International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 566-578 www.ijaem.net ISSN: 2395-5252

Analysis of Covid-19 Cases - A Review

Pooja Rajesh
Student, B.Tech Computer Engineering, Mukesh Patel School Of Technology, Management and Engineering NMIMS University, Mumbai, India

Date of Submission: 10-06-2023 Date of Acceptance: 20-06-2023

ABSTRACT
A unique coronavirus, called Covid-19 has been the topic of discussion lately. It is a global pandemic which has affected many countries and millions of people. As a result of the outbreak and dissemination of Covid-19, major change has been brought about in people's lives. Desperate times call for desperate measures and such is the case of Covid-19. This paper talks about the various Machine Learning approaches that were used for Covid-19. This paper talks about the far and wide scope of Machine Learning and how it has proven to be helpful during this global pandemic. This paper dives into the supervised and unsupervised machine learning algorithms, talks about more advanced approaches implemented as well which include bagging and boosting algorithms and also dives into the domain of integration of Machine Learning with Deep Learning. Our review helps break down each approach, what it is and how it helps in prediction of Covid-19, pros and cons of each approach and what scope each ofthem has.

Index Terms: Coronavirus, Covid-19, Machine Learning, Models

I. INTRODUCTION

Wuhan in the Chinese province of Hubei is where the first complaints of Covid-19, which initially surfaced in December 2019, originated. It was thought to result in acute respiratory distress syndrome, which could be lethal (ARDS). A large family of coronaviruses includes the RNA virus known as SARS-CoV-2. Primarily made up of four structural proteins namely, Spike (S) , Envelope (E) , Membrane (M) and Nucleocapsid (N) , it is a positive-sense single stranded RNA. Coronavirus Disease 2019 (Covid-19) was the term coined by the World Health Organization (WHO) [1]. Every day, tens of thousands of new people from all over the world are reported to be positive. The virus mainly spreads by close personal contact between individuals, respiratory droplets, or touching contaminated objects. The most difficult part of the virus' transmission is that a person might have it for a long time without experiencing any symptoms [13]. It is considered to be a tricky illness because of the various mutation of strains and also the intensity levels, which might be modest to severe. The range of severity canresult in organ failure or even death. The seriousness of infections ranges from mild respiratory tract illness to progressive pneumonia. It could also eventually lead to multiorgan failure or death [5]. It is the rapid spread of Covid-19 which has called for desperate measures which means rapid, simple and effective technologies which can help predict the disease better which could potentially lead to fewer mortality rates.

Artificial Intelligence has started gaining popularity and has attracted a variety of fields in the research aspect which include

- engineering, economy, medicine and psychology [5]. It has been discovered that artificial intelligence (AI) is able to provide Covid-19 solutions that are quick but also accurate. They also do not compromise on the accuracy and provide a high diagnostic accuracy. It is because of the popularity of AI that hospitals and research based centers use an automated system instead of manual analyzing of data [6].

A number of comprehensive studies and reviews, have shown that Machine Learning (ML) has been crucial for the identification and detection of Covid-19 Cases. Machine Learning (ML) is often integrated with other technologies like Artificial Intelligence (AI) and Deep Learning (DL) in order to produce good results with high accuracy.

Computers utilize statistical models and machine learning training algorithms to carry out a variety of activities without explicit commands. Due to their precision, machine learning algorithms are employed extensively for predictions. Machine learning methods, however, face minimal difficulties. It can be challenging to select the suitable parameters when a model hasto be trained, or when a selection has to be made of a machine learning model to provide us with the best prediction. Machine learning algorithms can be used for extraction of data analytics and various hidden patterns. Machine-learning algorithms are designed for identifying complicated patterns and linkages in data when there are uncertain and complex connection patterns among risk factors [9]. One of the key insights into the diagnosis of Covid-19 that was revealed by machine learning techniques were the distinctions between Covid-19 pneumonia and other viral pneumonia using lung computed tomography (CT) scan as the initial screening or as a replacement for the real-time inverse transcriptase-polymerase chain reaction (RT-PCR) [9].

Machine learning algorithms can be categorised into three categories: supervised (which do classification and regression tasks), unsupervised (which perform clustering and dimension reduction tasks), and reinforcement learning (which performs classification and control tasks) [9].

The imaging modes which provide for various diagnostic tools for Covid-19 which are mainly Computer vision based, will facilitate doctors in making better decisions in the global fight against disease, as it provides for an automated second reading. Medical professionals and Radiologists may find it challenging to distinguish between Covid-19 pneumonia and other types which include viral and bacterial, only based on diagnostic imaging. Various neural networks like Convolutional Neural Networks (CNN) and Deep learning Artificial Neural Networks (DLANN), both have excelled in various medical imagecategorization applications. [12].

In order to allow researchers to continue exploring, inventing, and designing newer ways of improving accuracies and improving existing algorithms that would not only prove beneficial for detection of Covid-19 but future challenges as well, we aim at providing the maximum comparison between the different methods. This review will provide a concise yet in-depth comparison of all algorithms implemented during the existence of Covid-19 in the domain of AI, ML and DL.

II. LITERATURE REVIEW

The main objective is comparison and contrast of the various machine learning and deep learning algorithms and doing anin depth review to understand how these have been helpful and successful in combating the Covid-19 pandemic.

The various writers of [17], describe how they ran an online poll as part of their study to better understand how the pandemic and lockdown had psychologically impacted students participating in undergraduate and graduate programmes at schools and universities. This gives us a thorough understanding of the methodology, tools and techniques implemented followed by the interpretation and results of the study conducted. The 38-question questionnaire used for the online survey focused primarily on five main categories: fundamental information about the responding students, such as details about their internet connectivity, social lives, experiences with online learning, engagements, and general mood and thoughts throughout the COVID-19-related lockdown. Later, this was analyzed and shown with the use of Python, R, and MS-Excel, among other tools.

