probability that the image should be categorized as 1, and when y = 0, the calculation is log p(y), or the log of the chance that the image should be classified as 0 (which is 1 - p(y)). Furthermore, employing log values penalizes bad choices more harshly, forcing the model to strive for the best option.

2.11 OPTIMIZERS IN DEEP LEARNING

While neural networks are the talk of the town right now, an optimizer is significantly more important to a neural network's learning process. While neural networks can learn without any prior knowledge on their own, an optimizer is a programme that runs alongside the neural network and speeds up learning. In a nutshell, it accomplishes this by modifying the neural network's settings in such a way that training with the neural network becomes much faster and easier. These optimizers allow neural networks to work in real-time, with training taking only a few minutes. Training would take days if they weren't there.

Some of the optimizers are Gradient Descent, Stochastic Gradient Descent, Mini- Batch Gradient Descent, Momentum Based Gradient Descent, Nesterov Accelerated Gradient (NAG), Adagrad, RMSProp, Adam.

2.12 LAYERS

Layers come in a variety of shapes and sizes, and we'll go over a few of them here. A layer, as you may recall, is a function that is applied to the input of an array of numbers in order to obtain a better understanding of that array and provide an output in the form of another array of numbers.

2.12.1 Input: A model's input is represented as an array with up to three dimensions by an input layer.

2.12.2 Convolution (2D): A filter is run across the image from left to right and top to bottom in a convolution layer. The numbers in a filter are multiplied by the numbers in the same size section of the input array as the filter. The resultant array holding the multiplication of those numbers is then averaged and placed in the output array at the pixel in the middle of the filter region. As a result, filters should be square and the length of the sides should be an odd number of pixels. This yields an array of numbers reflecting the filter's relationship to the area surrounding each pixel. This convolution is then applied to each of the arrays that have been passed through the convolution layer. The depth of the input determines the number of arrays that are input to the layer. The final output array is averaged from the results of each input array's convolution. Because most filters are larger than one pixel, they are unable to detect features at the array's edge. Padding attempts to address this by adding one or more layers of zeros to the array's exterior. Given that the filter has an odd side length, the number of layers of zeros inserted in padding is equal to where n is the length of the side of the filter. The convolutional layer's output can be written as:

$$xjl = f\left(\sum_{a=1}^{N} w_j^{l-1} * y_a^{l-1} + b_j^l\right) \text{---------Eq. (2)}$$

where x_j^l is the jth feature map in layer l,

$w^l{}_j{}^{-1}$ indicates jth kernels in layer $l-1$,

$y^l{}_a{}^{-1}$ represents the ath feature map in layer $l-1$,

b^l_j indicates the bias of the jth feature map in layer l,

N is number of total features in layer $l-1$,

and ($*$) represents vector convolution process.

2.12.3 Flatten: The flatten layer outputs a one-dimensional vector from the integers in the input array. This makes creating dense layers to arrive at a final probability for prediction much easier. In a flatten layer, no calculations are performed on numbers. It simply turns a multi-dimensional input into a one-dimensional output.

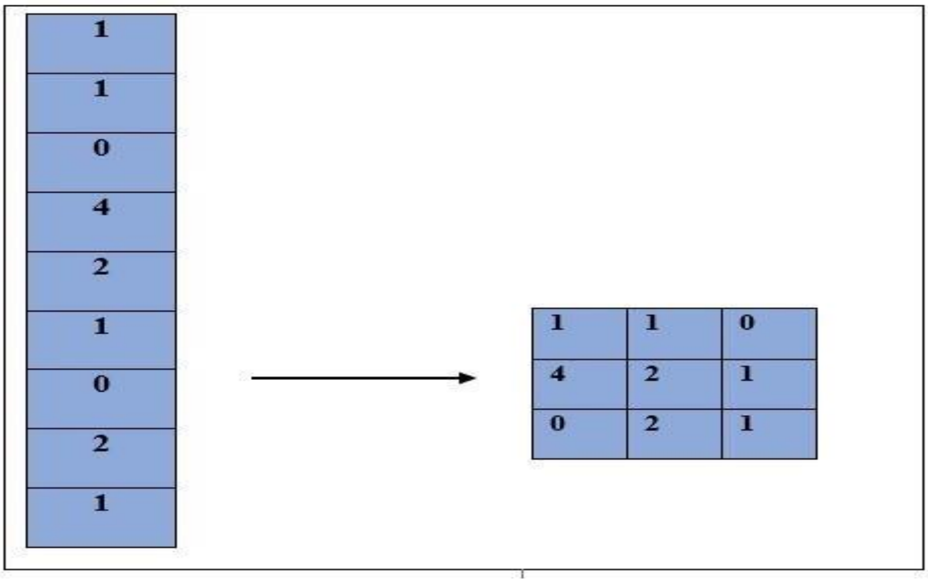

Fig. 2.5. Flattening Operation

2.12.4 Dense: A fully connected layer is sometimes known as a dense layer. Every neuron in the input is coupled to every neuron in the output in a dense layer, where a neuron is a number in a vector. When designing the architecture, the number of neurons in the output is specified. Each number in the preceding vector has a different weight, therefore a number in the output vector is a combination of all the other numbers in the previous vector. The influence of one neuron in the input on one neuron in the output is referred to as a weight. In addition to the bias parameter, each of these weights is a trainable parameter that will be optimised during model training.

2.12.5 Max Pool (2D): From left to right and top to bottom, a square window of a defined dimension sweeps across an array of numbers. You can also change the stride size (the number of pixels the window travels by each iteration). Each window's maximum value is taken and stored in an output array. The output array for a max pool layer with a dimension size of two and a stride of two is half the length and half the width of the input array. In a max pool layer, there are no weights or filters. It exists to reduce the size of the arrays by condensing the characteristics to allow for greater examination.

2.12.6 Average Pooling (2D): Average pooling and max-pooling are similar. Instead of choosing the highest value from each window, the average is used.

2.12.7 Concatenate: Concatenate takes multiple input arrays and concatenates them into a single set of arrays that may be supplied to subsequent levels. These arrays will be stacked in order of depth. To be concatenated, all of the inputs must have the identical width and height dimensions. In networks using inception modules, concatenate layers are frequently employed.

2.12.8 Global Average Pooling (2D): This method averages all of the numbers in each of the feature maps. This produces a vector with the same length as the number of feature maps. Overfitting can be reduced by using global average pooling.

2.12.9 Softmax: Softmax takes an input vector and calculates and normalises the exponential values of each item. This guarantees that they add up to one, allowing them to be transformed into probabilities. The Softmax equation is as follows:

$$\sigma(\vec{z}) = \text{--------------------} \quad \text{Eq. (3)}$$

$$\sum$$

where $\vec{z}$ is the input vector, $\sigma(\vec{z})$ is the output number for element i, e is the value of element i in vector $\vec{z}$, and e is the value of element j in vector $\vec{z}$.

2.12.10 Sigmoid: This function accepts a vector and applies a sigmoid function to it. A vector can be turned into a confidence prediction between two categories using this method. S(x) = ______

Where x is the value of an array [76]

2.12.11 Dropout: Dropout takes a number of arrays and removes a random number of them. This can be used to keep big neural networks from overfitting.

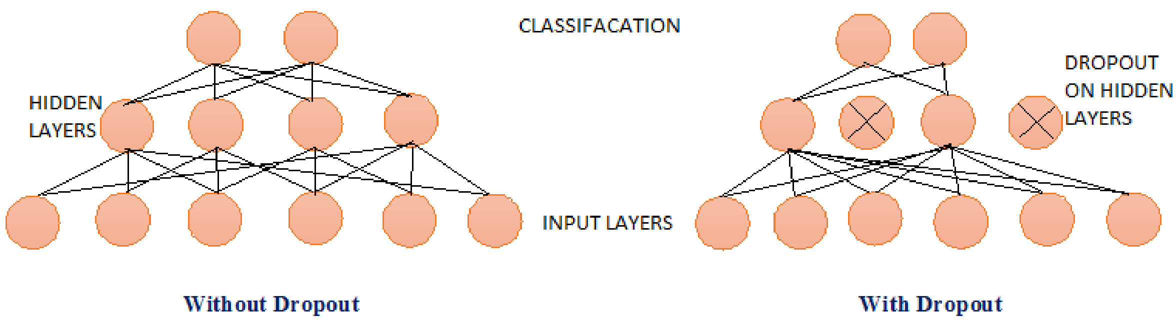

Fig. 2.6. Dropout Techniques

2.13 AUGMENTATION

To increase datasets, data augmentation could be employed during training. Various methods are used for data augmentation to alter the shape of the data. Some of the most popular augmentation techniques include flip, rotation, scale, crop, translation, and Gaussian noise. Utilizing augmentation is intended to produce new, unused data for training. The network will have more data for training, which will lower the risk of overfitting and increase model performance by having a generator that enhances each picture as it passes through the network.

The majority of machine learning and deep learning algorithms identify the most glaring characteristics that separate one class from another, hence performance is unaffected by dataset size.

2.14 BACKPROPAGATION

Using a concept known as the delta rule or gradient descent, the Backpropagation algorithm hunts for the lowest value of the error function in weight space. The weights that minimise the error function are then regarded a learning problem solution. Backpropagation is a short form for "backward propagation of mistakes" in a neural network. It's a common way to train artificial neural networks.

This method aids in calculating the gradient of a loss function with respect to all of the network's weights. The gradient of the loss function for a single weight is computed by the chain rule in the back propagation algorithm in neural networks. In contrast to native direct computation, it efficiently computes one layer at a time. It calculates the gradient but leaves it up to the user to decide how to use it. It expands upon the delta rule's computation.

2.15 LIBRARIES

One of the most important subfields in machine learning is deep learning. Machine learning is the study of algorithm creation based on the human brain model. Deep learning is gaining traction in disciplines such as robotics, artificial intelligence (AI), audio and video recognition, and image recognition. Deep learning approaches are built around artificial neural networks. Deep learning is supported by various libraries such as Theano, TensorFlow, Caffe, Mxnet, and others. Keras is one of the most powerful and easy-to-use python libraries for creating deep learning models, and it is built on top of popular deep learning libraries such as TensorFlow, Theano, and others.

Deep Learning is a popular subset of a larger family of machine learning algorithms based on data representations that is constantly evolving. As a relatively new concept, the sheer amount of resources available can be overwhelming for those considering entering or already working in the sector. Interacting with the community by participating and interacting with the deep learning open-source projects that are now available is a good method to remain up to date with the latest trends.

TensorFlow is the undisputed champion on all fronts. The top five are Keras, Caffe, Microsoft Cognitive Toolkit, and PyTorch.

In this research we have used three libraries that are Tensor Flow, Theanos and Keras.

2.15.1 TENSOR FLOW

Google Brain Team researchers and engineers produced Tensor Flow as part of the company's Machine Intelligence research department. The system is intended to facilitate machine learning research as well as the transition from a research prototype to a production system. TensorFlow comes with a lot of machine learning libraries and is well-documented. It provides a few key functionalities and ways for doing so. It combines computational algebra and optimization approaches to make many mathematical equations simple to calculate.

So, these are some of the TensorFlow's key features.

- ❖ It has a feature that quickly creates, optimises, and calculates mathematical expressions using multi-dimensional arrays known as tensors.
- ❖ It has deep neural network and machine learning programming support; • It has a high scalable characteristic of computation with varied data sets.
- ❖ TensorFlow employs GPU computing to automate management. It also has a one-of-a-kind feature that optimises the utilisation of the same memory and data.

2.15.2 KERAS:

TensorFlow, Theano, and Cognitive Toolkit are open-source machine libraries that Keras runs on top of (CNTK). Theano is a Python module for doing quick numerical computations. The most well-known symbolic math toolkit for constructing neural networks and deep learning models is Tensor-Flow. TensorFlow is extremely adaptable, with distributed computing as its key advantage. It employs Python, C#, and C++ libraries as well as standalone machine learning toolkits. Theano and TensorFlow are great libraries for building neural networks, but they are difficult to grasp.

Keras is based on a simple framework that makes it simple to build deep learning models using TensorFlow or Theano. Keras was created with the goal of easily defining deep learning models. Keras, on the other hand, is an excellent choice for deep learning applications.

Keras makes high-level neural network API easier and more performant by utilising multiple optimization approaches. It has the following capabilities:

- An API that is consistent, easy to use, and expandable.
- Simple construction - no frills required to obtain the desired result.
- It works with a variety of platforms and backends.
- It's a user-friendly framework that works on both the CPU and the GPU.
- Extremely scalable computation.
- Benefits:
- Keras is a robust and dynamic framework that offers the following benefits: • Greater community support.
- It is simple to test.
- Keras neural networks are built in Python, making them easier to use.

- ❖ Both convolutional and recurrent networks are supported by Keras.
- ❖ Because deep learning models are distinct components, they can be combined in a variety of ways.

2.15.3 THEANOS

Theano is a Python package that allows you to build Machine Learning mathematical expressions, optimise them, and evaluate them quickly by utilising GPUs in important areas. In most circumstances, it can compete with full-fledged C implementations. enables you to quickly define, optimise, and analyse multi-dimensional array mathematical expressions.

The creation of a Machine Learning model necessitates a large number of mathematical calculations. These typically necessitate mathematical computations, particularly when dealing with big matrices with several dimensions. For designing Machine Learning applications, we now leverage Neural Networks rather than classic statistical methodologies. Neural Networks must be trained on a massive amount of data. The training is carried out in batches of data that are of a manageable size. As a result, the learning procedure is iterative. Training the network can take several hours or even days if the computations are not done efficiently. As a result, optimizing the executable code is highly desirable. That is precisely what Theano offers.

Theano is a Python library that lets you define mathematical expressions used in Machine Learning, optimize these expressions and evaluate those very efficiently by decisively using GPUs in critical areas. It can rival typical full C-implementations in most of the cases.

Theano was written at the LISA lab with the intention of providing rapid development of efficient machine learning algorithms. It is released under a BSD license.

3.1 PROPOSED METHODOLOGY

So, this research is all about to make Covid-19 screening more efficient and effective for the patients. Covid-19 is an infectious disease caused by SARS Covid-2 virus. Anyone can get sick sick with Covid-19 and become seriously ill or even die from this disease. Its so many new variants have come so far. Different waves comes and go. The best way to prevent and slow down transmission of this disease is to be well informed about the disease and to do quick screening.

We need to build some faster algorithms and so we have hybridized the algorithms. We mixed the algorithms as to get better results. We have taken out features from GoogleNet, etc. from flatten layers and save it on the file and used machine learning algorithms like knn, naïve baise using PCA.

CNN (Convolutional Neural Network) which is a type of artificial neural network used in image recognition and processing that is used to process pixel data. CNN takes image’s raw pixel data, trains the model, then extract features automatically for better classification. Eg-VGG 16, ResNet, GoogleNet, etc.

Machine Learning is a field of study which gives computers the capability to learn without being explicitly programmed. We have used three libraries that are keras, theanos and TensorFlow. We have chosen a hybrid model. Epoch achieved was 20. In order to reduce complexity, we have used VGA. CNN works on different data sets of same size. We have taken data from Kaggle of different sizes and tried improving accuracy, and taken out images from flatten layer.

The process starts with feeding good quality of data then training the machines by building machine learning models using data and different algorithms. The feature extraction was done on the first-fully connected layers of the model. In this layer the feature vector was extracted from each training image. The feature vector was then fed into four classifiers. Then all the machine learning classifiers were hybridized to develop an ensemble of classifiers. The performance analysis is then performed in which the evaluation of the proposed system was evaluated in terms of Accuracy, Precision, F1-score and recall.

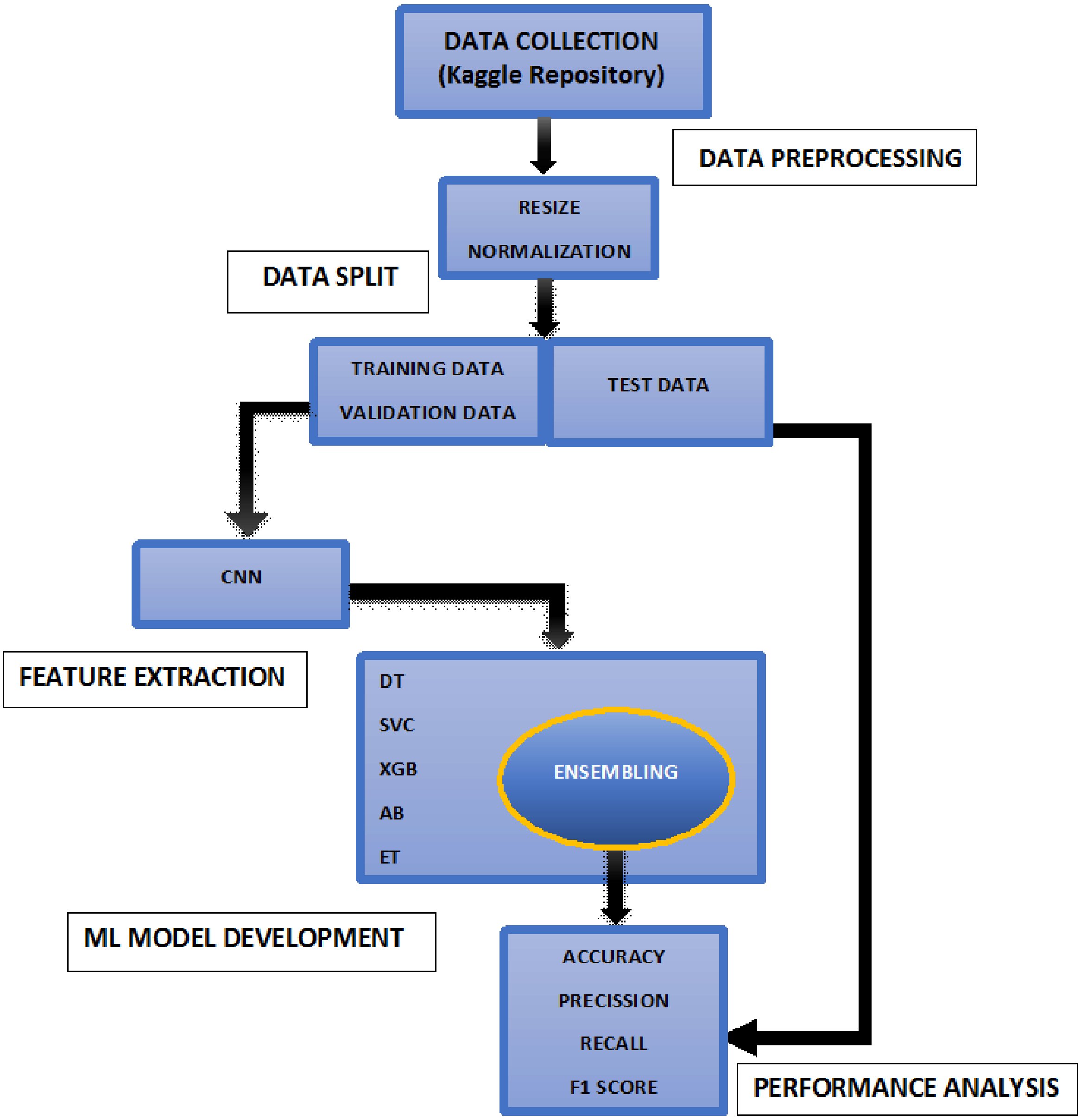

Fig. 3.1. Proposed methodology for Covid-19 Prediction

Figure 3.1 depicts the entire pipeline of a COVID-19 diagnosis system based on deep learning. Patients from the hospital are considered participants during the data gathering stage. Despite the fact that the data can take many various forms, COVID-19 is identified utilising imaging techniques such as CT and X-ray samples. The images that are collected are resized and data normalization was performed on the dataset for better learning of the system thus the dataset was thus prepared for feeding into the CNN network and for training the model and ease generalization. The data preparation stage, which entails putting the data into a usable format, is the next crucial step. Data pre-processing procedures include noise removal, scaling, and augmentation. The data partitioning approach divides the data into three sets for the experiment: training, validation, and testing. Data partitioning frequently use the cross-validation technique. The training data is used to create a specific model, which is then validated, and the model's performance is assessed using test data. Feature extraction and classification are the most important steps in a deep learning-based COVID-19 diagnosis. At this point, the deep learning system extracts the feature by repeating a number of methods, and the classification is finished with the use of class labels (healthy or COVID-19). V- Net can be used to extract various radiomics features, which requires extracting the greatest amount of features from medical images, such as size, shape, and textural aspects, while also incorporating spatial data. Finally, criteria such as accuracy, sensitivity, specificity, precision, F1-score, and others are used to evaluate the built system.

3.2DATA SET

3.2.1 DATA GATHERING

The lengthy and important process of data collection. Regardless of the research area, maintaining coherence depends on the accuracy of the data gathering. The collection of the data was a rigid and time-consuming process because patient clinical information was not publicly accessible. In order to obtain the most precise data, different hospitals and health institutes in Sweden and China were contacted, however due to the current situation at hospitals—a significant influx of patients with COVID-19—we were unable to obtain direct information. To acquire open source clinical data of patients diagnosed with COVID-19, a thorough search was made across a number of databases.

3.2.2 KAGGLE

This project will use data from the Kaggle repository, which will comprise Chest X-Ray images of Covid-19 infected, normal, and pneumonia patients. Rather than claiming the diagnostic ability of any Deep Learning model, this dataset will be compiled to evaluate multiple approaches to efficiently detecting Coronavirus infections using computer vision techniques. A normal, covid, and pneumonia training and validation set will be created from the data. Few samples of a normal case, covid, and pneumonia are considered during the validation phase. The PA view scans were determined to be consistent with our covid dataset.

In 2009, neural networks were employed for the first time for speech recognition, although Google didn't deploy them until 2012. Deep learning, often known as neural networks, is a type of machine learning that employs a computational model inspired strongly by brain structure.

"Deep learning is already in use in Google search and picture search; for example, you can use it to image-search a phrase like "hug." It's used to send Smart Replies to your Gmail account. It manifests itself in speech and vision. I believe it will soon be employed in machine translation "According to Geoffrey Hinton, the "Godfather of Neural Networks,"

Because of their multi-level structures, Deep Learning models are extremely useful in extracting sophisticated information from input photographs. Convolutional neural networks, unlike many other networks, can significantly cut calculation time by utilising graphics processing units (GPUs).

Let's take a closer look at IDP's deep learning-based image data preprocessing. To provide improved local and global feature identification, images must be prepared for subsequent analysis, which is how IDP enables straight-through processing and creates ROI for your organisation. The steps are as follows:

3.3 IMAGE CLASSIFICATION:

Using CNN to categorise images improves accuracy. First and foremost, your IDP solution will necessitate a set of pictures. In this case, the first training data set consists of pictures of cosmetics and pharmaceuticals. The most common picture data input parameters are the number of photos, image dimensions, number of channels, and number of levels per pixel.

3.4 DATA LABELING

It's preferable to label the input data manually so that the deep learning algorithm can learn to make predictions on its own in the future. Here are some off-the-shelf manual data labelling tools. At this phase, the major goal will be to identify the real object or text in a given image, determine whether the word or object is orientated incorrectly, and determine whether the script (if there) is in English or another language.

NLP pipelines can be used to automate the labelling and annotation of photographs. For non-linear activation functions, ReLU (rectified linear unit) is employed since it performs better and takes less time to train.

We may also use data augmentation to expand the training dataset by simulating and modifying existing images. We could reduce the size of the accessible photographs, blow them up, crop elements, and so on.

We may also use data augmentation to expand the training dataset by simulating and modifying existing images. We could reduce the size of the photographs provided, blow them up, crop components, and so on.

Data augmentation strategies can be used to increase the dataset volume. Common augmentation techniques such as flipping and rotating can be utilized to picture data with good results. Transfer learning is another technique for coping with insufficient data quantity while avoiding generalisation errors and over-fitting. To achieve satisfying results, a large number of studies incorporated data augmentation and transfer learning into deep learning algorithms. Few-shot learning and zero-shot learning can also be used to tackle the data insufficiency problem.

3.5 ALGORITHM CONFIGURATIONS

This section discusses the algorithms' settings and how they may be configured. Alterations made to the settings of the algorithm's setup may have an impact on the output.

- Support Vector Machines:
  SVC(kernel = ’linear’, random_state = 0)
- Convolutional Neural Network:
- Layers:
  - model.add(tf.keras.layers.Conv2D(64, kernel_size=(3,3), strides=(1,1), padding='valid', activation='relu', input_shape=(64,64,3), kernel_regularizer =tf.keras.regularizers.l1(l=0.001)))
  - model.add(tf.keras.layers.Conv2D(64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu', kernel_regularizer =tf.keras.regularizers.l1(l=0.001)))
  - model.add(tf.keras.layers.Conv2D(32, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu')))

- Compiling the CNN:

```
opt = Adam(0.001)
model.compile(loss='binary_crossentropy', optimizer=opt, metrics=['acc'])
```

- Xgboost:

```
clf = XGBClassifier(random_state=0)
```

- SVM:

```
clf = SVC (random_state=0)
```

- Decision Tree:

```
clf = DecisionTreeClassifier(random_state=0)
```

- Extra Trees:

```
clf = ExtraTreesClassifier(random_state=0)
```

- Adaboost:

```
clf = AdaBoostClassifier(random_state=0)
```

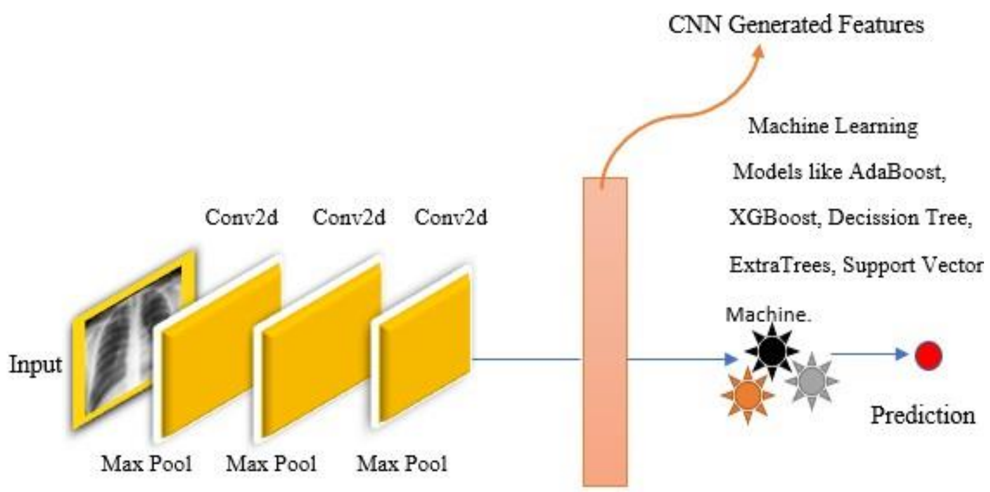

Fig.3.2 Hybrid Network Model Architecture (HN-Model)

The above figure depicts our Hybrid Network Model architecture, here the DNN Layer has only 3 layers and followed by the Machine Learning classification layer, where we have experimented with ML classification algorithms like Adaboost, Xgboost, Decision Tree, Extra Trees and SVM.

3.6 PERFORMANCE METRICS

The challenge of measuring the performance of a machine learning model is a crucial part of the process. Because our model relies on categorization, we decided to base performance on accuracy rather than any other parameter.

3.6.1 ACCURACY

In this particular experiment, the algorithms are judged according to how accurate their results are. It is the performance measure that is used the majority of the time when evaluating categorization strategies.

$$Accuracy = \frac{TP + TN}{TP + FT + FN + TN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

3.6.2 PRECISION

Precision is one of the ways in which the performance of a machine learning model can be evaluated. The precision metric evaluates how reliable a model is in providing optimistic forecasts for the foreseeable future.

$$Precision = \frac{TP}{TP + FP}$$

3.6.3 Recall

The term "recall" refers to the total number of good results that our ML model generates. Using the confusion matrix and the formula that is provided above, we are able to simply compute it.

$$Recall = \frac{TP}{TP + FN}$$

3.6.4 F1-score

This score will be used in the calculation of the harmonic mean of the accuracy and recall scores. F1's greatest value is 1 and its lowest value is 0. The F1 score may be determined by using the formula that is provided below:

$$\mathrm{F1-score} = \frac{2 \times (\mathrm{Precision} \times \mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}}$$

VALIDATION AND VERIFICATION OF PROPOSED MODEL

4.1 BEHAVIOUR OF STUDY

Currently, the detection of coronavirus disease 2019 (COVID-19) is one of the main challenges in the world, given the rapid spread of the disease. Recent statistics indicate that the number of people diagnosed with COVID-19 is increasing exponentially, with more than 1.6 million confirmed cases; the disease is spreading too many countries across the world. In this study, we analyses the incidence of COVID-19 distribution across the world. We present anmachine learning technique based to detect COVID19 patients using real-world datasets. Our system examines chest X-ray images to identify such patients.

Our findings indicate that such an analysis is valuable in COVID-19 diagnosis as X-rays are conveniently available quickly and at low costs. Empirical findings obtained from 1000 X-ray images of real patients confirmed that our proposed system is useful in detecting COVID-19 and achieves an F-measure range of 91–98%. Additionally, four forecasting methods—the SVM algorithm, CNN model, and AGboost—were adopted to predict the numbers of COVID-19 confirmations, recoveries, and deaths over days.

The prediction results exhibit promising performance and offer an average accuracy of 94.80% and 88.43% in Australia and Jordan, respectively. Our proposed system can significantly help identify the most infected criteria, and it has revealed that coastal areas are heavily impacted by the COVID-19 spread as the number of cases is significantly higher in those areas than in non-coastal areas.

4.2 CONTRIBUTION

- ❖ We propose an automated intelligent system for distinguishing COVID-19 patients from non-patients on the basis of chest X-ray images. Our system instantly reads the structure of a chest X-ray image, leverages hidden patterns to identify COVID-19 patients, and reduces the need for manual pre-processing steps.
- ❖ Empirical findings obtained from 1000 chest X-ray images of patients confirmed that our proposed system can detect COVID-19 patients with an accuracy of 92%–98%. ƒ
- ❖ We provide an intelligent prediction system for predicting the number of patients confirmed to have contracted the disease, recovered from the disease, and died from the disease over days using Support vector machine, Naive Bayes, and decision tree methods. Our proposed system has been trained and tested on datasets generated from real-world cases and has predicted the numbers of COVID-19 confirmations, recoveries.. ƒ
- ❖ We highlight the most affected part and show that coastal part are heavily impacted by COVID-19 infection and spread as the number of cases in those body parts is significantly higher than that in other non-coastal part.

Impact: After Quantification, we have moved to validation part for justification of proposed model using new methodology such as TOPSIS.

4.3 CHALLENGES IN MEDICAL HEALTH

Diagnosis and severity assessment of chronic diseases represent significant challenges to the health care system, affecting millions of patients each year. Despite remarkable progress in medical science, early detection is still a challenging task for the prevention or timely treatment of the disease.

Due to the complexity, increasing cost of the diagnostic methods, and the unavailability of adequate resources in certain areas, especially in rural areas, finding faster and efficient strategies are essential. Developing a model for the diagnosis could help physicians refine knowledge for the stratification of risk of chronic diseases concerning obscure information and categorize patients into different health patterns for effective treatments.

The following challenges (figure 4.1) motivated us to do this research work:-

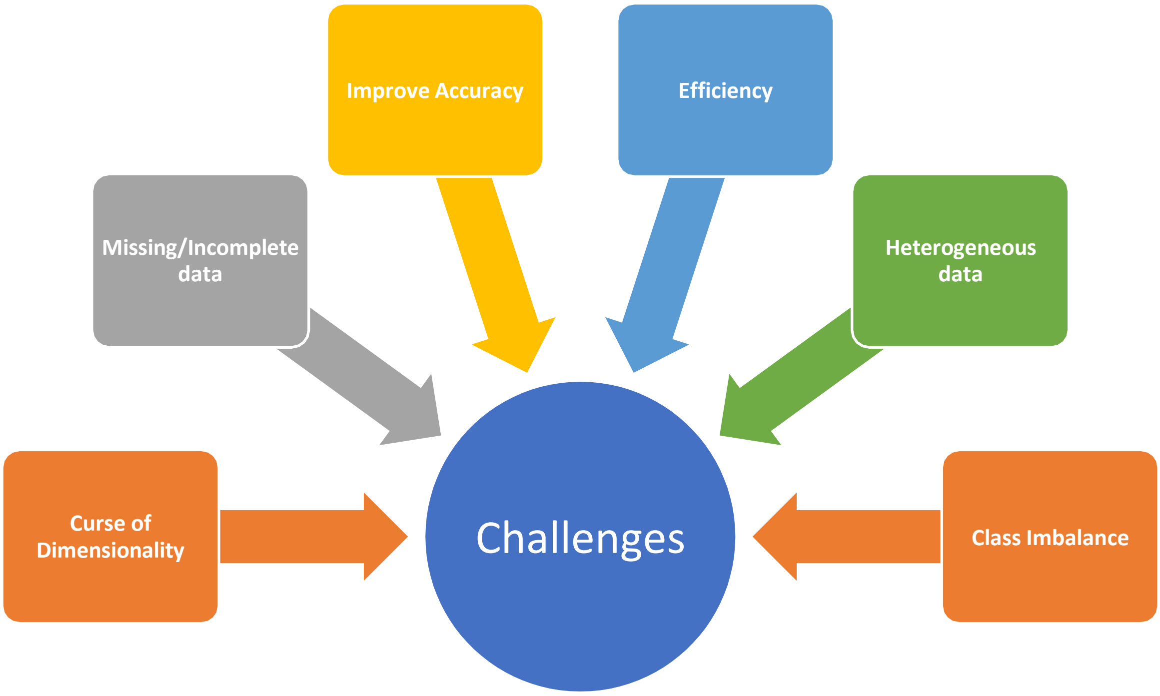

Figure 4.1: Health Challenges

- **Curse of Dimensionality** - The medical data contains a large number of features that adversely affect the performance of many of the data mining and machine learning algorithms. So, it is reasonable to keep only a limited number of input features and ignore others. It's a challenging task to decide which features to remove from the medical data to improve the mining performance.

- **Missing/Incomplete data** - This is a common problem associated with human and machine reading errors that occurred during the monitoring of the patient. There is a subtle need for effective imputation methods to fill the missing gaps.

- **Improve Accuracy -** The medical data contains a large number of irrelevant and redundant features, which decrease the performance of the classification algorithms. Dimensionality reduction methods using appropriate feature selection and feature extraction techniques can solve this problem by selecting relevant features and eliminating redundant features. Improve **Efficiency -** By applying the classification algorithm on the selected features can lead to reducing the computation time instead of using it to all the features/attributes.
- **Heterogeneous data** - Heterogeneity in medical data is one of the essential performance bottlenecks. The medical data sets generally contain a mix of different types of heterogeneous data, such as categorical data, continuous data, nominal data, and interval scaled data, etc. Effective data pre-processing techniques are required to deal with this data heterogeneity.
- **Class Imbalance -** Medical data sets sometimes suffer from class imbalance problem, where the majority of data instances belong to one target class. Appropriate under-sampling and over-sampling techniques are required to balance the target class distribution.

4.4 VALIDATION OF PROPOSED MODEL

Model verification and validation (V&V) is an enabling methodology for the development of computational models that can be used to make engineering predictions with quantified confidence. Model V&V procedures are needed by government and industry to reduce the time, cost, and risk associated with full-scale testing of products, materials, and weapon systems. Quantifying the confidence and predictive accuracy of model calculations provides the decision-maker with the information necessary for making high-consequence decisions. The development of guidelines and procedures for conducting a model V&V program are currently being defined by a broad spectrum of researchers.

Model verification and validation are the primary processes for quantifying and building credibility in numerical models. Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and its solution. Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. Both verification and validation are processes that accumulate evidence of a model's correctness or accuracy for a specific scenario; thus, V&V cannot prove that a model is correct and accurate for all possible scenarios, but, rather, it can provide evidence that the model is sufficiently accurate for its intended use.

Machine learning validation and system verification are crucial for ensuring the quality of machine learning applications. However, a rigorous technique for such tasks is yet to emerge. During the past decade, we have developed a machine learning system for investigating the classification of biological cells based on cell morphology which is captured in diffraction images. Proposed Model includes a collection of scientific software tools, machine learning algorithms, and a large-scale cell image repository.

In order to ensure the quality of machine learning system proposed model, we developed a framework for rigorously validating the massive scale image data as well as adequately verifying both the software tools and machine learning algorithms. The validation of machine learning is conducted by iteratively selecting the data using a machine learning approach. An experimental approach guided by a feature selection algorithm is introduced in the framework to select an optimal feature set for improving the machine learning performance.

The verification of software and algorithms is developed on the iterative metamorphic testing approach due to the non-testable property of the software and algorithms. A machine learning approach is introduced for developing test oracles iteratively to ensure the adequacy of the test coverage criteria in below figure.

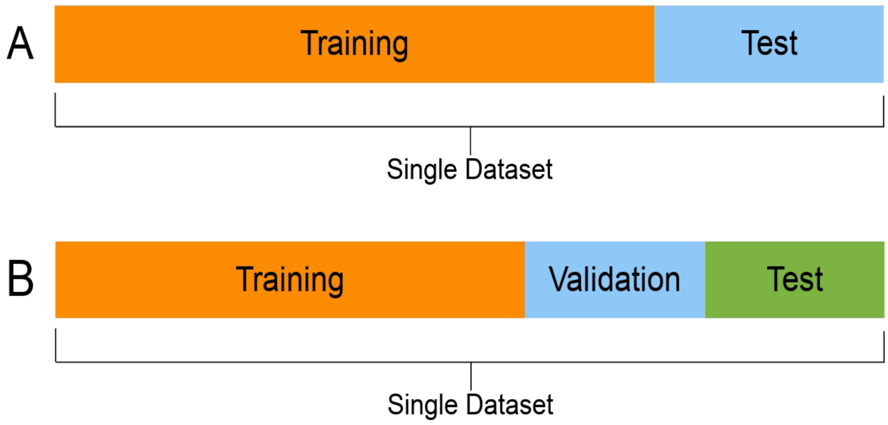

Fig .4.2 validation of proposed model

Performance of the machine learning algorithm is evaluated with a stratified TOPSIS validation and confusion matrix. We describe the design of the proposed machine learning verification and validation framework with proposed model as the case study, and demonstrate its effectiveness through verifying and validating the dataset, the software and the algorithms in proposed model.