The study in [17] shows some interesting results. To show the association rules, it mostly employs R's apriori method. There were 583 people that completed the survey in total. The percentage of hours that students spend using cellphones is highlighted in this report, along with their social lives. The typical amount of time spent on devices during lockdown is between two and four hours, while the majority of people are seen between four and eight hours. In addition, the amount of time spent on social media for non-academic activities ranges from 2 to 6 hours. The use of social media platforms like Whatsapp, Instagram, Youtube, Snapchat, and Netflix, to name a few, has also increased dramatically throughout this time. Students spent more time with their family, which was one of the major outcomes that was highlighted. Other activities provided them time to rest and unwind, reconnected them with old acquaintances, and taught them new hobbies like cooking. A large percentage of the population spent the majority of their time sleeping, which was both good and terrible because it indicated that the kids werenot according to a set schedule. Students also missed coming to school or college, seeing friends in person, missing events, shopping at the market or mall, and other activities. Many of the students disapproved of the idea that online education made them closer and more connected to one another.

The authors, scattered across various departments and universities, in [22] also discuss how the usage of ML algorithmshas been useful in predicting subjective stress as a result of the worldwide pandemic. Participants in this study were divided into two groups based on their perceived levels of stress: high and low. This study assisted in identifying those people who were more likely to experience psychiatric symptoms similar to PTSD. The datasets underwent data preprocessing, which also included feature engineering. The training and test sets of data had K fold cross validation performed on

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 566-578 www.ijaem.net ISSN: 2395-5252

them in collaboration of a variety of Machine Learning models which include, Support Vector Machine (SVM), Logistic Regression, Naive Bayes and Random Forests. A PSS score and the preset assumptions were then used to compare the results.

Like how [22], discussed about the levels of perceived stress because of the pandemic, The genuises who contributed to

[25] offered a machine learning strategy on user emotions and twitter debates. To begin with, tweets on the Covid-19 epidemic are gathered using hashtags. These tweets were gathered with the help of the Twitter API. A purposive sampling strategy was utilized to sample the data. Sampling, data collection, and preprocessing of the raw data were all parts of data preparation. The major concepts used in the data analysis stage were unsupervised ML, sentiment analysis, and theme qualitative analysis. Similar to the poll performed in [17] that focused on students' opinions on the pandemic, sentiment analysis is crucial in assessing people's tweets throughout the global pandemic. To acquire an accurate result, tweets were filtered; for example, duplicates and non-English tweets were eliminated, creating a dataset that was exclusively composed of tweets. It was made sure that special characters were eliminated in order to properly preprocess the data. Additionally, it was made sure that @users and hashtags were eliminated. LDA (Latent Dirichlet Allocation), a type of unsupervised learning that enables the analysis of unstructured data, was applied to this preprocessed dataset. Along with Natural Language Processing, sentiment analysis was employed to determine the user's emotions at the time the tweet was published. The bigrams in the dataset, which are terms connected to the pandemic, assisted in detecting the emotions.

The top five results were displayed. Tweets have noted other ways to refer to the infection, such as coronavirus and Covid. Second, the infection was referred to as the "China virus." As a result of the stigma that is being generated by the Covid-19 Outbreak, this could end up being harmful. Thirdly, there was a lot of discussion about the New York epidemic and links between Covid-19 and the Chinese Communist Party started to emerge, suggesting that Covid-19 may have had political ties. Fourth, the Covid-19 outbreak resulted in conflicted emotions of trust, rage, and terror, as well as anticipation for the new actions that would need to be taken. Fifth, losing loved ones portrayed a lack of trust and a sense of anxiety [25].

Talking about ML models and how they are used for measuring the accuracy and how effective they are in predicting Covid-19, The scholars who contributed to the research in [26] provide an explanation by machine learning by proposing three methods that were mainly used to predict the deterioration of patient's health by using the APACHE-II risk prediction score. For ML models, a number of metrics were utilized, including the ROC AUC, APACHE-II score, Sensitivity, Specificity, and Accuracy. Neural networks, random forests, and the classification and regression method were the three machine learning models. For Artificial Neural Network's (ANN) output layer to be activated, an activation function of Softmax type with a cross-entropy error function is required, while the hidden layer is activated by the usage of a hyperbolic tangent activation function. Three layers are composed by the model. 16 normalized variables comprise the input layer, and the other 2 layersare hidden layer and output layer. In order to perform validation, the method used is K fold cross validation. An ensemble of 42 randomized decision trees were used to create the second model, called Random Forest. Two crucial and effective toolsare bagging and random feature selection. Gini impurity is mostly used in the third model, the Classification and Regression model, to evaluate the gains from each split and choose the best cutoffs for continuous variables. Testing and validation are done using the K fold cross validation approach.

Additonally, many analyses of statistical kind were done, such as for non-paramateric comparisons and continuous normally distributed data, soem tests were performed which include the Mann-Whitney U test and Student's t test. In order to examine the categorical comparisons, the test performed was the Fisher's exact test. Calculations were done for the various performance parameters such as positive and negative predictive values (PPV) and (NPV), sensitivity, specificity and accuracy. The balance accuracy metrics, which include the F1 score and Matthews correlation coefficient (MCC), were compared with the aforementioned measurements in order to compare the prediction ability.

According to [26], ANN has an advantage over APACHE-II in that it has limited sensitivity and K fold cross validation to increase external validity. It also has an improvement of 11 percent in accuracy. The accuracy of APACHE-II score was 12 percent worse than that of Random Forests, thereby proving Random Forests to be a better model with a better accuracy and prediction. The values for the ROC and AUC curves for the three mentioned models were 0.92, 0.93 and 0.90

respectively.

The authors of [24] like [26] discuss the various ML techniques used to forecast Covid-19 instances as well as the accuracy of these models. Several parameters were derived to gauge accuracy and were described in [26] in relation to Artificial Neural Networks, Random Forests, Classification, and Regression. Regarding the dataset with epidemiological labels which contain the positive and neagtive Covid-19 cases of Mexico, ML models of supervised type are made use of which mainly include Logistic regression, Decision trees, Support vector machines, Naive Bayes, and Artificial neural networks [24]. When these models were used, it was found in [24] that the Decision Tree performed the best and in comparison to all the other models, exhibited the highest accuracy. Sensitivity was taken into account, and SVM emerged as the top model. Naive Bayes was the best model when specificity was taken into account.

The scholars in [23], who contributed to this research, propose an enhanced model based on machine learning that has been used to forecast it to nations around the world and make them well aware of the possible threat of Covid-19. We demonstrate that a better fit may be obtained to create a framework for prediction when fitting the Generalized Inverse Weibull distribution using iterative weighting. In order to account for a more accurate and faster prediction for the growth pattern of the pandemic, the deployment has been done on a cloud computing platform.

The writers in [20] create a machine learning strategy that learns from the records of 51,831 tested people. The test data includes information from the next week. High accuracy was demonstrated by the model with the presence of binary features which were a total of eight in number. The eight binary features mentioned about include gender, people of age 60 and above, being aware of the contact of a person with an infected individual and the presence of the initial five clinical symptoms. Eight simple questions are used to train this machine learning model on data. In order to make predictions, the decision-tree base-learners created a gradient-boosting model. In the field of machine learning, gradient boosting is utilized to the utmost by many effective algorithms which asserts the fact that it has high regards for predicting tabular data. The basic criteria for evaluation include an auROC, other elements akin to [26], [24] positive and negative predictive values (PPV) and (NPV), false-positive rate (FPR), false-negative rate (FNR), false discovery rate (FDR), sensitivity, specificity and overall accuracy.

The scholars, of [19], mainly explain how they have used several traditional and ensemble machine learning techniques, such as those covered in [20], [24], and [26]. The application of bagging, adaboost, multinomial naive bayes, SVM, decision trees, and logisitic regression is demonstrated in this study. A pertinent dataset is chosen after the collection of patient data. The various ML algorithms listed above are applied, together with feature engineering and data preprocessing. Using machine learning methods, four groups of viruses—ARDS, SARS, COVID, and both—are predicted (SARS,COVID).

The results demonstrate that in comparison to other ML algorithms, the algorithms, Multinomial Naive Bayes and Logistic Regression have better testing accuracy and i.e 96.2 percent. [19].

The writers of [13] demonstrate how supervised machine learning algorithms are put into practice. The demonstration ofthe application of Linear Regression (LR), Exponential Smoothing (ES), LASSO, and SVM is shown in this study, which are similar to those seen in [19], [20], [24], [26]. The goal is to develop a system for forecasting. Three significant variables are used in the predicting: the numbers for fatalities, recoveries and newly confirmed cases. John Hopkins has gotten the dataset needed for use. The crucial evaluation metrics for this study include the calculation of R square score, Adjusted R square score, Mean square and absolute errors, and Root mean square error.

The Covid-19 dataset is first gathered, and then data preparation is used on it. Following that, it is segregated as a trainingset and a testing set. Various evaluation parameters are produced when validation of the test set is done by the applying thesaid trained model. The various ML algorithms are applied on the training set to produce a trained model. The final results show that SVM performs badly in all scenarios, with ES performing best when projecting new confirmed cases, fatalities, and recoveries, followed by LR and LASSO [13].

The authors in [10], talk about implementation of various machine learning and deep learning algorithms. The study's primary focus is on the models - Deep neural networks, polynomial regression, support vector regression, and recurrent neural networks incorporating large short-term memory cells are among the methods used. Each of these is implemented using Sklearn and Keras, and the outcomes project the number of confirmed, recovered, and fatal cases globally. The number of Covid-19 examples that are accessible is used to calculate the RMSE score of the techniques.

It has been found that the size of the deviation affects how long it takes to train the LSTM model, with a bigger deviation having a greater impact. As a result, the minmax scaler was used to scale the number of cases in order to fit the LSTM model, and the invert minmax transform from the ”sklearn” python library was then used to scale the predicted instances back to the original range. The best strategy has been proven to be PR [10].

The writers in [9], talk about the importance and usage of machine learning algorithms. In terms of performance and accuracy, it has been discovered that supervised machine learning algorithms perform better than unsupervised machine learning algorithms. Numerous algorithms are examined, such as ANN, CNN, Linear Regression, K-Means, KNN, Naive Bayes, and Logistic Regression.

This study’s main discovery is that a lot of categorization techniques were employed throughout the various sources that were used to construct it. The most prevalent algorithms in that order are Linear regression, ANN, CNN, Logistic regression, K-Means, KNN, and Naive bayes. [9].