4.5 NEED FOR MODEL VALIDATION

Model validation is a set of processes and activities designed to ensure that an ML/AI model is performing as it should, including both its design objectives and its utility for the end user. An important part of validation is testing the model, but validation doesn't end there. As an integral part of model risk management, validation is designed to ensure that the model doesn't create more problems than it solves, and that it conforms to governance requirements. In addition to testing, the validation process includes examining the construction of the model, the tools used to create it and the data it used, to ensure that the model will run effectively.

After a model has been trained, a process is required to ensure that the model is performing the way it was intended and that it solves the problem it was designed to solve. This is the purpose of model validation. It's important that the validation be done in an unbiased manner. For this reason, the validation team is usually independent of the data-science team that trained the model and those who will be using the model. Often, smaller organizations will contract model validation out to a third party. In some sectors that are highly regulated, this is often a team that is familiar with regulations, to ensure compliance.

4.6 TOPSIS

Multicriteria decision-making (MCDM) or multiple-criteria decision analysis is an important branch of operations research that definitely uses multiple-criteria in decision making environments. In daily life and professional learning, there exist generally multiple conflicting criteria which need to be considered in making decisions and optimization. Price and spend are typically one of the main criteria with regard to a large amount of practical problems.

However, the factor of quality is generally another criterion which is in conflict with the price. For example, the cost, safety, fuel economy, and comfort should be considered as the main criteria upon purchasing a car. It is the most benefit for us to select the safest and most comfortable one which has the bedrock price simultaneously.

The best situation is obtaining the highest returns while reducing the risks to the most extent with regard to portfolio management. In addition, the stocks that have the potential of bringing high returns typically also carry high risks of losing money. In service industry, there is a couple of conflicts between customer satisfaction and the cost to provide service. Upon making decision, it will be compelling if multiple-criteria are considered even though they came from and are based on subjective judgment of human.

What is more, it is significant to reasonably describe the problem and precisely evaluate the results based on multiple-criteria when the stakes are high. With regard to the problem of whether to build a chemical plant or not and where the best site for it is, there exist multiple-criteria that need to be considered; also, there are multiple parties that will be affected deeply by the consequences. Constructing complex problems properly as well as multiple-criteria taken into account explicitly results in more reasonable and better decisions.

E. K. Zavadskas used elaborate interview techniques to deal with the problem in MCDM, which exist for eliciting linear additive utility functions and multiplicative nonlinear utility functions [77]. And there are many other methods, such as best worst method, characteristic objects method, fuzzy sets method, rough sets, and analytic hierarchy process **(W. Jiang, 2015).**

In G. Lee (2015), the authors aim to systematically review the applications and methodologies of the MCDM techniques and approaches, which is a good guidance for us to fully understand the MCDM. Technique for order preference by similarity to ideal solution (TOPSIS), which is proposed in **(D. Liang, 2015),** is a ranking method in conception and application [78].

The standard TOPSIS methodology aims to select the alternatives which have the shortest distance from the positive ideal solution and the longest distance from the negative ideal solution at the same time. The positive ideal solution maximizes the benefit attributes and minimizes the cost attributes, whereas the negative ideal solution maximizes the cost attributes and minimizes the benefit attributes.Which are widely used in the real decision situations (**Yang and Hung, 2007).** TOPSIS serves as one of the models that compromise subgroups subgroup subgroups compensation model and is itself adaptive [79].

The compensation model allowed the exchange between indices, for example, an indicator of weakness may be offset by other index score **(B. L. Golden et al., 2003).** Yang and Huang offered similar method for the best ideal solution [80]. This means that the option should be the shortest distance from the positive ideal solution and at the same time farthest from the ideal is negative **(Calabrò et al., 2014).**

Suppose desirability index is rising steadily or in other words only index are positive or negative aspects. The index of the positive aspects of profit and cost index, which has a negative aspect. It is easy to determine the ideal solution [81]. Therefore, the current value of the index indicates a positive ideal and the worst value of that particular ideal would be a negative (Sheng et al., 2002).

It is an approximation of the geometric point of view an option to be considered the minimum distance from the positive ideal solution and farthest from the negative ideal solution **(Dasturani, 2012).** TOPSIS assesses both distance option ideal solution both positive and negative ideal solution by the relative closeness to the ideal solution [82]. In fact TOPSIS a strong decision making method using qualitative and quantitative criteria for prioritizing by similarity and proximity to the ideal answer.

The option must be the shortest distance from the ideal answer. This method is useful when faced with a number of quantitative and qualitative factors [83]. The overlap some of the criteria in this way have any effect on application logic and conclusions intact **(Kermani et al., 2016)**. TOPSIS take into account information in a way that takes into account a set of weights for the desired criteria.

The answer depends on the weighting scheme that is given by the decision maker. Fortunately some reliable methods for evaluating the weights have been identified that will increase the desirability TOPSIS **(Leon et al., 2005).**

4.6.1 TOPSIS METHOD FOR A SINGLE DECISION MAKER:

Step 1: Construct the decision matrix and determine the weight of criteria. Let (X_{ij}) = x be a decision matrix and [W = $w_1, w_2 \ldots\ldots\ldots w_n$] a weight vector, where $x_{ij} \in \Re$,$w_j \in \Re$ and 1. w_1 + w_2 + + w_n = Criteria of the functions can be: benefit functions (more is better) or cost functions (less is better).

Step 2: Calculate the normalized decision matrix. This step transforms various attribute dimensions into non-dimensional attributes which allows comparisons across criteria. Because various criteria are usually measured in various units, the scores in the evaluation matrix X have to be transformed to a normalized scale. The normalization of values can be carried out by one of the several known standardized formulas. Some of the most frequently used methods of calculating the normalized valuen_{ij} are the following:

$$n_{ij} = \frac{n_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}$$

$$n_{ij} = \frac{x_{ij}}{\max x_{ij}}$$

$$n_{ij} = \frac{\dfrac{x_{ij} \min\limits_{i} x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}} \, if\ c_i\ is\ a\ benefit\ criterion}{\dfrac{\max\limits_{i} x_{ij} - x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}} \ if\ c_i\ is\ a\ cost\ criterion}$$

for i=1,…m; j =1,…,n.

Step 3. Calculate the weighted normalized decision matrix.

The weighted normalized value v_{ij} is calculated in the following way:

$v_{ij} = w_j n_{ij}$ for i=1,… m; j=1,…,n.

where w_j is the weight of the j-th criterion, $\sum_{j-1}^{n} w_j = 1$

Step 4. Determine the positive ideal and negative ideal solutions.

Identify the positive ideal alternative (extreme performance on each criterion) and identify the negative ideal alternative (extreme performance on each criterion). The ideal positive solution is the solution that maximizes the benefit criteria and minimizes the cost criteria whereas the negative ideal solution maximizes the cost criteriea and minimizes the benefit criteria.

Positive ideal solution A^+ has the form:

$$A^+ = (v_1^+, v_2^+, \dots, v_n^+) = (\max_i v_{ij} \mid j \in I), (\min_i v_{ij} \mid j \in J)$$

Negative ideal solution. A^- has the form:

$$A^- = (v_1^-, v_2^-, \dots, v_n^-) = (\min_i v_{ij} \mid j \in I), (\max_i v_{ij} \mid j \in J)$$

Where I is associated with benefit criteria and J with the cost criteria, i=1,…,m, j=1,…,n.

Step 5. Calculate the separation measures from the positive ideal solution and the negative ideal solution.

In the TOPSIS method a number of distance metrics can be applied*.

The separation of each alternative from the positive ideal solution is given as

$$d_i^+ = \left(\sum_{j-1}^{n} (v_{ij} - v_j^+)^2\right)^{\frac{1}{p}}, \qquad i=1,2,\dots,m.$$

The separation of each alternative from the negative ideal solution is given as

$$d_i^- = \left(\sum_{j-1}^{n} (v_{ij} - v_j^+)^2\right)^{\frac{1}{p}}, \qquad i=1,2,\dots,m.$$

$where\ p \geq 1$, for p=2 we have the most used traditional n – dimensional Euclidean metric.

$d_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^+)^2}$, i=1,21…,m, $d_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2}$, i=1,21…,m.

A proposed validation intends to choose the best from the set of models observation submitted by model. Four methods evaluate four issues using several criteria. In order to simplify the calculation, only four criteria are considered: Accuracy, Confusion Matrix, Prediction, Classification, respectively.

Based on Table 4.1, the decision matrixes for assigned weight are obtained.

Table 4.1: Assign Weight

Weight	0.92	0.95	0.88	0.92
	Accuracy	**PRECISSION**	**RECALL**	**F1-score**
XGB	1.6	1.9	2.5	2.9
DT	3.5	4.4	5.6	6.2
ADA	6.5	9.5	8.3	8.7
ET	1.9	4.2	3.1	2.1
SVC	2.5	3.6	7.9	8.2

Using the methods mentioned in the Materials and Methods and results obtained, decision matrix indicator is expressed in Table 4.1.The next step is to extract the relative importance implied by the table 4.1. That is, how important are the three alternatives when they are considered in terms of the proposed model criteria.In Given a table 4.2 with pairwise comparisons, the corresponding maximum weight criteria, normalized evaluation. An alternative approach for evaluating the relative priorities from a judgment matrix is based on a least squares formulation and is described between rank and weight. In the TOPSIS the pairwise comparisons in a table 4.2 are considered to be adequately consistent if the values are near to 1.

Table 4.2: Fuzzy Range for Criteria

Term	Fuzzy Number
V Low	1,1,3
Low	1,3,5
Avg	3,5,7
High	5,7,9
V High	7,9,9

The first evaluative criterion (table 4.1) has to do with the premise that a method which is accurate in multi-dimensional problems should also be accurate in single-dimensional problems. There is no reason for an accurate multi-dimensional method to fail in giving accurate results in single-dimensional problems, since single-dimensional problems are special cases of multi-dimensional ones. Because the first method for COVID- 19, gives the most acceptable results for the majority of single-dimensional problems, the result of the prediction was used as the standard for evaluating the other three methods in this context.

Table 4. 3 Calculate Normalized Matrix

	Accuracy	PRECISSION	RECALL	F1-score
XGB	0.211813726	0.178563227	0.242250792	0.261974524
DT	0.463342525	0.413514841	0.542641773	0.560083465
ADA	0.86049326	0.892816135	0.804272628	0.785923572
ET	0.24924598	0.372323454	0.295761675	0.192879838
SVC	0.346321515	0.327475771	0.665512449	0.675543811

This step transforms various attributes as Accuracy, Confusion Matrix, Prediction, Classification, dimensions into non-dimensional attributes which allows comparisons across Low, Mid and High. In table 4.3, normalized values of each criterion have quantified through above equation **(step 2).** Here i and j (step 2) are represented to criteria and alternatives segment. Normalized metrics values are more effective to further quantification.

$$\boldsymbol{nij} = \frac{\boldsymbol{xij}}{\sqrt{\sum_{i=1}^{m} \boldsymbol{xij}}}$$

$$\boldsymbol{nij} = \frac{\boldsymbol{xij}}{\mathbf{max}\, \boldsymbol{xij}}$$

Table 4. 4 Calculate the Normalized Weighted Matrix

	Accuracy	PRECISSION	RECALL	F1-score
XGB	0.194868628	0.169635066	0.213180697	0.241016562
DT	0.426275123	0.392839099	0.47752476	0.515276788
ADA	0.791653799	0.848175328	0.707759913	0.723049686
ET	0.229306302	0.353707281	0.260270274	0.177449451
SVC	0.318615793	0.311101983	0.585650955	0.621500306

Among the seven parameters discussed above, each has its own prominence. Every parameter will have different effects on the prospective indices of alternatives and hence cannot be allotted equal weights which shown in above equation (step 3). Thus, it becomes important to find out the primacies of each parameter. **In table 4.3**, we had been used for calculating an assured weightage for each parameter to define the individual prominence in a numeric value.

Table 4.5: Calculate the Euclidean from Ideal best

Si+	Si-
0.805454815	0.555786257
0.49703533	0.470297439
0.657396368	0.917275063
0.787337161	0.452991152
0.619374844	0.48807683

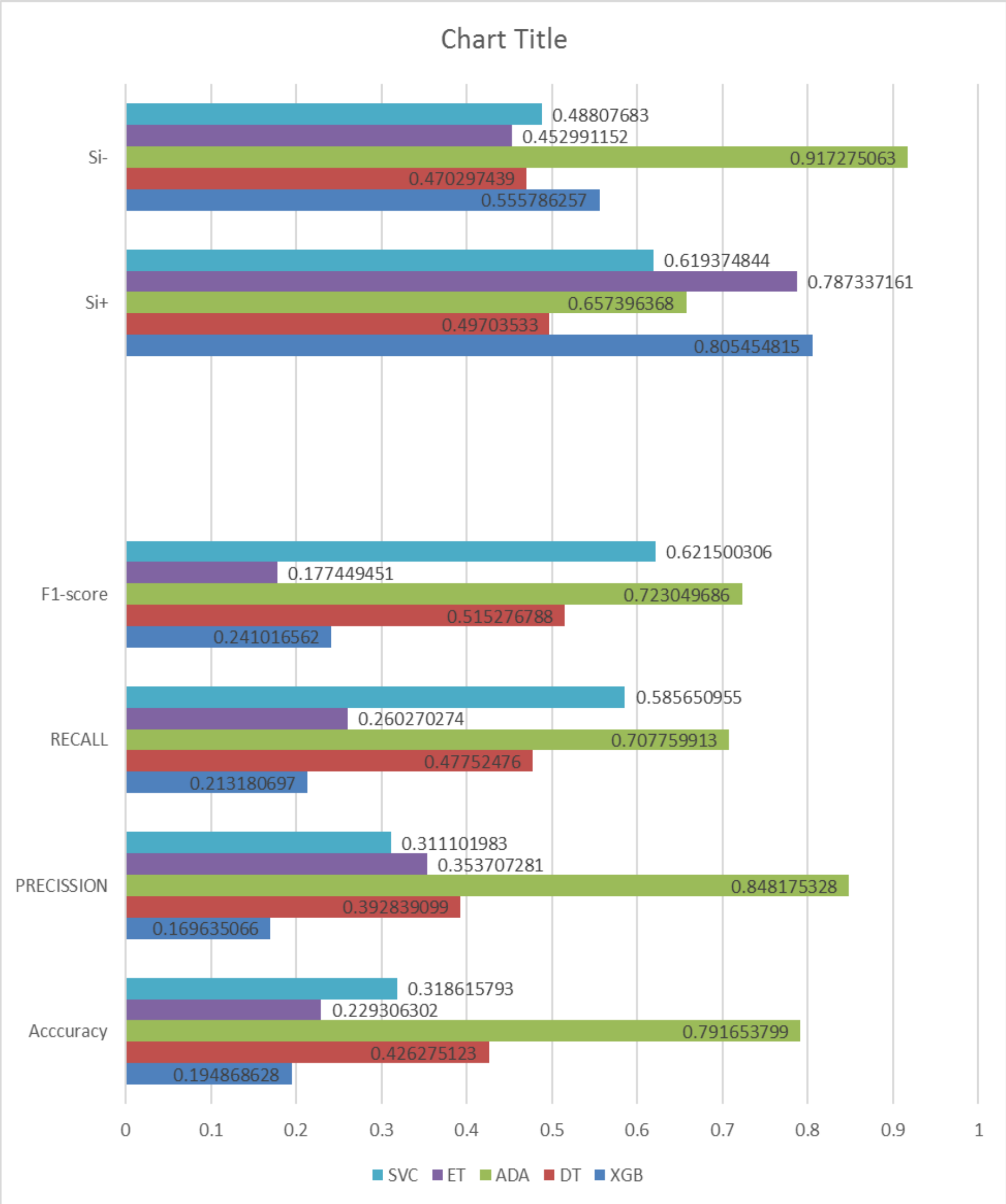

Figure 4.3: Graphical Structure

The normalized weighted decision matrixes are constructed as in **table 4.5**. By using step 4, maximum and minimum of each column are determined as Vi+, and Vi- respectively shown in figure 4.3. By using step 4, distance between the possible alternative v_{ij} and the positive ideal solution Vi $_{Max}$ and the negative ideal solution $V_{i}min$ are calculated and are depicted in tables 4.5.

$$v_i^+ = ((v)_1^+, (v)_2^+ - - - - - (v)_n^+) = \{(\max\langle v_{ij} \mid j \in I\rangle), (\min\langle v_{ij} \mid j \in J\rangle)\}$$

$$v_i^- = ((v)_1^-, (v)_2^- - - - - - (v)_n^-) = \{(\min\langle v_{ij} \mid j \in I\rangle), (\max\langle v_{ij} \mid j \in J\rangle)\}$$

Table 4.5: Calculate the Ideal best and Ideal worst Value

V+	0.791653799	0.415091751	0.213180697	0.723049686
V-	0.194868628	0.415091751	0.707759913	0.177449451

Figure 4.3: Graphical Structure of Table 4.5

In this section fuzzy TOPSIS is proposed for the reliable obstacle selection problem. Firstly, importance of criteria by using linguistic variables is evaluated decision makers shown in fig 4.2 and importance weights of the criteria determined by these three decision makers are shown in Table .42, To evaluate the rating of alternatives with respect to each criterion three decision makers use variables shown in **Table 4.2 and 4.3** and under criteria rating of five alternatives are shown in Table 4.5. To construct decision matrix and fuzzy weights of three alternatives shown in previous chapter, aggregate TOPSIS of alternative with respect to each criterion is calculated by using **(equation 5.1, chapter 5).** In this chapter, results of TOPSIS i.e., second alternative is preferred compared to other alternatives. So, we can say that obstacle detectionof proposed model is the most reliable version of roadside sensor.

TOPSIS outcome is preferred because on comparing criteria in terms of number of computations TOPSIS requires less computation than alternatives and works well for one-tier decision tree which is preferable for widespread hierarchies. In TOPSIS increase in number of criteria and alternative increases risk and pair wise comparison with reliability factors for obstacle valuations. The comparative experimental results indicates that TOPSIS scheme has higher reliable and many methods (Naïve Byes, CNN and SVM) accuracy rate for Covid detection and can recognize the main obstacles in the obstacle detection task. In addition, we further evaluate and test the scheme by using depth information, space layout information and their combination. The experiment starts from lane detection and continuing to detect COVID. **In fig. 4.3,** shows the comparison result of comparative chart of $V_i$value.

5.1 RESULTS & COMPARATIVE STUDY

This chapter presents the results that are achieved from this experiment. We evaluated the performance measures in terms of Accuracy, Precision, F1-score and Recall. In order to have some faster algorithms and better performance we hybridized the algorithms.

In this experiment, we have used five different models that are XGBoost, Decision Tree, Adaboost, Extra Trees and Support Vector Classifier and performed hybridization method.

5.1.1 APPLYING HN-SVM

So now we after using SVM along with CNN. Support vector machines, often known as SVMs, are a family of supervised learning algorithms that can categories data, carry out regression analysis, and locate outliers. The advantages of support vector machines include the following: efficient in high-dimensional situations. Maintains its utility even in circumstances in which the number of dimensions is greater than the number of samples. Table 5.1 shows the performance measures that this model has achieved. The HN-SVM model achieved the accuracy 82%, precision of 77%, recall of 91% and f1-score of 83%.

Table 5.1 Performance results (%) of HN-SVM model

	PRECISION	RECALL	F1-score	support
0.0	0.79	0.78	0.78	259
1.0	0.77	0.78	0.78	250
Accuracy			0.78	509
Macro-avg	0.78	0.78	0.78	509
Weighted-avg	0.78	0.78	0.78	509

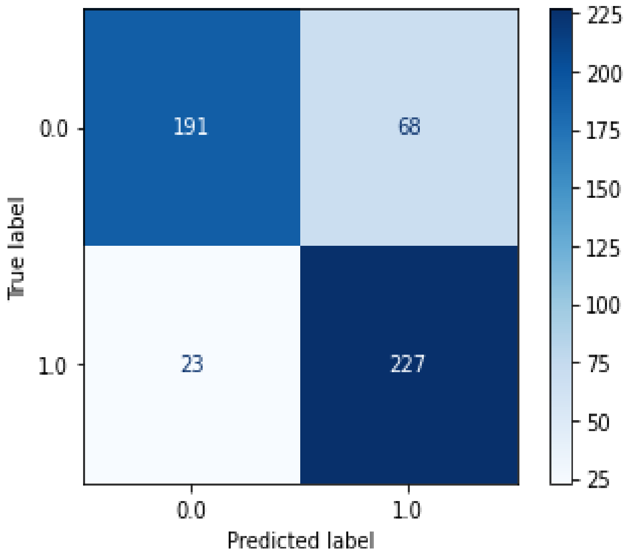

Fig. 5.1. Confusion matrix representation for HN-SVM model

5.1.2 APPLYING HN-DECISION TREE

A decision tree is a technique for supervised machine learning that may be used to categorise or predict data based on the responses that have been provided in the past to questions that have been asked. The nature of the model is supervised learning, which indicates that it is trained and evaluated using data sets that include the necessary classification in order to get the desired results. This 5.2 Table shows the performance evaluation of this model. It has achieved the accuracy of 78%, precision of 77%, recall of 78% and f1-score of 78%.

Table 5.2. Performance results (%) of HN-Decision Tree model

	PRECISION	RECALL	F1-score	support
0.0	0.89	0.74	0.81	259
1.0	0.77	0.91	0.83	250
Accuracy			0.82	509
Macro-avg	0.83	0.82	0.82	509
Weighted-avg	0.83	0.82	0.82	509

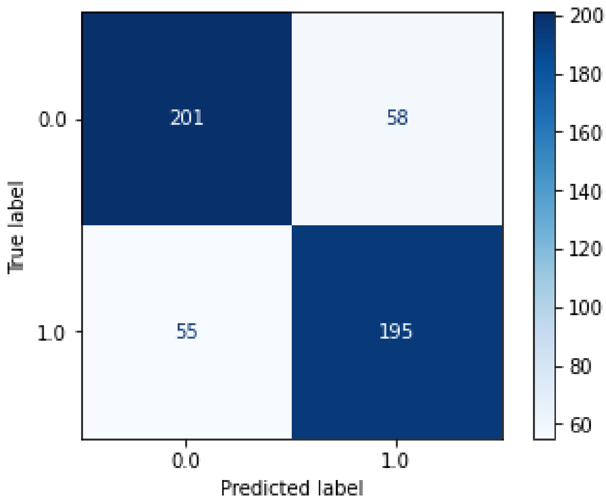

Fig 5.2 Confusion matrix representation for HN-Decision Tree model

5.1.3 APPLYING HN-EXTRATREES

Extremely Randomized Trees, often known as Extra Trees, is a machine learning ensemble algorithm. Applying CNN along with Extra Trees is what our HN-ExtraTrees model does. It is an ensemble of decision trees specifically, and it is connected to other ensembles of decision tree techniques like bootstrap aggregation (bagging) and random forest. Table 5.3 shows the performance evaluation of this model. It has achieved the accuracy of 88%, precision of 85%, recall of 93% and f1-score of 89%.

Table 5.3 Performance results (%) of HN-Extra Trees model

	PRECISSION	RECALL	F1-score	support
0.0	0.92	0.84	0.88	259
1.0	0.85	0.93	0.89	250
Accuracy			0.88	509
Macro-avg	0.89	0.88	0.88	509
Weighted-avg	0.89	0.88	0.88	509

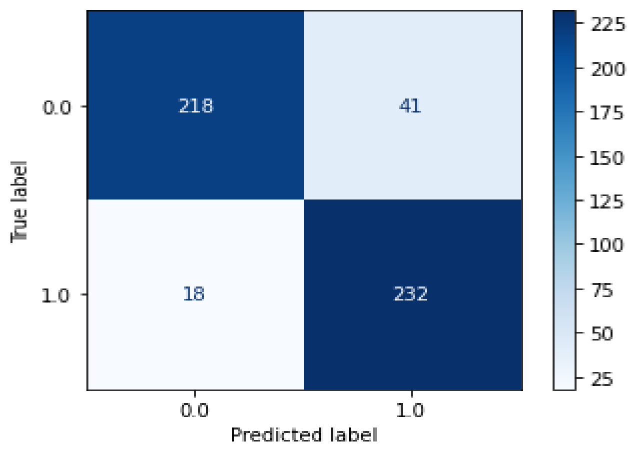

Fig 5.3 Confusion matrix representation for HN-Extra Trees model

5.1.4 APPLYING HN-ADABOOST

AdaBoost is an ensemble learning technique that was initially developed to boost the performance of binary classifiers (sometimes referred to as "meta-learning"). AdaBoost uses an iterative process to improve poor classifiers by learning from their errors. Table 5.4 shows the performance evaluation of this model. It has achieved the accuracy of 84%, precision of 83%, recall of 86% and f1-score of 84%.

Table 5.4 Confusion Matrix representation for HN-Adaboost model

	PRECISION	RECALL	F1-score	support
0.0	0.86	0.83	0.84	259
1.0	0.83	0.86	0.84	250
Accuracy			0.84	509
Macro-avg	0.84	0.84	0.84	509
Weighted-avg	0.84	0.84	0.84	509

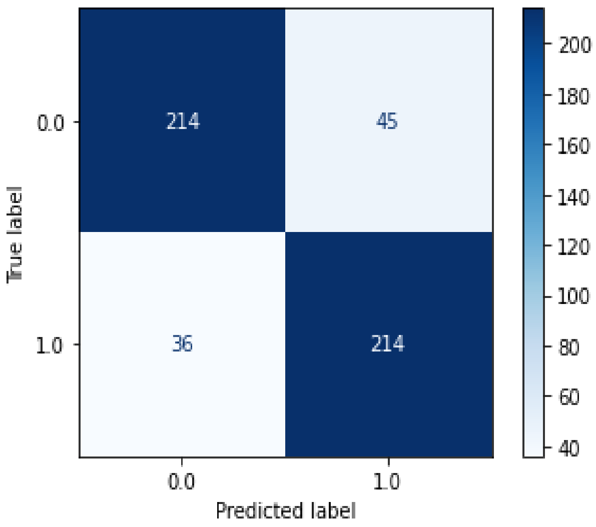

Fig. 5.4 Confusion Matrix representation for HN-Adaboost model

5.1.5 APPLYING HN-XGBOOST

The Extreme Gradient Boosting framework, often known as XGBoost, is a gradient-boosted decision tree (GBDT) machine learning system that is scalable and distributed. It provides parallel tree boosting and is often considered to be the best machine learning library for solving problems including regression, classification, and ranking. Xgboost when applied together with CNN, in a hybrid model, achieves the highest accuracy of 92%, precision of 89%, recall of 95% and f1-score of 92%. Table 5.5 shows the performance measures of every set of records achieved by HN-Xgboost model.

Table 5.5 Performance results (%) of HN-Xgboost model

	PRECISION	RECALL	F1-score	support
0.0	0.95	0.88	0.92	259
1.0	0.89	0.95	0.92	250
Accuracy			0.92	509
Macro-avg	0.92	0.92	0.92	509
Weighted-avg	0.92	0.92	0.92	509

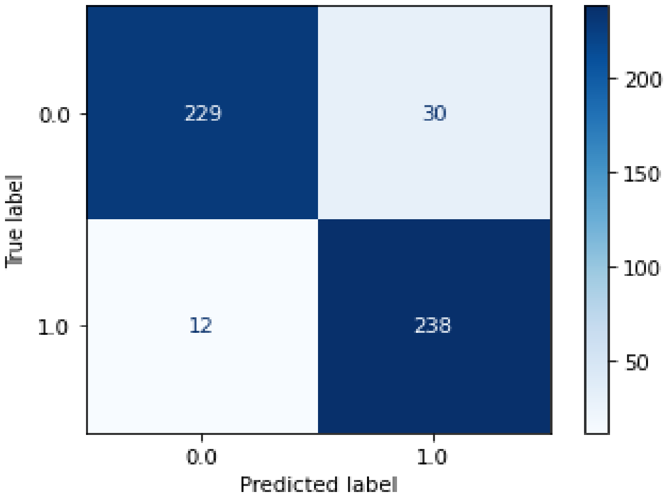

Fig. 5.5 Confusion matrix representation for HN-Xgboost model

COMPARISON PLOT

Below fig. depicts the comparison chart drawn with respect to the different models used in this hybridization method. The highest value of the accuracy is acquired by Xgboost model.

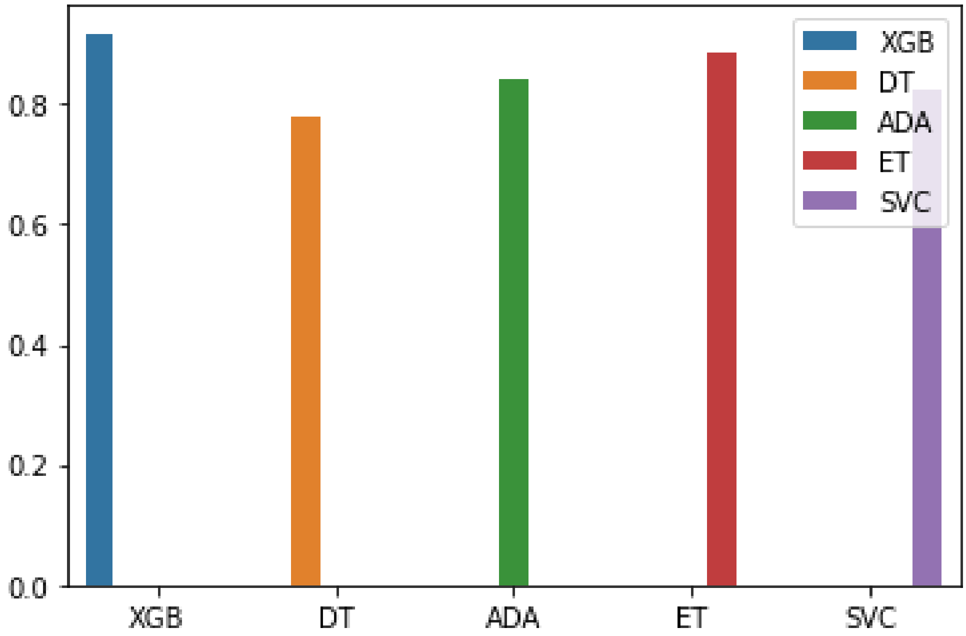

Fig:5.6 Comparison plot of all the models used.

Table 5.6 Comparison drawn along with recent works

Authors	Dataset Details	Partitioning	Techniques	Performance (%)
Wu et al. [86]	495 (COVID-19 =368, other pneumonia=127)	Training=80% Validation=10% Testing=10%	ResNet50	Accuracy=76, Sensitivity=81.1, Specificity=61.5, AUC=81.9
Hemdan et al. [87]	26(COVID-19, Pneumonia)	Training=80% Testing=20%	COVIDX-Net	Accuracy= 90% Precision=83% Recall=100% F1-score=91%
Zhang et al. [88]	1531 (COVID-19 = 100, Normal = 1431)	-	18 layer residual CNN	Accuracy=72.31% Sensitivity=90% Specificity=87.84%

Wang et al. [89]	1065 (COVID-19=740, COVID-19=325)	Random partition-ing	Model Inception	Accuracy=79,3, Sensitivity=83, Specificity=67, Precision=55, NPV=90, F1-Score=63, AUC=81, Kappa=48, Yoden index=50
He et al. [90]	746 (COVID-19=349, non-COVID-19=397)	Training=60% Validation=15% Testing=25%	CRNet	Accuracy=86, F1-Score=85, AUC=94
Zheng et al. [91]	630 (COVID-positive, COVID-negative)	Training=80% Testing=20%	DeCoVNet	Accuracy=90.1, Sensitivity=90.7, Specificity=91.1, Precision=84, NPV=98.2, AUC=95.9
Amyar et al. [92]	1044 (COVID-19=449, non-COVID-19=595)	Training=80% Validation=10% Testing=10%	Encoder-Decoder with multi-layer perception	Accuracy=86, Sensitivity=94, Specificity=79, AUC=93
HN-Xgboost	2529 (COVID-19=1262, non-COVID=1267)	Training=80% Testing=20%	Cnn + Hybrid Model	Accuracy=92% Precision=95% Recall=88% F1-score=92%

CONCLUSION

The COVID-19 epidemic is currently causing severe suffering throughout the world. Numerous people have already passed away as a result of inadequate care, inadequate facilities, or a lack of early detection. This hybrid model can aid in the automatic detection of COVID-19 from chest X-ray pictures in infected patients. It takes high-level characteristics from X-ray pictures using CNN. With great accuracy, the ensemble model distinguishes between COVID-19 and normal cases. Compared to other recent studies, the dataset has a substantial number of COVID-19 photos. An thorough experiment demonstrates that a hybrid model that combines the Xgboost and deep learning algorithms performs better, with 92 percent accuracy, 95 percent precision, 88 percent recall, and 92 percent F1-score.

epidemic. Taking into account all of these experimental findings, the Xgboost model, when combined with a deep learning algorithm, can be a valuable tool for doctors and a potential replacement for manual radiography analysis in the automatic detection of COVID-19 in the current epidemic. Although it has significant drawbacks (such as the potential to incorrectly label some COVID-19-positive cases as negative), it can serve as an alternative to manual radiological analysis and help medical professionals detect COVID-19 automatically from chest X-ray pictures. It has a great deal of potential for reducing the strain on front-line doctors and nurses, enhancing early diagnosis and treatment, and assisting in epidemic management.

FUTURE WORK

Machine learning has a lot of potential in the healthcare industry. It is advised to focus future study on calibrated and ensemble approaches that could address peculiar issues more quickly and effectively than the current algorithms. Additionally, an AI-based application can be created employing a variety of sensors and features to recognise and assist with the diagnosis of diseases.

The article suggests a methodology for early detection and prediction of COVID-19 suspects in order to prevent infection and mortality. The suggested method uses machine learning to identify and predict COVID-19 suspects by gathering data from the Kaggle repository. A actual dataset obtained from the Kaggle repository is used to evaluate the framework using machine learning methods. We employed the support vector machine, decision tree, naive bayes, logistic regression, and neural network among the five proposed machine learning techniques. According to the testing findings, XGBoost has surpassed 91 percent accuracy. The decision tree has a success rate of 75%, the support vector machine has a success rate of 86%, and Extra Trees have a success rate of 89%. The ADaboost has an accuracy rate of 86%. Through an early detection and prediction system, the suggested framework has the ability to stop and slow the spread of illness. Healthcare practitioners can readily access the cloud-stored data to further analyse it and gain new insights into the nature of sickness. In the future, we'll concentrate on proposing ensemble techniques to train our algorithms, like random forest and other gradient boosting algorithms. The dataset utilised in the work mentioned above is not so large that ensemble learning or other techniques would be beneficial. Deep learning approaches will also be tested in order to improve the model's performance metrics.

REFERENCES

1. [1] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., Author correction: a new coronavirus associated with human respiratory disease in China, Nature 580 (7803) (2020). E7–E7.
2. [2] M.S. Satu, M.I. Khan, M. Mahmud, S. Uddin, M.A. Summers, J.M. Quinn, M. A. Moni, Tclustvid: a novel machine learning classification model to investigate topics and sentiment in covid-19 tweets, Knowl. Base Syst. 226 (2021) 107126.
3. [3] Worldometer, Covid-19 Coronavirus Cases, 2021. https://www.worldometers.info/coronavirus/coronavirus-cases. accessed: 2021-08-05.
4. Tang YW, Schmitz JE, Persing DH, Stratton CW. Laboratorydiagnosis of COVID-19: current issues and challenges. J Clinical Microbiol. 2020;58(6).
5. Zhang JJ, Cao YY, Dong X, Wang BC, Liao MY, Lin J, etal. Distinct characteristics of COVID-19 patients with initial rRTPCR-positive and rRT-PCR-negative results for SARS-CoV-2. Allergy. 2020;75(7):1806—1812.
6. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). Clin Infect Dis. 2020;71(15):778–785.
7. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, MeredithHR, et al. The incubation period of coronavirus disease 2019(COVID-19) from publicly reported confirmed cases: estimation and application. Annals Int Med. 2020;172(9):577–582.
8. X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, ‘‘Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: Relationship to negative RT-PCR testing,’’ *Radiology*, vol. 296, no. 2, pp. E41–E45, Aug. 2020.

9. G. Lippi and M. Plebani, “A six-sigma approach for comparing diagnostic errors in healthcare—where does laboratory medicine stand?” *Annals of Translational Medicine*, vol. 6, no. 10, 2018.
10. G. Lippi, A.-M. Simundic, and M. Plebani, “Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19),” *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 58, no. 7, p. 1, 2020.
11. C. Sheridan, “Coronavirus and the race to distribute reliable diagnostics,” *Nature Biotechnology*, vol. 38, no. 4, p. 382, 2020.
12. O. Wolach and R. M. Stone, “Mixed-phenotype acute leukemia,” *Current Opinion in Hematology*, vol. 24, no. 2, pp. 139–145, 2017.
13. J. P. Kanne, ‘‘Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: Key points for the radiologist,’’ *Radiology*, vol. 295, no. 1, pp. 16–17, Apr. 2020.
14. G. D. Rubin, C. J. Ryerson, L. B. Haramati, and N. Sverzellati, ‘‘The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner Society,’’ *Radiology*, vol. 296, no. 1, pp. 172–180, Jul. 2020.
15. D. Dong *et al.*, ‘‘The role of imaging in the detection and management of COVID-19:A review,’’ *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 16–29, Jan. 2021.
16. H. Kim, H. Hong, and S. H. Yoon, ‘‘Diagnostic performance of CT and reverse transcriptase polymerase chain reaction for coronavirus disease 2019: A meta-analysis,’’ *Radiology*, vol. 296, no. 3, pp. E145–E155, Sep. 2020.
17. Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, ‘‘Chest CT manifestations of new coronavirus disease 2019 (COVID-19): A pictorial review,’’ *Eur. Radiol.*, vol. 30, no. 8, pp. 4381–4389, Aug. 2020.