All the authors in [5], who contributed to this study, mainly talk about how they used AI to produce a response to combat the pathogen. Some of the deep learning methods that are proposed in [10] include Extreme Learning Machine (ELM), Long/Short Term Memory (LSTM), and Generative Adversarial Networks (GANs).The speedier detection and treatment of the Covid-19 illness is a significant benefit of AI-based platforms. There have been numerous ways suggested for finding Covid-19. Many ways were put out in order to leverage AI-based techniques while also incorporating deep learning. Some of them include employing an ELM model to predict medications, LSTM ANN to accurately classify the best treatment option, and RNN to forecast the spread of infection.

This study largely advances our knowledge of the numerous deep learning and machine learning techniques used to overcome Covid-19. Large-scale data on Covid-19 patients can be integrated and analysed by advanced machine learning algorithms to help with a deeper understanding of viral spread patterns, speed up and more accurately diagnose patients, develop new, efficient therapeutic approaches, and even identify people who, based on their physiological and genetic characteristics, aremost susceptible to the disease. [5].

The writers in [11], in order to accurately anticipate Covid-19 in Hungary, they recommend a hybrid machine learning approach in substitution of susceptible-infected-resistant (SIR)-based models. To predict time series of infected individuals and death rate, the hybrid machine learning approaches of adaptive network-based fuzzy inference system (ANFIS) and multi- layered perceptron-imperialist competitive algorithm (MLP-ICA) are presented.

The dataset mostly includes statistical records on Covid-19 cases and death rates in Hungary. In order to provide a platform for prediction of Covid-19 Cases and fatality rates in Hungary, MLP-ICA is made use of which is a solid hybrid algorithm.The ICA is an evolutionary computation technique that looks for the best solution to various optimization issues. Cost function is specified when the model is merged with a neural network. The Cost function is specified for the error in the network that is produced. As weights and biases are adjusted, the network’s output is enhanced and the resulting error is reduced. The category of artificial neural networks includes ANFIS (ANNs). Fuzzy logic known as the Takagi-Sugeno fuzzy system is incorporated into the ANFIS architecture to enhance it. The mean absolute percentage error, root mean square error, and determination coefficient standard values were obtained in order to perform assessments. These components assess the target values and generate an index score to further assess the model’s effectiveness.

Both models demonstrate encourage results when the prediction of time series is talked about without the usage of presumptions that is necessary by epidemiological models. Both machine learning methods demonstrated promise as an alternative to epidemiological approaches for predicting the COVID-19 epidemic and calculating total mortality. MLP-ICA, however, fared better than ANFIS in terms of providing reliable results on validation samples [11].

The authors in [6] describe the creation of a decision matrix that is compared using several machine learning models and the entropy and TOPSIS methodologies. Naive Bayes, Neural Networks, Radial Basis Function, KNN, Stochastic Gradient Descent, Random Forests, Adaboost, Decision Tree, SVM(Polynomial), and CN2 rule inducer method are some of the comparison models that have been used and are fairly comparable to those that have been stated earlier in [10], [13], [24], [26]. Similar evaluation criteria including

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 566-578 www.ijaem.net ISSN: 2395-5252

Precision, Recall, F1 score, and ROC curves are also seen. X-ray images constitute the majority of the Covid-19 Dataset and those images include Covid-19 instances and those of ARDS, MERS and SARS patients.

The decision matrix for the Covid-19 dataset is developed, and it primarily reflects the aforementioned evaluation standards. The primary sub-criteria in the reliability group of criteria are represented by the parameters in a parameter matrix that is constructed. According to the aforementioned specifications, Three confusion matrices, each including the specific Covid-19 class parameters, are created from the confusion matrix (non Covid-19 and Covid-19).

The Decision matrix is processed using Integrated Entropy and TOPSIS techniques, and the results are compared to the models indicated earlier. For each criterion, TOPSIS determined the COVID-19 diagnosis model's best and worst performances. The SVM (linear) performs better than the other eleven diagnosis models, which were the furthest from the negative answerand the closest to the ideal solution, respectively. Linear SVM after thorough analysis and as per results, is considered as the top Covid-19 diganosis model. [6].

The authors in [7] are motivated to research about the usage of Artificial Intelligence (AI) to recognise Covid-19 from chest X- ray pictures fast and correctly. They want to make a recommendation for a trustworthy technique to recognise Covid-19 in chest X-ray images. For this detection, deep learning algorithms that have already been trained are used. Several previously learned deep convolutional neural networks were trained and tested using the transfer learning technique and picture augmentation (CNNs). Specificity, sensitivity, accuracy, and precision were the evaluation criteria.

(AP) which stands for Anterior to Posterior and (PA) which stands for Posterior to Anterior, chest X-ray images were stored in the database. In this study, eight distinct CNN models which were trained previously, underwent training, validation, and evaluation. The experiments were divided into two groups: the two-class problem which included augmentation with and without picture, and the three-class problem. The chest X-ray images underwent data preprocessing, or were scaled. Rotation and translation were the augmentation methods used. The deep layer characteristics were thoroughly examined. All of these layers maintained mixing the features recognised by the preceding layers to provide a forecast as the CNN continued to train to identify other features, such as edges or colour.

Inorder to provide validation for the training algorithms, the method employed was K fold cross validation. It was found that when picture augmentation was utilized to train the CNN models, DenseNet201 outperformed other deep CNN networks. Even without picture augmentation, the DenseNet derivative CheXNet was outperforming other networks. Dense201, a more advanced variant of DenseNet, outperforms CheXNet when trained on a large augmented dataset [7].

The writers in [14] use chest X-ray images like [6], [7] and it seeks to create COVID-CAPS, an alternative modeling framework known as the Capsule Network. Similar evaluation criteria to [6] and [7] include accuracy, specificity, sensitivity, and AUC. To possibly enhance COVID-diagnosis, CAPS's pre-training and transfer learning of the frontal chest x-ray pictures is carried out.