18. Xu B, Meng X A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)

19. Ajlan AM, Ahyad RA, Jamjoom LG, Alharthy A, Madani TA (2014) Middle East respiratory syndrome coronavirus (MERS-CoV) infection: chest CT findings. Am J Roentgenol 203(4):782–787.

20. Kanne JP (2020) Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist.

21. Rajpurkar, Pranav. In: (2017). URL: https://arxiv.org/pdf/1711.05225.pdf.

22. 2020. URL: https://canvas.kth.se/courses/17663/pages/lectures.

23. 2020. URL: https://canvas.kth.se/courses/17088/pages/online-educationalresources-about-deep-learning.

24. CR Rao and Venkat N Gudivada. Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications. Elsevier, 2018.

25. Mathkunti NM, Rangaswamy S. Machine learning techniques to identify dementia. SN Comput Sci. 2020;1:118. https ://doi. org/10.1007/s4297 9-020-0099-4.

26. Mitchell T (1997) Machine learning. McGraw Hill0–07–042807–7

27. Rustam F, et al. COVID-19 future forecasting using supervised machine learning models. IEEE Access. 2020. https ://doi. org/10.1109/acces s.2020.29973 11.

28. Sadiq H, Muhammad LJ, Yakubu A. Mining social media and DBpedia data using Gephi and R. J Appl Comput Sci Math. 2018;12(1):14–20.

29. Saravanan R, Sujatha P (2018). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In: 2018 Second International Conference on Intelligent Computing and Control Systems

(ICICCS), Madurai, India, 945–949, doi: https: //doi.org/10.1109/ ICCON S.2018.86631 55.

30. Rasheed OA, Mohammed E, Iris S et al. (2020) Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform 136.
31. Taiwo Oladipupo Ayodele. Types of machine learning algorithms. New advances in machine learning, pages 19–48, 2010.
32. Muhammad LJ, Islam MM, Usman SS, et al. Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. Springer Nat Comput Sci. 2020. https: //doi.org/10.1007/ s4297 9-020-00216 -w.
33. Daniel R, Schrider A, Kern D. Supervised machine learning for population genetics: a new paradigm. Trend Genet. 2018;34–4:301–12.
34. Muhammad LJ, et al. Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano–Wudil Highway. Int J Database Theory Appl. 2017;10(11):197–208.
35. Taiwo Oladipupo Ayodele. Types of machine learning algorithms. New advances in machine learning, pages 19–48, 2010. Publisher: InTech.
36. Dianbo L, Leonardo C, Canelle P et al. (2020) A machine learning methodology for real-time forecasting of the 2019–2020 COVID19 outbreak using Internet searches, news alerts, and estimates from mechanistic models, https ://arxiv .org/abs/2004.04019.
37. Adrien PG, Yunpeng G, Gregory M, et al. A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. Ecol Inform. 2020;58:101090.

38. Ahmad IS, Bakar AA, Yaakub MR, et al. A survey on machine learning techniques in movie revenue prediction. SN Comput Sci. 2020;1:235. https ://doi.org/10.1007/s4297 9-020-00249 -1.

39. Asadi H, Dowling R, Yan B, Mitchell P, et al. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS ONE. 2014;9:2.

40. Singh P (2019) Supervised machine learning. In: Learn PySpark. Apress, Berkeley.

41. Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. Journal of artificial intelligence research, 4:237–285, 1996.

42. Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin

43. Khanday AMUD et al (2020) Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 12(3):731–739

44. Rajpurkar, Pranav. In: (2017). URL: https://arxiv.org/pdf/1711.05225.pdf

45. Joseph Paul Cohen Paul Morrison, Lan Dao. “COVID-19 Image Data Collection”. In: (2020). URL: https://arxiv.org/pdf/2003.11597.pdf.

46. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43:635–640

47. R.M.Pereira, D.Bertolini, L.O.Teixeira, C.N.Silla, and Y. M. G.Costa, ‘‘COVID-19 .identification in chest X-ray images on flat and hierarchical classification scenarios,’’Comput. Methods Programs Biomed. vol.194, Oct. 2020, Art. no. 105532.

48. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning—An MIT Press Book. Cambridge, MA, USA: MIT Press, 2016.

49. Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, pp. 436–444, May 2015.
50. M. Wainberg, D. Merico, A. Delong, and B. J. Frey, ‘‘Deep learning in biomedicine,’’ Nat. Biotechnol., vol. 36, no. 9, pp. 829–838, Oct. 2018.
51. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, ‘‘A guide to deep learning in health care,’’Nature-Med., vol.25, no.1, pp. 24–29, Jan.2019.
52. H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, ‘‘The rise of deep learning in drug discovery,’’ Drug Discovery Today, vol. 23, no. 6, pp. 1241–1250, Jun. 2018.
53. D.Shen,G.Wu,andH.Suk,‘‘Deep learning in medical image analysis,’’ Annu. Rev. Bi-omed. Eng., vol. 19, pp. 221–248, Jun. 2017.
54. Li M. Chest CT features and their role in COVID-19. Radiol Infect Dis. 2020;7(2):51–54.
55. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) outbreak: what the department of radiology should know. J Am Coll Radiol. 2020;17(4):447–451.
56. Kanne JP, Little BP, Chung JH, Elicker BM, Ketai LH. Essentials for radiologists on COVID-19: an update—radiology scientific expert panel. Radiological Society of North America. 2020;296(2): E113–E114.
57. Xu B, Xing Y, Peng J, Zheng Z, Tang W, Sun Y, et al. Chest CT for detecting COVID-19: a systematic review and metal analysis of diagnostic accuracy. EurRa-diol.2020;30(10):1.

58. WuJ, WuX, ZengW, GuoD, FangZ, ChenL,etal. ChestCT findings in patients with coronavirus disease 2019 and its relationship with clinical features. Investigative Radiology. 2020;55 (5):257.

59. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology. 2020;296(2):200432.

60. Weinstock MB, Echenique A, DABR JWR, Leib A, ILLUZZI FA. Chest x-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: a normal chest x-ray is no guarantee. J Urgent Care Med. 2020;14(7):13–8.

61. Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849, 2020.

62. Nanshan Chen, Min Zhou, Xuan Dong, Jieming Qu, Fengyun Gong, Yang Han, Yang Qiu, Jingli Wang, Ying Liu, Yuan Wei, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. The Lancet, 395(10223):507–513, 2020.

63. Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, and Bo Xu. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID19). preprint, Infectious Diseases (except HIV/AIDS), February 2020.

64. Li Yan, Hai-Tao Zhang, Yang Xiao, Maolin Wang, Chuan Sun, Jing Liang, Shusheng Li, Mingyang Zhang, Yuqi Guo, Ying Xiao, et al. Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan. medRxiv, 2020.

65. Sina F Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R Varkonyi-Koczy, Uwe Reuter, Timon Rabczuk, and Peter M Atkinson. Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188, 2020.

66. Dawei Wang, Bo Hu, Chang Hu, Fangfang Zhu, Xing Liu, Jing Zhang, Binbin Wang, Hui Xiang, Zhenshun Cheng, Yong Xiong, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. Jama, 2020.

67. Halgurd S. Maghdid, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, Kevin Curran, and Khaled Rabie. A novel ai-enabled framework to diagnose coronavirus covid 19 using smartphone embedded sensors: Design study, 2020.

68. Goossens, Michel, Mittelbach, Frank, and Samarin, Alexander. *"The Companion"*. Reading, Massachusetts: Addison-Wesley, 1993.

69. Pan, Sinno Jialin and Yang, Qiang. "A Survey on Transfer Learning". In: (). URL: https://www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf

70. Macukow, B. (2016). Neural networks – state of art, brief history, basic models and architecture. Computer Information Systems and Industrial Management, 3–14. https://doi.org/10.1007/978-3-319-45378-1_1

71. Kaushik, V. (2021, August 26). 8 applications of Neural Networks. Analyt-ics Steps. Retrieved February 19, 2022, from https://www.analytics- steps.com/blogs/8-applications-neural-networks

72. Victor Zhou. Machine Learning for Beginners: An Introduction to Neural Networks, December 2019. Library Catalog: towardsdatascience.com.

73. Support Vector Machine Machine learning algorithm with example and code, January 2019. Library Catalog: www.codershood.info Section: Machine learning.

74. Yanjun Qi. Random forest for bioinformatics. In Ensemble machine learning, pages 307–323. Springer, 2012.

75. Tony Yiu. Understanding Random Forest, August 2019. Library Catalog: towardsdatascience.com.

76. Wood, T. (2019, May 17). Softmax Function. DeepAI. Retrieved February 18, 2022, from https://deepai.org/machine-learning-glossary-and- terms/softmaxlayer

77. E. K. Zavadskas, J. Antucheviciene, S. H. R. Hajiagha, and S. S. Hashemi, "The interval-valued intuitionistic fuzzy MULTIMOORA method for group decision making in engineering," Mathematical Problems in Engineering, vol. 2015, Article ID 560690, 13 pages, 2015.

78. W. Jiang, Y. Luo, X.-Y. Qin, and J. Zhan, "An improved method to rank generalized fuzzy numbers with different left heights and right heights," Journal of Intelligent and Fuzzy Systems, vol. 28, no. 5, pp. 2343–2355, 2015.

79. G. Lee, K. S. Jun, and E.-S. Chung, "Robust spatial flood vulnerability assessment for Han River using fuzzy TOPSIS with cut level set," Expert Systems with Applications, vol. 41, no. 2, pp. 644–654, 2014.

80. D. Liang, W. Pedrycz, D. Liu, and P. Hu, "Three-way decisions based on decision-theoretic rough sets under linguistic assessment with the aid of group decision making," Applied Soft Computing Journal, vol. 29, pp. 256–269, 2015.

81. B. L. Golden, E. A. Wasil, and P. T. Harker, Analytic Hierarchy Process, vol. 113, Springer, New York, NY, USA, 2003.

82. Calabrò, F; Spina, L (2014). The Cultural and Environmental Resources for Sustainable Development of Rural Areas in Economically Disadvantaged Contexts-Economic-Appraisals Issues of a Model of Management for the Valorization of Public Assets. In Advanced Materials Research. Vol. 869, pp. 43-48.

83. Dasturani, M; Karimian, A; Ebrahimi, M; Rustam, MH (2012). Review method (TOPSIS) and its application in assessing the ecological potential areas for watershed management, Eighth National Conference on Science and Watershed Engineering, University of Lorestan.

84. Kermani, M.A.M.A; Badiee, A; Aliahmadi, A; Ghazanfari, M; Kalantari, H (2016). Introducing a procedure for developing a novel centrality measure (Sociability Centrality) for social networks using TOPSIS method and genetic algorithm. Computers in Human Behavior, 56, pp.295-305.

85. Leon, Y (2005). Rurality Development in Europe: A Research for Agricultural Economists, European review of Agricultural Economists, Vol. 23(3): 301-317.

86. X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, and X. Yang, ‘‘Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study,’’ Eur. J. Radiol., vol. 128, Art. no. 109041, Jul. 2020.

87. Hemdan EED, Shouman MA, Karar ME. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv:200311055. 2020.

88. Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest X-ray images using deep learning based anomaly detection. 2020. aeXiv.2002.12338 vol. 1.

89. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, and M. Cai, ‘‘A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19),’’ MedRxiv, Jan. 2020. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.02.14.20023028v5
90. X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, ‘‘Sample-effi-cient deep learning for COVID-19 diagnosis based on CT scans,’’ MedRxiv, Jan. 2020. [Online]. Available: https://www. medrxiv.org/content/10.1101/2020.04.13.20063941v1
91. C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, and W. Liu, ‘‘Deep learning-based detection for COVID-19 from chest CT using weak label,’’ MedRxiv, Jan. 2020. [Online]. Available: https://www. medrxiv.org/content/10.1101/2020.03.12.20027185v2
92. A. Amyar, R. Modzelewski, H. Li, and S. Ruan, ‘‘Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation,’’ Comput. Biol. Med., vol. 126, Nov. 2020, Art. no. 104037.

A review based on machine learning techniques for the diagnosis of COVID-19

Fatima Alvi[1], Faiyaz Ahmad[2], Mohammad Zunnun Khan[3]
Department of CSE Integral University, Lucknow, India

E- mail : [1]alvifatima25@gmail.com [2]faiyaz@iul.ac.in

[3]zunnun@iul.ac.in

Abstract
There are several uses for machine learning in prediction, which is a novel method. The COVID19 pandemic necessitates the use of this approach to detect high-risk individuals, as well as their mortality rates and other irregularities. It can be used to foresee future issues and learn more about the structure of the infection. Infected individuals have been effectively diagnosed at an early stage using imaging modalities such as computed tomography. There is an introduction to the pandemic-fighting machine learning here. Other COVID-19 datasets are also examined. Prefavorable reporting elements of systematic reviews and meta-analyses (PRISMA) standards are being used to conduct a scoping review on AI in support of COVID-19 Relevant papers published between January 1, 2020 and March 2, 2022, were searched for in a literature review. Numerous machine learning algorithms have been **investigated, and a thorough study of the proposed taxonomy has been conducted.**

Keywords- Machine learning, COVID-19 diagnosis, Artificial intelligence, Forecasting, Coronavirus.

1. Introduction

For complex issues, ML uses statistical models and little prior knowledge for difficult challenges (Gale, 1988). Regression methods include K-means clustering, support vector machines, decision trees, random forests, logistic regression, linear regression, and Naive Bayes models. In order to learn data patterns, this branch of machine learning creates neural network (NN) models utilizing feed-forward and reverse propagation. In-depth Learning are crucial in the battle against COVID-19 as well. These tests allow for both the diagnosis and prediction of a virus's severity. These models are used in the pharmaceutical sector to research the genetics and mutations of COVID-19 in an effort to improve medication prediction and immunization. In order to anticipate outbreaks, monitor the public, detect epidemics, and follow patients, AI has learned a variety of disease transmission models (Bagad P, 2020). First, we'll look at a couple publications that discuss COVID-19 machine learning techniques. The World Health Organization (WHO) declared the coronavirus disease COVID-19 a pandemic in December 2019 after it spread quickly over the world (Shuvo SB, 2020). Pandemic effects are being felt throughout all elements of society and the environment, including the economy (Wang F, 2020). Patients who are infected must be diagnosed as quickly as possible for both infection control and patient care reasons. SARS-COV-2, often known as Coronavirus, is a highly contagious illness that has spread around the world. These viruses are all members of the same family, Coronavirdae. All of the following symptoms can be caused by COVID-19: Cough, fever, breathing difficulties, headaches, aches and pains in the muscles, and a loss of taste and smell. Beta-COV genus, a rodent and bat infecting virus, has been tied to this virus, but its origins are still unclear. In December of this year, Wuhan, China, reported finding the first instance. There have been 3,947,630 fatalities and 182,276,267 verified cases as of June 29, 2021. Numerous industries in the health care profession are researching creative approaches to address this disease. Data Science is presently one of the hottest issues in the contemporary time. It includes AI, ML, DL, algorithms, modelling, stats, and sims (Morgan DJ, 2019). Artificial Intelligence (AI) may help fight this highly contagious pathogen. It aids academic and clinical research. ML has applications in engineering, multidisciplinary research, psychology, social analysis, earth observations, hazard mitigation, urban areas, etc. ML theories were implemented in applications such as face recognition, drone disinfection, automated pharmaceutical and food delivery, COVID-19 detection, and drug research.

1.1 Literature Survey

Khabir Uddin Ahameda et al, They utilized a preprocessing approach on the picture dataset to allow our deep learning model to analyse it accurately and quickly. They created a deep learning model based on ResNet50V2 that can precisely identify and classify COVID-19 patients. In their study, they demonstrated that our algorithm can accurately recognize COVID-19 patients using chest CT-scan and X-ray image datasets. To increase the model's accuracy and durability, advanced image processing techniques like hybrid filtering must be used. They started with ResNet50V2 and added layers. By adding layers to the current levels, the recommended design becomes deeper and more complex. Deeper models extract more characteristics, but they require a long time to train on huge datasets. As a consequence, they want to create a deep learning model that is simple, practical, and resilient.

Rachna Jain et al , For both covid-19 patients and healthy individuals, the PA view of chest x-ray pictures was used in this investigation. After image cleaning and data augmentation, they compared the performance of CNN models based on deep learning. They compared Inception V3, Xception, and ResNeXt. Kaggle repository contributed 6432 chest x-ray scan samples for training and validation. The Xception model accurately identifies Chest X-rays (97.97%). This study examines various covid-19 categorization algorithms and makes no medical claims.

Lin Li et al, In 2019, To detect the coronavirus in chest CT images, they applied deep learning. They discovered that this model has good specificity (90%) and sensitivity (90%) using data from independent testing (96 percent). AUC was 0.96 in CAP and COVID-19. (95% CI: 0.94, 0.99). COVNet can extract two-dimensional local features. The system's backbone is RestNet50, which takes in CT slices and develops features for them. Each slice receives a maximum pooling of the recovered characteristics. A completely linked layer with a softmax activation function is created from the resulting feature map (COVID-19, CAP, and non-pneumonia). Using U-net, they perform preprocessing on a 3D CT image to isolate the lung area as the region of interest (17). After that, the image is sent to our COVNet for prediction. The results also reveal that a convolutional network model can identify COVID-19 from CAP.

Himadri Mukherjee et al, Compared to deep CNN designs, this study proposes a shallow CNN architecture with only four layers. The main objective was to build a light architecture with few parameters (weights) to save calculation time. Researchers examined CXRs that were healthy, positive for Covid-19, and positive for pneumonia. In order to evaluate the durability of COVID19, 5-fold cross validation was performed on balanced and unbalanced datasets. A popular deep learning tool like MobileNet, InceptionV3, and ResNet50 was compared to the suggested shallow CNN-tailored architecture, as was Covid-19 identification using CXRs. To examine for Covid19 positive patients in chest X-rays, the proposed shallow CNN customised architecture might be applied.

Tianyang Li et al, As a technique for COVID-19 assisted screening using chest X-rays, they suggest discriminative cost-sensitive learning (DCSL). DCSL combines cost-sensitive learning with fine-grained categorization. Initial DCSL conditional centre loss DCSL uses cost-sensitive learning at the score level to lower the cost of misclassifying COVID-19 data. Any DNN, 239 from confirmed COVID-19 cases, 1,000 from bacterial or viral pneumonia patients, and 1,000 from healthy individuals may use DCSL. Tests on three-class classification demonstrate their strategy outperforms existing methods. It is 97.01 percent accurate in terms of precision, sensitivity, and F1 score. These results confirm their large-scale COVID-19 screening approach.

Cheng Jin et al, Patients with influenza A/B, non-viral CAP, and non-pneumonia were included in the study. On two publicly accessible datasets, the CC-CCII and the MosMed Data, the multiway classification accuracy is 97.81 percent, with AUCs of 92.99 percent and 93.25 percent. On a reader research with five radiologists, the artificial intelligence system outperformed them all by two orders of magnitude in progressively demanding tasks. When comparing the diagnostic performance of CXR and CT, there are several differences.

2. Impact of Machine learning in disease prediction

Companies in the fields of health care and manufacturing use data mining to their advantage. Indepth decision assistance typically necessitates new technologies, according to the experts. Using this innovative technology, trends and predicted patterns in data may be discovered, hypotheses can be developed and tested, and eye-opening visualizations can be created(O'Flynn-Magee K, 2021). End users benefit from data mining because it makes it possible for them to glean meaningful information from massive datasets. Data warehouses, sometimes known as "Data Mountains," hold these massive datasets, which are then accessible to data mining software. When it comes to data warehouses, one may build the mountains of data. Extracting previously unidentified and possibly usable information from a data mountain is referred to as "data mining." This data mining is not industry-specific; it necessitates intellectual technologies and a desire to discover the potential of the hidden knowledge that lies in the data. Sometimes known as knowledge discovery in databases (KDD) [20], data mining is also referred to as. The goal of data mining is to uncover hidden patterns in large amounts of corporate data in order to generate predictions about how that data will be used in the future. It's the process of extracting relevant information from enormous databases that isn't always apparent.

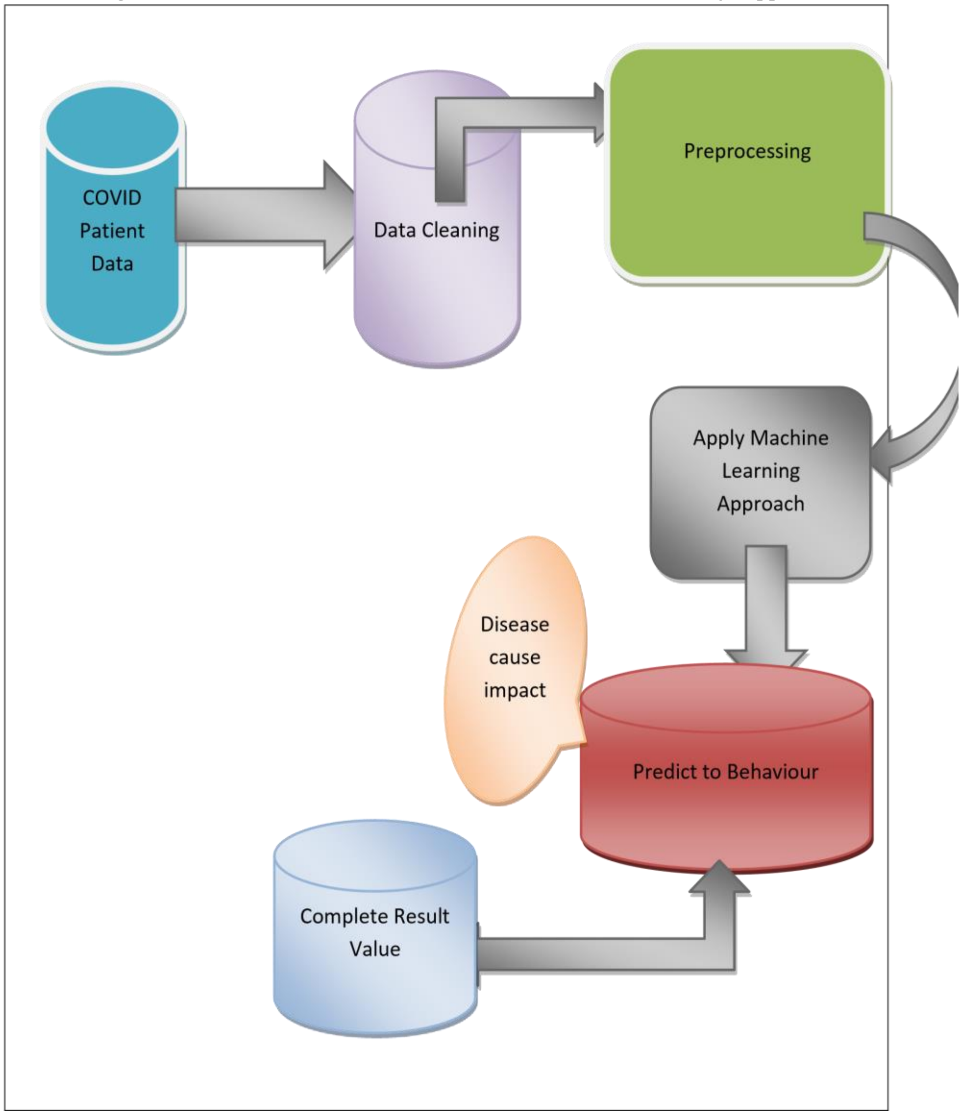

Fig 1: Process for Evaluation Using Machine Learning Concepts

3. COVID-19

Corona virus 2019 (COVID-19), Since its discovery at the end of November, the SARS-Cov-2 disease, which is a new pathogen, has generated a worldwide pandemic crisis by spreading around the globe. The World Health Organization (WHO) April 2021 report estimates that since the outbreak started in 2019, there have been close to 150 million confirmed illnesses and more than 3 million fatalities (D. Cucinotta, 2020). There have been about 32.5 million illnesses and 500,000 fatalities in the US alone, making it the country with the largest cumulative toll. Many healthcare facilities have been badly impacted by these enormous numbers, particularly due to the virus's propensity to create more genetic variations and transmit more easily among individuals. A new form of COVID-19 has caused a large number of cases in India, one of the world's largest suppliers of vaccinations, and the country is now seriously affected by the pandemic. More than 17.5 million cases have been verified, making it the second-worst struck nation behind the United States (C. Huang, 2020). Symptoms in COVID-19 individuals may vary from asymptomatic to life-threatening, with pneumonia being the most common. Virus incubation may last anywhere from one to fourteen days in the majority of instances before any signs of infection show up. In patients who tested positive for COVID-19, coughing, wheezing, fever, fatigue, and other acute respiratory distress syndromes have all been documented (ARDS). The majority of persons who get infected only have mild to moderate viral symptoms before fully recovering. People over the age of 60 with disorders including diabetes, cardiovascular disease (CVD), high blood pressure (hypertension), and cancer are more likely to suffer from severe pneumonia. Most of the time, early detection of COVID-19 aids in avoiding the infection from spreading and progressing to a more severe state. Steps like early patient isolation and contact tracking are often used to accomplish this goal. Additionally, prompt medicine and quick treatment minimises symptoms and lowers the pandemic's fatality rate. Human civilisation has been put at risk by COVID-19, one of the most terrible diseases ever discovered. Innovative approaches to disease detection, prevention, and management have been developed as a result of the advancements in contemporary technology during the last few decades [27]. Different imaging modalities, such as CT and X-ray, are thought to be among the most helpful for diagnosing COVID-19 specifically. While X-rays are more economical and commonly accessible, M. M. Islam et al. : Review on Machine Learning Approaches for the Diagnosis of Novel Coronavirus (COVID-19) techniques recommend CT screening when it is available because of its adaptability and three-dimensional lung perspective [28]. When it comes to combating the epidemic, standard medical imaging technologies are very essential. Tables 1 show the results of our investigation of the numerous theories and variants of the COVID-19 theory from the standpoint of machine learning.

Table 1: Contribution Table

References	Methodology	Disease Constraints
In [1]	Deep-learning based multimodal system	Covid-19
In [2]	Deep convolutional neural network with multi-feature channel	COVID-19 disease
In [3]	Sound Based	COVID-19
In [4]	Ensemble learning approach	COVID-19
In [5]	CNN based observation	Lung auscultation sounds
In [6]	Health care demand interpretable	Artificial intelligence
In [7]	Artificial Intelligence	Predicting hospital readmissions
In [8]	Clinical practice	Covid-19
In [9]	Deep Learning and Radionics	Covid-19
In [10]	Deep Learning Prognosis Model	Covid-19
In [11]	Neural networks	COVID-19 and cardiovascular disease
In [12]	Neural networks	Classification of COVID- 19 patients
In [13]	CT halo sign	COVID-19
In [14]	Feature Analysis	COVID-19
In [15]	Deep Bayes	COVID-19
In [16]	Deep neural network	COVID-19

4. Relational of study

This research [17] reviews and discusses deep learning-based COVID-19 diagnosis systems employing CT or X-ray images. Despite the fact that many of the issues raised in the literature are addressed in this study, more research is needed to solve some of the remaining issues. There is no foundation information on deeplearning approaches emphasizing mathematical representations, hence the only COVID-19 diagnostic systems utilising these techniques are reported [18-21]. For this assignment, it is expected that the student has a working knowledge of a certain area. The reader is then referred to pertinent sources to learn more about particular characteristics of the neural networks under investigation, such as the number and specification of layers employed, learning rate, number of used epochs, batch size, dropout layer, optimizer, and loss function. Despite the fact that this assessment considers the COVID-19 diagnostic from a computer vision standpoint, no qualitative diagnosis results are displayed in CT or X-ray pictures. A fourth point to make is that, regardless of whether a model is pre-trained or customised, most of the examined systems have an accuracy of 90 percent or more on CT and Xray scans with little or big amounts of data [25-27]. Last but not least, this study does not provide any implemented examples or computer code that may be used to replicate some of the most prominent outcomes in the COVID-19 diagnostic systems under consideration.

5. Medical significance

According to the study's statistical findings, COVID-19 and healthy people had different ischemic heart disease comorbidities. Studies relate COVID-19 to heart dysfunction and this is consistent with this [28]. COVID-19 has been shown to cause myocardial damage, cardiac arrhythmia, and acute coronary syndromes in animal studies. In addition, a number of respiratory health conditions, including fever, cold, and cough, were found to be significant in distinguishing COVID-19 subjects from healthy individuals. According to WHO data, 14.8% of COVID-19 patients in China experienced muscular discomfort. This research didn't find a link between COVID-19 and diarrhoea. Smartphone-based breathing recordings combined with deep learning may offer a noninvasive, zero-cost, and quick COVID-19 screening method. RT-PCR may be beneficial in nations where healthcare, economic, and political concerns prevent universal testing [29]. Furthermore, the suggested structure enables faster, more affordable, and more effective largescale detection, particularly for counties and districts that have logistical difficulties with RTPCR testing. In addition to the rapidity of this approach, reducing the demand on clinicians or nurses could significantly revive many healthcare services. Detection of asymptomatic subjects is made easier, which reduces the need for additional equipment and the associated costs of additional medication after a viral infection develops in patients.

6. Discussion

Our article summarises the most recent studies on the use of machine learning to forecast various outcomes in dermatology, including skin cancer. The results of all previous investigations on this subject have been positive [30]. ML approaches allow dermatologists to more correctly anticipate when precision medicine is a goal, the clinical results and prognoses of their patients with skin conditions [31]. Unlike conventional statistical approaches, ML focuses on prediction rather than inference. In more layman's terms, machine learning algorithms are built to anticipate future behaviour rather than merely detecting connections in the data they already have. Datasets with a lot of input variables can also be used for machine learning [32]. Because of this, ML algorithms are able to create more accurate inferences from the data and hence deliver more dependable findings when they have a larger sample size to work with. When the quantity of input variables was small to moderate, conventional statistical techniques were anticipated to be the most precise and efficient [33]. Statistical models lose precision as the number of inputs rises. COVID-19 continues to be a pandemic, with new records being set every day in terms of worldwide infection and death toll. Pre-trained models with COVID-19 diagnostic architectures are described in the review. According to the taxonomy, the two views of machine learning and imaging were examined at two levels. Useful datasets that can be easily accessible and understood by researchers are included in this document. The lack of gold standards is a fundamental obstacle for COVID-19 diagnostic systems based on deep learning. In addition, researchers are encouraged to come up with viable answers to the existing issues in order to inspire and motivate them to continue working in this field. It's important to keep in mind that imaging modalities and machine learning approaches only provide a partial picture of the infected individuals. The present study does not imply that machine learning approaches can take the place of doctors or clinicians in clinical diagnosis.

References

1. Sait U, KV GL, Shivakumar S, Kumar T, Bhaumik R, Prajapati S, et al. A deeplearning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images. Applied Soft Computing. 2021; p. 107522.
2. Kumar LK, Alphonse P. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. Alexandria Engineering Journal. 2021;
3. Bagad P, Dalmia A, Doshi J, Nagrani A, Bhamare P, Mahale A, et al. Cough against COVID: Evidence of COVID-19 signature in cough sounds.
4. ohammed EA, Keyhani M, Sanati-Nezhad A, Hejazi SH, Far BH. An ensemble learning approach to digital corona virus preliminary screening from cough sounds. Scientific Reports. 2021;11(1):1–11. pmid:34321592
5. Shuvo SB, Ali SN, Swapnil SI, Hasan T, Bhuiyan MIH. A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram. IEEE Journal of Biomedical and Health Informatics. 2020;.
6. Wang F, Kaushal R, Khullar D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? Ann Intern Med. (2020) 172:59.
7. Morgan DJ, Bame B, Zimand P, Dooley P, Thom KA, Harris AD, et al. Assessment of machine learning vs standard prediction rules for predicting hospital readmissions. JAMA Netw Open. (2019) 2:e190348.
8. O'Flynn-Magee K, Hall W, Segaric C, Peart J. GUEST EDITORIAL: The impact of Covid-19 on clinical practice hours in pre-licensure registered nurse programs. Teaching and Learning in Nursing. 2021;16(1):3. pmid:32874165
9. Li C, Dong D, Li L, Gong W, Li X, Bai Y, et al. Classification of Severe and Critical Covid-19 Using Deep Learning and Radiomics. IEEE Journal of Biomedical and Health Informatics. 2020;24(12):3585–3594.
10. Meng L, Dong D, Li L, Niu M, Bai Y, Wang M, et al. A Deep Learning Prognosis Model Help Alert for COVID-19 Patients at High-Risk of Death: A Multi-Center Study. IEEE Journal of Biomedical and Health Informatics. 2020;24(12):3576–3584. pmid:33108303
11. Nishiga M, Wang DW, Han Y, Lewis DB, Wu JC. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. Nature Reviews Cardiology. 2020;17(9):543–558.
12. D. Singh, V. Kumar, Vaishali, and M. Kaur, ''Classification of COVID- 19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks,'' Eur. J. Clin. Microbiol. Infect. Dis., vol. 39, no. 7, pp. 1379–1389, Jul. 2020.
13. X. Li, X. Zeng, B. Liu, and Y. Yu, ''COVID-19 infection presenting with CT halo sign,'' Radiol., Cardiothoracic Imag., vol. 2, no. 1, Jan. 2020, Art. no. e200026.
14. A. Farid, G. I. Selim, and H. A. A. Khater. (2020). A Novel Approach of CT Images

    Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19). [Online]. Available: https://www.preprints.org/manuscript/202003.0284/v1
15. F. Ucar and D. Korkmaz, ''COVIDiagnosis-net: Deep Bayes- squeezenet based diagnosis of the coronavirus disease 2019 (COVID-from X-ray images,'' Med.
    Hypotheses, vol. 140, Jul. 2020, Art. no. 109761.
16. I. Khan, J. L. Shah, and M. M. Bhat, ''CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images,'' Comput. Methods Programs
    Biomed., vol. 196, Nov. 2020, Art. no. 105581.
17. M. Rahimzadeh and A. Attar, ''A modified deep convolutional neural net- work for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of xception and ResNet50 V2,'' Informat. Med. Unlocked, vol. 19, 2020, Art. no. 100360.
18. K. R. Himadri Mukherjee, Subhankar Ghosh, Ankita Dhar, Sk. Md. Obaidullah, KC Santosh. (2020). Shallow Convolutional Neural Network for COVID-19 Outbreak Screening using Chest X- rays. [Online]. Available:

    https://www.techrxiv.org/articles/preprint/Shallow_Convolutional_Neural_Network_for_COVID-19_Outbreak_Screening_using_Chest_X-rays/12156522/1
19. T. Li, Z. Han, B. Wei, Y. Zheng, Y. Hong, and J. Cong, ''Robust screening of COVID19 from chest X-ray via discriminative cost-sensitive learning,'' 2020.
20. D. Cucinotta and M. Vanelli, ''WHO declares COVID-19 a pandemic,'' Acta Biomed., vol. 91, no. 1, pp. 157–160, 2020.
21. Worldometers. Accessed: Feb. 11, 2021. [Online]. Available: https://www.worldometers.info/coronavirus/

22. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, and L. Zhang, ‘‘Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,’’ Lancet, vol. 395, pp. 497–506, May 2020.
23. P. Vetter, D. L. Vu, A. G. L’Huillier, M. Schibler, L. Kaiser, and F. Jacquerioz, ‘‘Clinical features of COVID-19,’’ BMJ, vol. 4, Apr. 2020, Art. no. m1470.
24. T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, and Q. Tao, ‘‘Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-
19) in China: A report of 1014 cases,’’ Radiology, vol. 296, Feb. 2020, Art. no. 200642.
25. X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, ‘‘Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: Relationship to negative RT-PCR testing,’’ Radiology, vol. 296, no. 2, pp. E41–E45, Aug. 2020.
26. A. Kumar, P. K. Gupta, and A. Srivastava, ‘‘A review of modern technolo- gies for tackling COVID-19 pandemic,’’ Diabetes Metabolic Syndrome,
Clin. Res. Rev., vol. 14, no. 4, pp. 569–573, Jul. 2020.
27. D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, ‘‘Digital tech nology and COVID-19,’’ Nature Med., vol. 26, no. 4, pp. 459–461,
2020.
28. Khabir Uddin Ahameda et al, A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images, Computers in Biology and Medicine, Volume 139, December 2021, 105014
29. Rachna Jain et al, Deep learning-based detection and analysis of COVID-19 on chest X-ray images, Springer Science+Business Media, LLC, part of Springer Nature 2020
30. Li, Lin; Qin, Lixin; Xu, Zeguo; Yin, Youbing; Wang, Xin; Kong, Bin; Bai, Junjie; Lu, Yi; Fang, Zhenghan; Song, Qi; Cao, Kunlin; Liu, Daliang; Wang, Guisheng; Xu, Qizhong; Fang, Xisheng; Zhang, Shiqin; Xia, Juan; Xia, Jun (2020). Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology, (), 200905– doi:10.1148/radiol.2020200905
31. Mukherjee, H., Ghosh, S., Dhar, A. et al. Shallow Convolutional Neural Network for COVID-19 Outbreak Screening Using Chest X-rays. Cogn Comput (2021).
32. Tianyang Li et al, Robust Screening of COVID-19 from Chest X-ray via Discriminative Cost-Sensitive Learning, Preprint submitted to Elsevier, 2020
33. Jin, C., Chen, W., Cao, Y. et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 11, 5088 (2020).

Service Discover mechanism for the diagnosis of Covid-19 using Convolutional Neural Network and hybrid ml classifiers

Fatima Alvi[1], Faiyaz Ahamad[2], Mohammad Zunnun Khan[3]
Department of CSE Integral University, Lucknow, India

E- mail : [1]alvifatima25@gmail.com [2]faiyaz@iul.ac.in
[3]zunnun@iul.ac.in

ABSTRACT

There are several uses for machine learning in prediction, which is a novel method. The COVID-19 pandemic necessitates the use of this approach to detect high-risk individuals, as well as their mortality rates and other irregularities. Utilizing it is one way to have a better knowledge of the structure of the virus and to anticipate potential issues in the future. Machine Learning along with Convolutional Neural Network are the major techniques that can shed light on well-known data sets that are used to train these networks. Imaging methods such as computed tomography have made it possible to accurately diagnose infected people at an earlier stage in the evolution of the illness. In this research we have collected datasets in the form of CT-scans and chest X-rays of Covid-19 positive patients and healthy people from Kaggle repository. The images that are collected are resized and data normalization was performed on the dataset for better learning of the system. Data pre-processing procedures include noise removal, scaling, and augmentation. The data partitioning approach divides the data into three sets for the experiment: training, validation, and testing. Feature extraction and classification are the most important steps. In this experiment, we have used five different models that are XGBoost, Decision Tree, Adaboost, Extra Trees and Support Vector Classifier and performed hybridization method in our HN model. Finally, criteria such as accuracy, precision, F1-score, recall and others are used to evaluate the built system. Xgboost achieves the highest accuracy of 92%, precision of 89%, recall of 95% and f1-score of 92%.