Each layer of the capsule network is made up of a number of capsules, each of which uses a number of neurons to represent an image at a particular place. Three Capsule layers and four convolutional layers make up the proposed Capsule network. Following batch normalization, average pooling, and reshaping the fourth layer to create the first capsule layer, the four convolutional layers are processed. As a result, the COVID-CAPS contains three integrated Capsule layers to carry out the routing by agreement process. The final layer of the capsule contains the instantiation settings for the two classes of positiveand negative Covid-19.

3D X-ray pictures are the network's inputs. There is an inclusion of four distinct class labels, namely, Normal, Bacterial, Covid-19 Viral, and Covid-19, in the dataset. Predicting whether a person is positive or negative is the main objective. The findings collected demonstrate that the COVID-CAPS performs satisfactorily with a small number of trainable parameters. Accuracy, specificity, and AUC might all be further enhanced by pre-training [14].

The genuises in [15] describe CovidGAN, a model, which is a technique and it facilitates for the creation of chest X-ray pictures which are fake. An Auxiliary Classifier Generative Adversarial Network (ACGAN) was used for the creation of CovidGAN. Artificial visuals are employed that can be useful for enhancing CNN's performance and this is done because of CovidGAN. The dataset constitutes of chest X-ray photographs of healthy people as well as Covid-19 patients. The architecture of the VGG16 network, which is employed to detect Covid-19, has twelve convolutional layers. Data augmentation employs a GAN variant known as ACGAN. The discriminator and generator architectures are used in the CovidGAN image enhancement. An adaptation of CGAN called ACGAN aids in stabilising the training procedure to produce high-quality images. A hyperbolic tangent (tanh) activation function, a kernel of size (5, 5), a stride of (2, 2), and ReLU activation are among the methods used in the generator architecture model's output layer. The discriminator model uses a stride that alternately shifts from (1, 1) to(2, 2), a kernel with a size of 1, and a LeakyReLU activation function (3, 3).

The discriminator model lies below the generator model. CovidGAN helps in synthesizing pictures of Covid-CXR and Normal-CXR. It has been noted that artificial enhancements created by CovidGAN aid in improving CNN's performance and accuracy. With normal data, accuracy was found to be 85 percent; with synthetic augments, accuracy was found to be 95 percent [15].

The authors in [1], Offer a completely automatic, quick, and accurate method that is machine independent that aids in identifying and quantifying the diseased areas in CT images from various sources. They fused a stack of 2D models to create 2.5D models to do 3D segmentation, which was inspired by 2D and 3D approaches. The key advancements made by the suggested methodology involve addressing data scarcity difficulties. To this end, a CT scan simulator is suggested, which performs dynamic alterations by fitting patient data at various time points. Additionally, a novel approach is suggested for breaking down the 3D segmentation into three 2D ones, which lowers the model's parameter requirements. The major component of the total procedure is training the scans, which results in a standardized dataset that is subsequently segmented. The testing data is preprocessed to create a standardized dataset, which is then used to apply additional segmentation. The Harbin and Riyadh databases are the main ones utilized.

The fundamental component of the preprocessing is normalization, which includes both spatial and signal normalization. Given that distinct CT scans have varying resolutions, resolution and dimension is unified with the help of spatial normalization. The X-ray attenuation coefficients of air and water are the basis of The Hounsfield Units used for signal normalization. Theyare normalized linearly. The next step is data augmentation. The main methods of data augmentation include simulation and modeling of the infection's dynamic changes. It focuses on three things: developing models, fitting real-time series of CT scan data, and enhancing data through simulation. As previously noted, the suggested 3D segmentation model is divided into three 2D models to facilitate visualization and trim the model's parameters.

Dice, recall, RMSE, and Pearson's correlation coefficient are the primary evaluation criteria employed (PCC). Comprehensive tests on datasets from many hospitals, countries, and machines showed that the proposed segmentation model performs much better than the other comparison models with respect to dice, recall, and worst-case performance as well as speed [1].

The writers in [2] like [1] also discuss finding Covid-19 using CT scans. A deep learning system which is of weakly- supervised nature, is used for the localization of lesion and for Covid-19 Classification. 3D Ct volumes were used for its development. The divison of each patient's lung region was facilitated by a UNet whihc was trained previously. To determineif COVID-19 will spread, a 3D deep neural network was given the segmented 3D lung region. First, participants in this study who had significantly enhanced CT chest images and were thought to have Covid-19 were included. The dataset would consist of this. A specified date was taken into account while dividing the data for the deep learning model into training and testing groups. According to this theory, scans performed before and after the date created the training data and the testing data, respectively. All participating patients had their pictures taken with a gemstone CT scanner. In order to diagnose Covid-19 and present it, various requirements are supposed to be fulfilled whihc includes epidemiological traits like travel or contact history, clinical signs and symptoms, chest CT, laboratory results, and real-time RT-PCR for SARS-CoV-2 nucleic acid testing. The suggested DeCovNet which is a neural network i.e a 3D deep convolutional one, is used for identification purposes from CT volumes for Covid-19. A trained UNet was able to acquire a 3D lung mask. The DeCovNet receives its input from this 3Dlung mask and a CT volume.

The main idea of a Covid-19 lesion localization which is weakly-supervised is that there is a combination of the activation zones that are produced by DeCovNet which is the deep convolutional neural network. Along with that, a lung segmentation method which is unsupervised is also employed. Strong memory was demonstrated by the DeCoVNet activation zones, however they frequently made incorrectly positive predictions. Preprocessing of UNet, preprocessing of DeCovNet, and data augmentation are the primary components of data preprocessing and augmentation. A constant learning rate of 1e-5, being used at 100 epochs, by usage of an Adam optimizer, is used for training of the network. It is followed by testing procedures. The trained DeCoVNet input each patient's preprocessed CT-Mask volume throughout the testing process. both the COVID-positive and COVID-negative probabilities. Then, for statistical analysis, the projected probability for each patient and their matching ground-truth labels were gathered. The main focus of statistical analysis is on ROC AUC curve analysis as well as evaluation factors like specificity and sensitivity.

The accuracy of the suggested DeCoVNet was at least 10 percent higher than that of the conventional techniques. This study offered a common and effective approach for creating medical AI for new disorders like COVID-19 [2].

The writers in [3] depicts an effort to use chest CT images to do highly interpretable and accurate Covid-19 screening. According to [1] and [2], for the detection of Covid-19, CT scans are the primary unit. In this article, we propose a technique called deep 3D multi-instance learning approach based on attention (AD3D-MIL), used for identifying each instance of a 3D chest CT at the patient level. The likely infection area can be followed by semantically generated deep 3D instances using AD3D-MIL. In order to make learning more approachable, Finally, the Bernoulli distributions of the bag-level labels have been discovered via AD3D-MIL. Multiple Instance Learning (MIL) and Attention Mechanism with MIL are the techniques used in this study. The deep 3D multi-instance learning approach based on attention was used in this work (AD3D-MIL). It consists of input that a deep instance generator utilising attention-based pooling processes, which displays a bag representation before performing a transformation function and generating a Bernoulli distribution that forecasts Covid-19.

The advantages of AD3D-MIL include scalability and interpretability. As a result of the attention-based MIL pooling approach, the AD3D-MIL can detect multiple crucial instances as opposed to only one critical instance. The AD3D-MIL algorithm is easier to interpret when multiple instance learning is used since the found key instances can pinpoint where COVID-19 infection zones are located. Comprehensive studies have shown that AD3D-MIL is capable of producing excellent but understandable outcomes [3].

The authors in [8] is a review that discusses how different strategies were compared and how none of them seemed to have provided satisfactory results. A large number of papers were filtered for this study, and the main task involved weeding out duplicates and reducing the number of studies to a manageable quantity for comparison. Deep learning and machine learning techniques were the only ones highlighted in the publications. Both private and public databases were utilized in this investigation. The majority of the private data came from countries including Belgium, France, Hong Kong, the United States, and Mainland China.

Many deep learning and machine learning techniques, such as VGG-16, ResNet-18, and DenseNet-121, were emphasized. The majority of articles categorized photos into one of three categories: Covid-19 pneumonia, non Covid-19 pneumonia, or normal. The performances of ResNet and DenseNet were noticeably superior to those of the other models evaluated. In order to find Covid-19, a method was employed by numerous articles which was Ct imaging. The application of machine learning methods with convolutional neural networks was also highlighted in numerous articles. SVM, Adaboost, random forests, logistic regression, linear regression, classification, and many other machine learning algorithms are among those that havebeen compared.

Both internal and external data analysis are performed. I order to internally validate data, development data was enoughthat originated from the same source. To externally validate the data, it was necessary for it to originate from various sources. Inclusion of both external and internal validation helps in better understanding of the generalizability of the algorithm. It isalso done to evaluate models. This contains the cross validation to assess how well the aforementioned models work.

After thorough analysis and according to the results, it is shown that there exist common methodological and reportingerrors, and the literature actually surveyed failed to match up to the robustness and reproducibility that was required to justifyits use in clinical practice. There are various factors in study design which can affect researchers and those include Bad data, poor application of machine learning methods, poor reproducibility, and biases [8].

The writers of [12], have a goal to use sophisticated deep learning image classification models to give overworked medical personnel a second set of eyes. The model of choice was a type of artificial neural network, popularly known as the Convolutional Neural Network (CNN), and this choice felt appropriate after initial comparison analysis of various well-known CNN models. The chosen VGG19 model is optimized for imaging modalities in order to demonstrate its applicability on the difficult and rare Covid-19 datasets. An image pre-processing phase is recommended for smooth testing of the various deep learning models and also the creation of a dataset which is trustworthy and capable of doing so.

Data comes from a variety of sources. Covid-19 labeled data makes up many datasets and these datasets mainly consists of pictures which include Normal, Pneumonia, Covid-19 and other non Covid-19 ones gathered from various sources and they are readily available because of the various investigations undertaken and it being a global pandemic. The usage of actual X-ray, ultrasound and Ct scan data was considered as the primary objective of the study. It did not take into account or employ synthetic data. A classification pipeline and model consideration were incorporated in the model construction. Reading the color image, converting it to gray, applying N-CLAHE, converting it back to color, scaling it to the default size, applying augmentations, and finally training/testing the model made up the majority of the preprocessing and classification pipeline. The selection of an appropriate deep learning model which is CNN-based out of the wide array of choices as it has proven to be extremely useful, needed for categorization of multimodal images was the main objective of this study.

The first step is making sure the model which has been selected, in this case the VGG19 model has its performance optimized. In order to make sure that is fulfilled for every experiment, a variety of parameters are adjusted including the learning rate, batch size, node size, and the drop rate. The accuracy metric of the experiment is influenced by the hyperparameter and learning rate selection of the hidden layer. It was noticed that when classfication is performed for Covid-19 and pneumonia vs normal, the greatest results are yielded in the ultrasound mode. Once again when this is performed, the same results are demonstrated i.e the best results yielded when classfication for Covid-19 pneumonia vs normal is done and in the ultrasound mode. It was demonstrated that categorization of Covid-19 scans against non Covid-19 scans performed better in case of Ct imaging modality. [12].