KEYWORDS- Machine learning, COVID-19, Artificial intelligence, Forecasting, Corona virus.

1. INTRODUCTION

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, emerged in 2019 as a serious infectious disease that can cause severe respiratory distress. Since its discovery in Wuhan, China, the disease has claimed the lives of millions of people all over the world [1,2]. COVID-19 has become widespread, since the propagation of the virus through human-to-human transmission. Viruses can manifest themselves long before symptoms appear. At the time of writing, over 130 billion people all around the world had been infected [3], posing a massive healthcare burden. Using the RT-PCR assay, COVID-19 was shown to be present [4]. Because the coronavirus is an RNA virus, the generation of complementary DNA from its genetic material requires that it first go through the process of reverse transcription (cDNA). Even However, research has demonstrated that RT-PCR might give findings that are incorrectly interpreted as negative [5].The gold standard RT-PCR test for COVID-19 has flaws that make the condition difficult to detect. RTPCR is time-consuming, labor-intensive, and costly [6]. The procedure needs a lab kit, which many nations can't acquire during crises and epidemics.[7]. This procedure is not error-free and biased, much like all diagnostic and laboratory methods used in healthcare.[8] The nose and throat mucosa must be sampled by an expert laboratory technician, which is a painful procedure [9–12]. Different imaging modalities, such as methods [13-15], are used for COVID-19 diagnosis. Because of its adaptability and three-dimensional pulmonary imaging, CT screening is preferred over X-ray screening when possible [16, 17], despite the fact that X-rays are more affordable and more widely available. Pandemic control relies heavily on the use of traditional medical imaging tools.[18]

The data is analyzed using CNN, a machine learning tool. This study looks at how CNN models can be used to categories chest X-ray pictures in coronavirus-infected individuals. The software tools and machine learning algorithms that we built to guarantee that the proposed model for the machine learning system is of high quality were also extensively validated. [19] Traditionally we design the algorithm then we submit the input data then after we get the result, but in machine learning, first, we trained the machine by submitting input as well as output then machine design the algorithm [20].

At the conclusion of the process, all of these models are combined to produce a hybrid model. This results in decision-making that is superior than that produced by any particular classifier. The following is a list of the significant contributions that the paper makes:

1. The purpose of this study was to present a hybrid model that detects COVID-19 patients by extracting deep features from CT-scans and Chest X-Ray images. The model uses CNN along with other machine learning models.

2. A hybrid model is built for the purpose of categorization. The goal of the hybridization procedure is to improve the model's accuracy or degree of precision. The fact that one model performed very well in a particular test set does not guarantee that it will consistently perform well in all subsequent test sets. When applied to various datasets, different models may provide different results.

3. There are a total of 1262 photos in the COVID-1262 class, of which 80% are training data and 20% are test data; conversely, there are 1267 images in the COVID-1267 class where contemporary emerging systems have done their study using very tiny COVID-19 datasets.

4. The Xgboost model had the best overall performance, with 92 percent accuracy, 95 percent precision, 88 percent recall, and 92 percent F1- score respectively.

2. Related Work

Chest X-rays are used to determine the presence of COVID-19 in studies that include either binary or many categories. In certain areas of study, the use of raw data is common, whereas in others, feature extraction is more common. Investigations use a variety of different quantities of data points, which also varies. It has been discovered that the most often used method in investigations is the use of convolutional neural networks (CNN). The use of COVID-19 to cause pneumonia was used by. In particular, the method often referred to as transfer learning has been used. Transfer learning allows for the accurate identification of a wide variety of defects, even in very small datasets of medical images, and the results are often quite spectacular [21].

This project intends to construct a deep learning-based model that can identify COVID-19 with high sensitivity, offering rapid and reliable scanning [22]. The model will identify COVID-19 using chest X-rays. CNN approaches can learn proper representations of data in many layers consecutively [23, 24]. Deep learning algorithms are frequently employed in biomedicine [25], smart healthcare [26], drug development [27], medical image analysis [28], and other medical systems.

COVID-19 can be diagnosed using radiography pictures, such as chest X-ray and computed tomography (CT), because the disease primarily affects the human respiratory system [29]. A typical pneumonia [30] and organizing pneumonia [31] are the most common abnormalities on chest X-rays. Ground Glass Opacity (GGO), which refers to a region of enhanced attenuation in the lung, is the most prevalent finding in chest radiography pictures. An X-ray of the chest reveals that the image is not totally black but rather has a hazy grey colour with many microscopic white blood vessels. This is seen on the X-ray. On the other hand, CT scans reveal the existence of GGO as well as consolidation in cases when the condition is more severe. "Crazy paving" refers to GGO along with interlobular and intralobular septal hypertrophy on chest imaging.

It is essential to bring to your attention the fact that chest CT is thought to have a higher sensitivity [32] than chest X-rays when it comes to the early diagnosis of COVID-19. A CT scan would show a distinctive pattern of GGO and consolidation in the lungs. CT scans are yet another diagnostic method that may be used to ascertain the severity of the ailment [33]. According to a recent research that was carried out on 51 COVID-19 patients, CT shown a sensitivity of 98 percent for COVID-19 infection, but RT-PCR demonstrated a sensitivity of only 71 percent [34].

The key issue is that similar findings are reported in pneumonia cases as well as COVID-19 cases. Many moderate COVID-19 cases have symptoms that are comparable to those of a regular cold, and in some cases, the lungs appear to be normal. Despite the fact that research published in [35] found that the radiographic image of COVID-19-affected lungs varies from that of bacterial pneumonia-affected lungs. In medical images inception module is one of the great innovations which helps in identifying the features more accurately. Each layer in a standard CNN model sends its output to one of the other layers as input. This is repeated until the output layer is reached, in a pipeline-like approach.

3. Methodology

Therefore, the purpose of this study is to make the Covid-19 screening process more user-friendly and productive for patients. The SARS Covid-2 virus is the infectious agent that causes the sickness known as Covid-19.

We need to build some faster algorithms and so we have hybridized the algorithms. We mixed the algorithms as to get better results. We have taken out features from GoogleNet, etc. from flatten layers and save it on the file and used machine learning algorithms like knn, naïve baise using PCA.

CNN processes pixel data. CNNs are image identification and processing artificial neural networks. CNN gathers raw image pixel input, trains a model, and automatically extracts characteristics for classification. VGG 16, GoogleNet, etc. Machine learning is a subfield of computer science that focuses on giving computers the capacity to learn on their own without being given specific instructions. We have used three libraries that are keras, theanos and TensorFlow. We have chosen a hybrid model. Epoch achieved was 20. In order to reduce complexity, we have used VGA. CNN works on different data sets of same size. We have taken data from Kaggle of different sizes and tried improving accuracy, and taken out images from flatten layer.

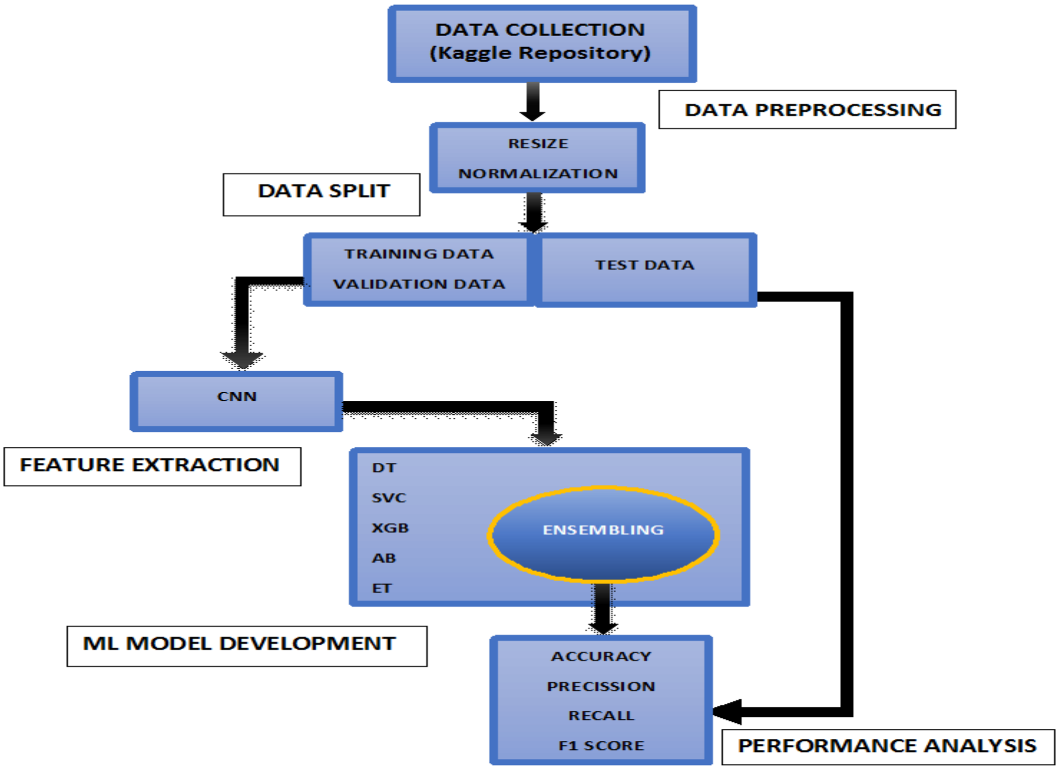

Fig.3.1. Proposed Methodology for Covid-19 prediction model

The process begins with the introduction of data of a high quality, followed by the training of machines via the construction of machine learning models through the use of data and a variety of methods. The initial completely linked layers of the model were the ones on which the feature extraction was performed. The feature vector was extracted from each training image before moving on to the next layer. After that, the feature vector was sent through a total of four different classifiers. After that, each of the machine learning classifiers was subjected to hybridization so that an ensemble of classifiers could be developed. After that, the performance analysis is carried out, during which the suggested system is assessed according to its recall, accuracy, and precision, as well as its F1-score.

The whole workflow of a COVID-19 diagnostic system that is based on deep learning is shown in figure 3.1. During the stage when the data is being collected, patients from the hospital are deemed to be participants. Imaging methods such as CT and X-ray samples are used in order to identify COVID-19. This is the case in spite of the fact that the data might exist in a wide variety of formats. The photos that are gathered are scaled, and data normalisation was conducted on the dataset. This was done so that the system could learn more effectively, and the dataset was then prepared for input into the CNN network, training the model, and making it easier to generalise the results. The next critical step is the one known as "data preparation," which involves getting the data into a format that can be used. The steps of removing noise, scaling, and augmenting the data are included in the pre-processing operations. The experiment's data are partitioned into training, validation, and testing sets. Data partitioning uses cross-validation. Training data is utilised to develop a model, which is subsequently verified and tested. Feature extraction and classification are crucial for deep learning-based COVID-19 diagnosis. The deep learning algorithm collects features by repeating a number of procedures, and class labels complete the categorization (healthy or COVID-19). V-Net may be used to extract radiomics properties, such as size, shape, and texture, while adding spatial data. Finally, criteria such as accuracy, sensitivity, specificity, precision, F1-score, and others are used to evaluate the built system.

3.1 Description of Datasets

We tried to acquire access to direct information by contacting a number of hospitals and health institutes in India, but owing to the current scenario at hospitals, which involves a high volume of patients, we were unable to do so. In order to collect open source clinical information of patients who had been diagnosed with COVID-19, a comprehensive search was carried out on a variety of databases. The original datasets is collected from Kaggle, which include X-ray images or C-T scans of Covid-19 positive patients and Covid- 19 negative patients. have collected 1268 images of Covid patients, whereas 1272 images of Covid negative patients from We Kaggle Repository. The available datasets were of different sizes, so we tried improving the accuracy.

3.2 Data Preprocessing

The sample photographs were all of different sizes. As a result, the photos were scaled to 408×408 pixels in size. Data standardization was carried out in order to improve the system's learning. With this data in place for use in training the CNN network, we were ready to proceed with the experiment. Normalization speeds up the training of a CNN model and improves the stability of the gradient descent [36].

3.3 Data Partitioning

The data for the experiment are partitioned using the data partitioning technique, which creates three separate sets: training, validation, and testing. Data partitioning typically uses cross-validation. After training data verification, the model's performance is validated using test data. Model development uses training data.

3.4 Feature Extraction

Extraction of features and categorization of those features are the two processes in a deep learning-based COVID-19 diagnosis that are most important. At this stage, the CNN will extract the feature by going through a series of operations many times, and the classification will be finished with the application of class labels (healthy or COVID-19). V- Net can be used to extract various radiomics features, which needs extracting the greatest number of features from medical images, such as size, shape, and textural qualities, while also including spatial data.

3.5 Algorithm Configurations

This section discusses the algorithms' settings and how they may be configured. Alterations made to the settings of the algorithm's setup may have an impact on the output.

- Support Vector Machines:

  SVC(kernel = ’linear’, random_state = 0)

- Convolutional Neural Network:
- Layers:
  - model.add(tf.keras.layers.Conv2D(64, kernel_size=(3,3), strides=(1,1), padding='valid', activation='relu', input_shape=(64,64,3), kernel_regularizer =tf.keras.regularizers.l1(l=0.001)))
  - model.add(tf.keras.layers.Conv2D(64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu', kernel_regularizer =tf.keras.regularizers.l1(l=0.001)))
  - model.add(tf.keras.layers.Conv2D(32, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu')))

- Compiling the CNN:
  opt = Adam(0.001)
  model.compile(loss='binary_crossentropy', optimizer=opt, metrics=['acc'])

- Xgboost:
  clf = XGBClassifier(random_state=0)

- SVM:
  clf = SVC(random_state=0)

- Decision Tree:
  clf = DecisionTreeClassifier(random_state=0)

- Extra Trees:
  clf = ExtraTreesClassifier(random_state=0)

- Adaboost:
  clf = AdaBoostClassifier(random_state=0)

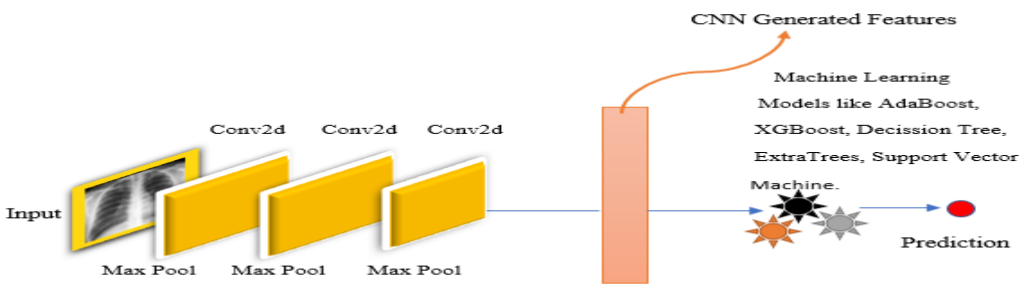

Fig.3.2 Hybrid Network Model Architecture (HN-Model)

The above figure 3.2 depicts our Hybrid Network Model architecture, here the DNN Layer has only 3 layers and followed by the Machine Learning classification layer, where we have experimented with ML classification algorithms like Adaboost, Xgboost, Decision Tree, Extra Trees and SVM.

3.6 Performance Metrics

The challenge of measuring the performance of a machine learning model is a crucial part of the process. Because our model relies on categorization, we decided to base performance on accuracy rather than any other parameter.

3.6.1 Accuracy

In this particular experiment, the algorithms are judged according to how accurate their results are. It is the performance measure that is used the majority of the time when evaluating categorization strategies.

$$Accuracy = \frac{TP + TN}{TP + FT + FN + TN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

3.6.2 Precision

Precision is one of the ways in which the performance of a machine learning model can be evaluated. The precision metric evaluates how reliable a model is in providing optimistic forecasts for the foreseeable future.

$$Precision = \frac{TP}{TP + FP}$$

3.6.3 Recall

The term "recall" refers to the total number of good results that our ML model generates. Using the confusion matrix and the formula that is provided above, we are able to simply compute it.

$$Recall = \frac{TP}{TP + FN}$$

3.6.4 F1-score

This score will be used in the calculation of the harmonic mean of the accuracy and recall scores. F1's greatest value is 1 and its lowest value is 0. The F1 score may be determined by using the formula that is provided below

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

4. Result and Discussion

This chapter presents the results that are achieved from this experiment. We evaluated the performance measures in terms of Accuracy, Precision, F1-score and Recall. In order to have some faster algorithms and better performance we hybridized the algorithms. We extracted features using CNN and then applied machine learning.

In this experiment, we have used five different models that are XGBoost, Decision Tree, Adaboost, Extra Trees and Support Vector Classifier and performed hybridization method. Firstly we have applied normal CNN and then applied various machine learning models and combined them in our hybrid network model.

4.1 Applying HN-SVM

So now we after using SVM along with CNN. Support vector machines, often known as SVMs, are a family of supervised learning algorithms that can categorise data, carry out regression analysis, and locate outliers. The advantages of support vector machines include the following: efficient in high-dimensional situations. Maintains its utility even in circumstances in which the number of dimensions is greater than the number of samples. Table 4.1 shows the performance measures that this model has achieved. The HN-SVM model achieved the accuracy 82%, precision of 77%, recall of 91% and f1-score of 83%.

Table 4.1. Performance results (%) of HN-SVM model

	PRECISION	RECALL	F1-score	support
0.0	0.89	0.74	0.81	259
1.0	0.77	0.91	0.83	250
Accuracy			0.82	509
Macro-avg	0.83	0.82	0.82	509
Weighted-avg	0.83	0.82	0.82	509

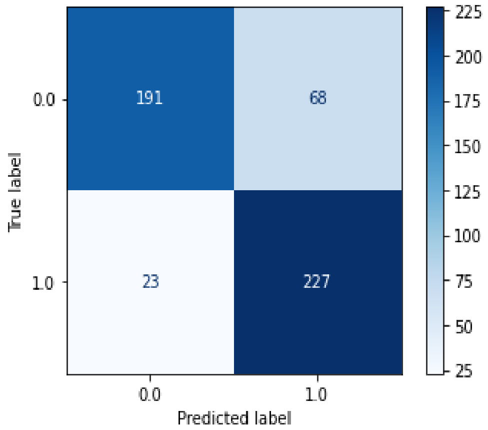

Fig. 4.1. Confusion matrix representation for HN-SVM model

4.2 Applying HN-Decision Tree

A decision tree is a technique for supervised machine learning that may be used to categorise or predict data based on the responses that have been provided in the past to questions that have been asked. The nature of the model is supervised learning, which indicates that it is trained and evaluated using data sets that include the necessary classification in order to get the desired results. This Table 4.2 shows the performance evaluation of this model. It has achieved the accuracy of 78%, precision of 77%, recall of 78% and f1-score of 78%.