The authors in [16], suggest the usage of CT images inorder to diagnose Covid-19 from a number of characteristics which have been previously observed and already studied in [1], [2], [3]. A unified latent representation was proposed which could comprehensively encode all the information and offers a class strucutre which has a potential to exhibit seperability. The encoding of information is done for the various features which includes all the different aspects of the too. This research utilizes and suggests the usage of a thorough diagnostic pipeline for Covid-19 from community-acquired pneumonia. The procedure incorporates latent representation learning, which boosts stability, generalisation, and resilience.

In this work, the extraction from each image pre-processing, by facilitating a V-net model is showcased, The extraction includes the lung, lung lobes and pulmonary segments. Segmenting was also done with the infected lesions. Multi-view machine learning approaches are used since there are various heterogeneous feature types from CT scans that provide additional information to diagnose COVID-19.

There are several actions. The first is learning complete and structured representations, which primarily emphasizes com- pleteness and structure for latent representation. Fitting of diversity of data by effective and adaptive means into a space which is low-dimensional in nature is the first objective. This should be done keeping in mind, the data should comprise all different sorts of features for the latent representation which is intended. The second objective is to organise the latent representation that has been learned in a way that is appropriate for these two different pneumonia conditions. The second stage involves learning how to transfer original features onto latent representations. Training a latent based classifier that can discriminate between Covid-19 and CAP clearly is the final step. The testing stage is of utmost importance and it is considered crucial because this is where two of the most important components are used, which include, the latent representation regressor and the latent representation based classifier. This is in contrast to several models like SVM, LR, KNN, etc. Highest accuracy is said to have been exihibited by the proposed method which comes out to be 95.50 percent. On comparison of all the baseline approach to the proposed one which is the latent representation based model, it is observed that the diagnostic accuracy is improved by the proposed model to 19.9 percent from 6.1 percent. The latent representation model also portrays the fact that it gives a better performance than its competitors in terms of parameters like sensitivity and specificity, clearly winning by increasing it to

21.22 percent from 4.61 percent [16].

The authors in [18], propose the creation of a novel Covid-19 Lung Infection Segmentation Deep Network which they term as Inf-Net. The chest CT slices are utilized for automatic identification of the infected areas. The utilization of a parallel partial decoder is made in the Inf-Net in order to produce a world map from the aggregated high level features. Semi supervised segmentation approach is used for the Covid-19 infection segmentation. Explicit edge attention and implicit reverse attention are used to make sure that the representations along with the modelling of borders is improved upon.

Lung Infection Segmentation Network is the suggested technique (Inf-Net) Observed, two convolutional layers are first fed CT images to extract high-resolution, low-level (i.e., semantically weak) information. Three convolutional layers receive the low-level information as input and extract the high-level features. The Edge Attention Module is the next, and it improves the representation of objective region boundaries clearly. Edge data may offer helpful restrictions to direct the feature extraction process for segmentation. for the production of an edge map, a single filter feature is explicitly fed into a convolutional layer. The next step is the Parallel Partial Decoder, which combines the high level characteristics. The Reverse Attention Module comes next. Reverse attention is learnt by means of adaptive learning all at once by the three high level features instead of performing aggregation of features at all the levels.

Because it augments the training dataset with a substantial number of CT images, semi-supervised InfNet is utilized to addressthe issue of a lack of CT images. The framework's benefits can be divided into two categories. First of all, the training and selecting technique is straightforward and simple to apply. It is threshold-free and does not require measurements to evaluatethe projected label. Second, by avoiding overfitting, this approach can perform better than existing semi-supervised learning techniques. The Multi-Class Infection Labeling has also been expanded. Apart from the assessment conducted, Evaluation which is quantitative in nature, which may help in the assessment of various lung infections in a clinical environement may be of interest to clinicians. For providing richer data for the corresponding diagnosis and for treating Covid-19, a Semi Inf-Net is expanded to a multi class lung infection framework. Semi Inf extension is built on a framework for multi-class labelling that is infection region guided.

The suggested approach can distinguish between infections and healthy tissues using objects with low intensity contrast [18].

The writers of [21], suggest two optimization techniques, for diagnosing Covid-19, which include feature selection and classfication. Three cascading phases make up the suggested framework.

The gathering of data is the initial step. In this, primary datasets contain data which is photos, pertaining to both Covid-19 and non Covid-19. Inclusion of Axial images of patients that is done by Multi-Detector Computerized Tomography and onlyof high resolution, is there in the dataset. Dataset balancing is the next action to do. The features that were extracted might have issues with class imbalance. The methods employed for dealing with the class imbalance issue include Smote and LSH Smote. The method named Smote for any given minority class instance, lets say A, will help identify randomly, the k nearest minority class neighbours. The automatic hierarchical feature representation of CNN makes it a popular choice for object and pattern detection in images.

Three steps make up the suggested framework. The CNN training methods are used in the feature engineering process in the initial phase. The second phase consists of applying the LSH-SMOTE technique and the suggested SFS-Guided WOA to allow the selected features to be appropriately balanced and to employ feature selection respectively. The infected instances are categorised using the PSO-Guided WOA voting classifier method, which is indicated for the features selected in the second phase, in the third and final phase.

Three scenarios make up this article's experiments section. The first scenario is based on the suggested model's initial phase. Emphasis is made on the importance feature extraction exhibits because the effectiveness on usage of multiple CNN modelsof the experiment is proven in the subsequent stage. The recommended feature selection approach is thoroughly assessed and comparison is drawn with all the competing algorithms which makes up the second instance. Assessment of the proposed vote optimizer and see whether or not it can help improve the precision for categorization of Covid-19 instances makes up the third and final scenario.