Table 4.2. Performance results (%) of HN-Decision Tree model

	PRECISION	RECALL	F1-score	support
0.0	0.79	0.78	0.78	259
1.0	0.77	0.78	0.78	250
Accuracy			0.78	509
Macro-avg	0.78	0.78	0.78	509
Weighted-avg	0.78	0.78	0.78	509

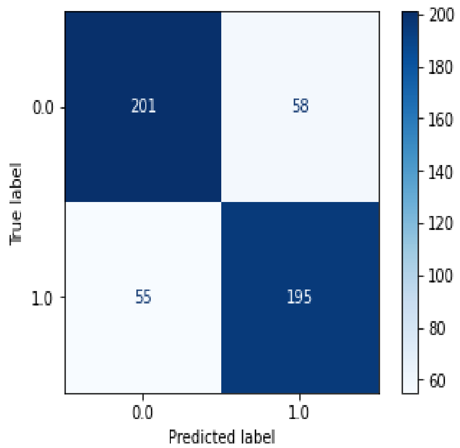

Fig 4.2. Confusion matrix representation for HN-Decision Tree model

4.3 Applying HN-ExtraTrees

Extremely Randomized Trees, often known as Extra Trees, is a machine learning ensemble algorithm. Applying CNN along with Extra Trees is what our HN-ExtraTrees model does. It is an ensemble of decision trees specifically, and it is connected to other ensembles of decision tree techniques like bootstrap aggregation (bagging) and random forest. Table 4.3 shows the performance evaluation of this model. It has achieved the accuracy of 88%, precision of 85%, recall of 93% and f1-score of 89%.

Table 4.3. Performance results (%) of HN-Extra Trees model

	PRECISION	RECALL	F1-score	support
0.0	0.92	0.84	0.88	259
1.0	0.85	0.93	0.89	250
Accuracy			0.88	509
Macro-avg	0.89	0.88	0.88	509
Weighted-avg	0.89	0.88	0.88	509

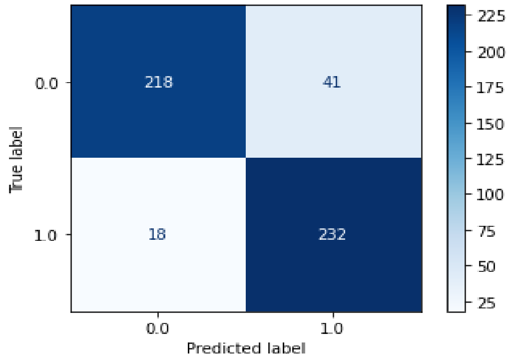

Fig 4.3. Confusion matrix representation for HN-Extra Trees model

4.4 Applying HN-Adaboost

AdaBoost is an ensemble learning technique that was initially developed to boost the performance of binary classifiers (sometimes referred to as "meta-learning"). AdaBoost uses an iterative process to improve poor classifiers by learning from their errors. Table 4.4 shows the performance evaluation of this model. It has achieved the accuracy of 84%, precision of 83%, recall of 86% and f1-score of 84%.

Table 4.4. Performance results (%) of HN-AdaBoost model

	PRECISION	RECALL	F1-score	Support
0.0	0.86	0.83	0.84	259
1.0	0.83	0.86	0.84	250
Accuracy			0.84	509
Macro-avg	0.84	0.84	0.84	509
Weighted-avg	0.84	0.84	0.84	509

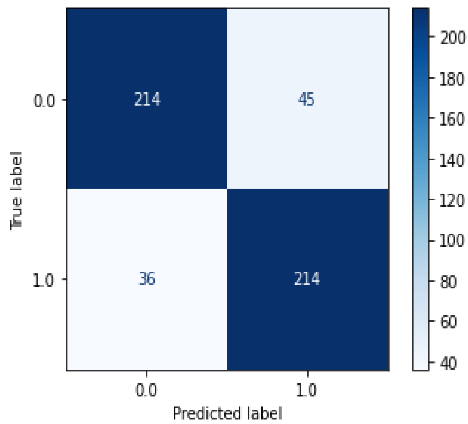

Fig. 4.4. Confusion Matrix representation for HN-Adaboost model

4.5 Applying HN-Xgboost

The Extreme Gradient Boosting framework, often known as XGBoost, is a gradient-boosted decision tree (GBDT) machine learning system that is scalable and distributed. It provides parallel tree boosting and is often considered to be the best machine learning library for solving problems including regression, classification, and ranking. Xgboost when applied together with CNN, in a hybrid model, achieves the highest accuracy of 92%, precision of 89%, recall of 95% and f1-score of 92%. Table shows the performance measures of every set of records achieved by HN-Xgboost model.

Table 4.5. Performance results (%) of HN-Xgboost model

	PRECISION	RECALL	F1-score	support
0.0	0.95	0.88	0.92	259
1.0	0.89	0.95	0.92	250
Accuracy			0.92	509
Macro-avg	0.92	0.92	0.92	509
Weighted-avg	0.92	0.92	0.92	509

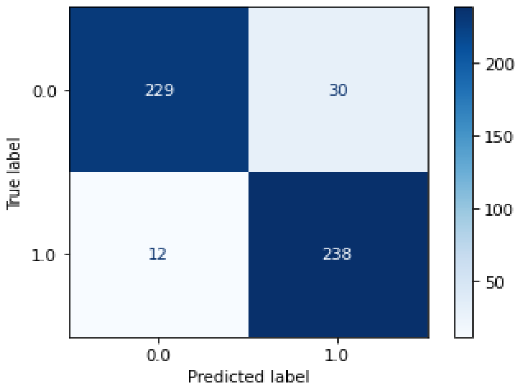

Fig. 4.5. Confusion matrix representation for HN-Xgboost model

4.6 Comparison Plot

Below fig. depicts the comparison chart drawn with respect to the different models used in this hybridization method. The highest value of the accuracy is acquired by HN-Xgboost model.

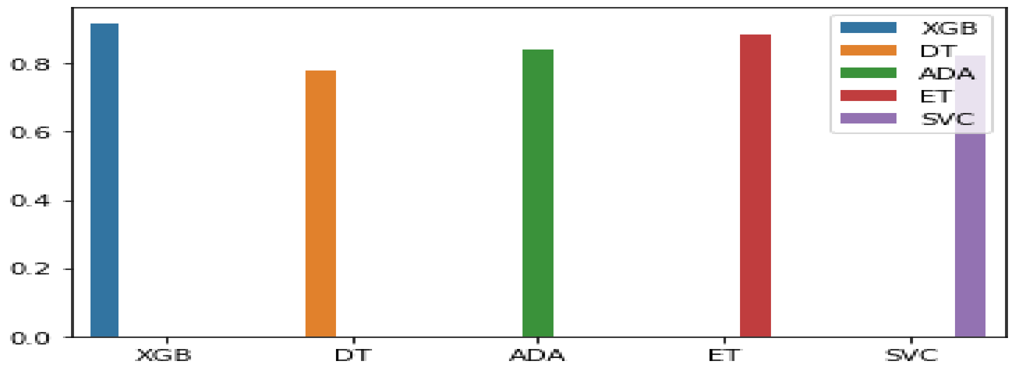

Fig 4.6. Comparison plot of all the models used.

Table 4.6. Comparison drawn along with recent works

Authors	Dataset Details	Partitioning	Techniques	Performance (%)
Wu et al. [37]	495 (COVID-19 =368, other pneumonia=127)	Training=80% Validation=10% Testing=10%	ResNet50	Accuracy=76, Sensitivity=81.1, Specificity=61.5, AUC=81.9
Hemdan et al. [38]	26(COVID-19, Pneumonia)	Training=80% Testing=20%	COVIDX-Net	Accuracy= 90% Precision=83% Recall=100% F1-score=91%
Zhang et al. [39]	1531 (COVID-19 = 100, Normal = 1431)	-	18 layer residual CNN	Accuracy=72.31% Sensitivity=90%

				Specificity=87.84%
Wang et al. [40]	1065 (COVID-19=740, COVID-19=325)	Random partitioning	Model Inception	Accuracy=79,3, Sensitivity=83, Specificity=67, Precision=55, F1-Score=63
He et al. [41]	746 (COVID-19=349, non-COVID-19=397)	Training=60% Validation=15% Testing=25%	CRNet	Accuracy=86, F1-Score=85, AUC=94
Zheng et al. [42]	630 (COVID-positive, COVID-negative)	Training=80% Testing=20%	DeCoVNet	Accuracy=90.1, Sensitivity=90.7, Specificity=91.1, Precision=84
Amyar et al. [43]	1044 (COVID-19=449, non-COVID-19=595)	Training=80% Validation=10% Testing=10%	Encoder-Decoder with multi-layer perception	Accuracy=86, Sensitivity=94, Specificity=79, AUC=93
HN-Xgboost	2529 (COVID-19=1262, non-COVID=1267)	Training=80% Testing=20%	Cnn + Hybrid Model	Accuracy=92% Precision=95% Recall=88% F1-score=92%

5. Conclusion

The COVID-19 epidemic is currently causing severe suffering throughout the world. Numerous people have already passed away as a result of inadequate care, inadequate facilities, or a lack of early detection. This hybrid model can aid in the automatic detection of COVID-19 from chest X-ray pictures in infected patients. It takes high-level characteristics from X-ray pictures using CNN. With great accuracy, the ensemble model distinguishes between COVID-19 and normal cases. Compared to other recent studies, the dataset has a substantial number of COVID-19 photos. An thorough experiment demonstrates that a hybrid model that combines the Xgboost and Convolutional Neural Network performs better, with 92 percent accuracy, 95 percent precision, 88 percent recall, and 92 percent F1-score.

epidemic. Given the results of these studies, it seems that the HN-Xgboost model, when used in conjunction with CNN, might be a useful medical tool for the automated identification of COVID-19 in the present pandemic, replacing manual radiography analysis. There are substantial disadvantages (such as inaccurately labelling certain COVID-19 positive cases as negative) yet it may be an alternative to manual radiological analysis and assist medical professionals discover COVID-19 from chest X-ray photographs. Early diagnosis and treatment can be improved, as well as epidemic control. This technology has a lot of promise to ease the burden on front-line medical professionals.

6. Future Work

Machine learning has a lot of potential in the healthcare industry. It is advised to focus future study on calibrated and ensemble approaches that could address peculiar issues more quickly and effectively than the current algorithms. Additionally, an AI-based application can be created employing a variety of sensors and features to recognize and assist with the diagnosis of diseases.

In order to avoid infection and death, the author proposes a method for detecting and predicting COVID-19 suspects early on. In order to identify and forecast COVID-19 suspects, the Kaggle repository is used to acquire data for the proposed technique. The framework is evaluated using machine learning algorithms on a dataset from the Kaggle source. Among the five machine learning approaches we presented, we used support vector machines, decision trees, naive bayes, logistic regression, and neural networks. XGBoost has been shown to be more accurate than 91% in the tests. The success rate of the decision tree is 75%, the success rate of the support vector machine is 86%, and the success rate of Extra Trees is 89%. The Adaboost has an accuracy rate of 86%. Through an early detection and prediction system, the suggested framework has the ability to stop and slow the spread of illness The dataset utilized in the work mentioned above is not so large that ensemble learning or other techniques would be beneficial. CNN approaches will also be tested in order to improve the model's performance metrics.

References

1. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.- H. Tian, Y.-Y. Pei, et al., Author correction: a new coronavirus associated with human respiratory disease in China, Nature 580 (7803) (2020). E7–E7.
2. M.S. Satu, M.I. Khan, M. Mahmud, S. Uddin, M.A. Summers, J.M. Quinn, M. A. Moni, Tclustvid: a novel machine learning classification model to investigate topics and sentiment in covid-19 tweets, Knowl. Base Syst. 226 (2021) 107126.
3. Worldometer, Covid-19 Coronavirus Cases, 2021. https://www.worldometers. info/coronavirus/coronavirus-cases. accessed: 2021-08-05.
4. Tang YW, Schmitz JE, Persing DH, Stratton CW. Laboratorydiagnosis of COVID-19: current issues and challenges. J Clinical Microbiol. 2020;58(6).
5. Zhang JJ, Cao YY, Dong X, Wang BC, Liao MY, Lin J, etal. Distinct characteristics of COVID-19 patients with initial rRTPCR-positive and rRT-PCR-negative results for SARS-CoV-2. Allergy. 2020;75(7):1806—1812.
6. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). Clin Infect Dis. 2020;71(15):778–785.
7. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, MeredithHR, et al. The incubation period of coronavirus disease 2019(COVID-19) from publicly reported confirmed cases: estimation and application. Annals Int Med. 2020;172(9):577–582.
8. Service mechanism for clinical decision support system for an intensive care unit F Ahmad, MR Beg, R Asthana - IMPACT-2013, 2013
9. G. Lippi and M. Plebani, "A six-sigma approach for comparing diagnostic errors in healthcare—where does laboratory medicine stand?" *Annals of Translational Medicine*, vol. 6, no. 10, 2018.
10. G. Lippi, A.-M. Simundic, and M. Plebani, "Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19)," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 58, no. 7, p. 1, 2020.
11. C. Sheridan, "Coronavirus and the race to distribute reliable diagnostics," *Nature Biotechnology*, vol. 38, no. 4, p. 382, 2020.
12. O. Wolach and R. M. Stone, "Mixed-phenotype acute leukemia," *Current Opinion in Hematology*, vol. 24, no. 2, pp. 139–145, 2017.
13. J. P. Kanne, ‘‘Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: Key points for the radiologist,’’ *Radiology*, vol. 295, no. 1, pp. 16–17, Apr. 2020.

14. G. D. Rubin, C. J. Ryerson, L. B. Haramati, and N. Sverzellati, ‘‘The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner Society,’’ *Radiology*, vol. 296, no. 1, pp. 172–180, Jul. 2020.
15. D. Dong *et al.*, ‘‘The role of imaging in the detection and management of COVID-19: A review,’’ *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 16–29, Jan. 2021.
16. H. Kim, H. Hong, and S. H. Yoon, ‘‘Diagnostic performance of CT and reverse transcriptase polymerase chain reaction for coronavirus disease 2019: A meta-analysis,’’ *Radiology*, vol. 296, no. 3, pp. E145–E155, Sep. 2020.
17. Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, ‘‘Chest CT manifestations of new coronavirus disease 2019 (COVID-19): A pictorial review,’’ *Eur. Radiol.*, vol. 30, no. 8, pp. 4381–4389, Aug. 2020.
18. S. Saqib, F. Ahmad (2021) An Integrative Decision Support Model for Smart Agriculture Based on Internet of Things and Machine Learning. Journal of Informatics Electrical and Electronics Engineering, Vol. 02, Iss. 02, S. No. 007, pp. 1-19, 2021. https://doi.org/10.54060/JIEEE/002.02.007
19. Ajlan AM, Ahyad RA, Jamjoom LG, Alharthy A, Madani TA (2014) Middle East respiratory syndrome coronavirus (MERS-CoV) infection: chest CT findings. Am J Roentgenol 203(4):782–787.
20. Haroon, Mohammad, Manish Madhava Tripathi, and Faiyaz Ahmad. "Application of Machine Learning In Forensic Science." Critical Concepts, Standards, and Techniques in Cyber Forensics. IGI Global, 2020. 228-239.
21. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43:635–640
22. R.M.Pereira, D.Bertolini, L.O.Teixeira, C.N.Silla, and Y. M. G.Costa, ‘‘COVID-19 .identification in chest X-ray images on flat and hierarchical classification scenarios,’’Comput. Methods Programs Biomed. vol.194, Oct. 2020, Art. no. 105532.
23. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning—An MIT Press Book. Cambridge, MA, USA: MIT Press, 2016.
24. Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, pp. 436–444, May 2015.
25. M. Wainberg, D. Merico, A. Delong, and B. J. Frey, ‘‘Deep learning in biomedicine,’’ Nat. Biotechnol., vol. 36, no. 9, pp. 829–838, Oct. 2018.
26. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, ‘‘A guide to deep learning in health care,’’NatureMed., vol.25, no.1, pp. 24–29, Jan.2019.
27. H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, ‘‘The rise of deep learning in drug discovery,’’ Drug Discovery Today, vol. 23, no. 6, pp. 1241–1250, Jun. 2018.
28. D.Shen,G.Wu,andH.Suk,‘‘Deep learning in medical image analysis,’’ Annu. Rev. Biomed. Eng., vol. 19, pp. 221–248, Jun. 2017.
29. Li M. Chest CT features and their role in COVID-19. Radiol Infect Dis. 2020;7(2):51–54.
30. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) outbreak: what the department of radiology should know. J Am Coll Radiol. 2020;17(4):447–451.

31. Kanne JP, Little BP, Chung JH, Elicker BM, Ketai LH. Essentials for radiologists on COVID-19: an update—radiology scientific expert panel. Radiological Society of North America. 2020;296(2): E113–E114.
32. Xu B, Xing Y, Peng J, Zheng Z, Tang W, Sun Y, et al. Chest CT for detecting COVID-19: a systematic review and metal analysis of diagnostic accuracy. EurRadiol.2020;30(10):1.
33. WuJ, WuX, ZengW, GuoD, FangZ, ChenL,etal. ChestCT findings in patients with coronavirus disease 2019 and its relationship with clinical features. Investigative Radiology. 2020;55 (5):257.
34. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology. 2020;296(2):200432.
35. Weinstock MB, Echenique A, DABR JWR, Leib A, ILLUZZI FA. Chest x-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: a normal chest x-ray is no guarantee. J Urgent Care Med. 2020;14(7):13–8.
36. Z.N.K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu, Brain tumor classification for MR images using transfer learning and fine-tuning, Comput. Med. Imaging Graph. 75 (2019) 34–46.
37. X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, and X. Yang, ''Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study,'' Eur. J. Radiol., vol. 128, Art. no. 109041, Jul. 2020.
38. Hemdan EED, Shouman MA, Karar ME. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv:200311055. 2020.
39. Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest X-ray images using deep learning based anomaly detection. 2020. aeXiv.2002.12338 vol. 1.
40. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, and M. Cai, ''A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19),'' MedRxiv, Jan. 2020. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.02.14.20023028v5
41. X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, ''Sample-efficient deep learning for COVID-19 diagnosis based on CT scans,'' MedRxiv, Jan. 2020. [Online]. Available: https://www. medrxiv.org/content/10.1101/2020.04.13.20063941v1
42. C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, and W. Liu, ''Deep learning-based detection for COVID-19 from chest CT using weak label,'' MedRxiv, Jan. 2020. [Online]. Available: https://www. medrxiv.org/content/10.1101/2020.03.12.20027185v2
43. A. Amyar, R. Modzelewski, H. Li, and S. Ruan, ''Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation,'' Comput. Biol. Med., vol. 126, Nov. 2020, Art. no. 104037.

fatima

by Similarity Report

Submission date: 19-Jul-2022 11:57PM (UTC-0500)
Submission ID: 1869641310
File name: THESISREPORTFATIMAALVI.docx (866.13K)
Word count: 18312
Character count: 101972

fatima

ORIGINALITY REPORT

13%	16%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.hindawi.com Internet Source	3%
2	www.softcomputing.net Internet Source	3%
3	www.diva-portal.se Internet Source	2%
4	www.ajol.info Internet Source	2%
5	Submitted to Vilnius Gediminas Technical University Student Paper	2%
6	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Student Paper	2%

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%