Positive and Negative Predictive value (PPV) and (NPV), Precision, Accuracy, Sensitivity, Specificity and F-score are amongthe performance measures calculated for the first phase. The average error, average fitness, mean, best fitness, worst fitness, and standard deviation are the performance measures for the second scenario. AUC and MSE are two third scenario performance indicators.

Of the compared CNN models, the first scenario exhibits the highest classification accuracy. The efficiency of the suggested feature selection is examined and it proves to outperform the other examined techniques, some of which include the WOA algorithm. Covid-19 datasets, having calculated the lowest fitness value from the extracted features for which feature selection has been suggested, the decision has been made and this makes up the second case. Display of effectiveness is achieved by the suggested classification system. A value of 0.995 is observed for an AUC with binary predictions and a MSE of 2.495, it has been declared that the suggested voting classifier which incorporates the LSH Smote preprocessing method performs better than any of the other ensemble learning techniques that have been compared in the study and that makes up the thirdcase [21].

III. FUTURE WORK

Despite an extensive research on the variety of Machine Learning and and Deep Learning models along with a few custom approaches, there is still scope for future work which includes incorporation of more deep learning and also using a larger dataset in some cases which could help predict better and also mimic real time data.

IV. CONCLUSION

I can infer that there has been extensive research in the area of machine learning. As observed by the papers, I can conclude that a good deal of Machine Learning and Deep learning algorithms have been studied which have helped during the tough times of Covid-19. Machine Learning has shown breadth and depth while helping detecting Covid-19 and has sown various ways and forms it has been sued in. The integration of Machine learning with other technolgies like cloud computing, IOT and slightly advanced algorithms which are a part of deep learning, has added to the efficiency and helped improve the accuracy in prediction of Covid-19 cases. CT scans and chest X-ray pictures are the two methods of detection of Covid-19 that are most frequently detected. A fair amount of image acquistion and data augmentation has also been portrayed in the various papers. Deep Learning appraoches like CNN have been deemed as popular and efficient for identification of Covid-19 cases. The amount of comparision of models done, also gives a clear idea about when and which algorithms performs the best.

REFERENCES

[1]. Zhou, L., Li, Z., Zhou, J., Li, H., Chen, Y., Huang, Y., Xie, D., Zhao, L., Fan, M., Hashmi, S. and Abdelkareem, F., 2020. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. IEEE transactions on medical imaging, 39(8), pp.2638-2652.

[2]. Wang, X., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W. and Zheng, C., 2020. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. IEEE transactions on medical imaging, 39(8), pp.2615-2625

[3]. Han, Z., Wei, B., Hong, Y., Li, T., Cong, J., Zhu, X., Wei, H. and Zhang, W., 2020. Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. IEEE transactions on medical imaging, 39(8), pp.2584-2594.

[4]. Imran, A., Posokhova, I., Qureshi, H.N., Masood, U., Riaz, M.S., Ali, K., John, C.N., Hussain, M.I. and Nabeel, M., 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics in Medicine Unlocked, 20, p.100378.

[5]. Jamshidi, M., Lalbakhsh, A., Talla, J., Peroutka, Z., Hadjilooei, F., Lalbakhsh, P., Jamshidi, M., La Spada, L., Mirmozafari, M., Dehghani, M. and Sabet, A., 2020. Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. Ieee Access, 8, pp.109581-109595.

[6]. Mohammed, M.A., Abdulkareem, K.H., Al-Waisy, A.S., Mostafa, S.A., Al-Fahdawi, S., Dinar, A.M., Alhakami, W., Abdullah, B.A.Z., Al-Mhiqani, M.N., Alhakami, H. and Arbaiy, N., 2020. Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. Ieee Access, 8, pp.99115-99131.

[7]. Chowdhury, M.E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al Emadi, N. and Reaz,

M.B.I., 2020. Can AI help in screening viral and COVID-19 pneumonia?. IEEE Access, 8, pp.132665-132676.

[8]. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A.I., Etmann, C., McCague, C., Beer, L. and Weir-McCall, J.R., 2021. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), pp.199-217.

[9]. Kwekha-Rashid, A.S., Abduljabbar, H.N. and Alhayani, B., 2021. Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Applied Nanoscience, pp.1-13.

[10]. Punn, N.S., Sonbhadra, S.K. and Agarwal, S., 2020. COVID-19 epidemic analysis using machine learning and deep learning algorithms. MedRxiv.

[11]. Pinter, G., Felde, I., Mosavi, A., Ghamisi, P. and Gloaguen, R., 2020. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. Mathematics, 8(6), p.890.

[12]. Horry, M.J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M. and Shukla, N., 2020. COVID-19 detection through transfer learning using multimodal imaging data. Ieee Access, 8, pp.149808-149824.

[13]. Rustam, F., Reshi, A.A., Mehmood, A., Ullah, S., On, B.W., Aslam, W. and Choi, G.S., 2020. COVID-19 future forecasting using supervised machine learning models. IEEE access, 8, pp.101489-101499.

[14]. Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K.N. and Mohammadi, A., 2020. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. Pattern Recognition Letters, 138, pp.638-643.

[15]. Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F. and Pinheiro, P.R., 2020. Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. Ieee Access, 8, pp.91916-91923.

[16]. Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., Jiang, H., Wu, D., Sui, H., Zhang, C. and Shen, D., 2020. Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. IEEE transactions on medical imaging, 39(8), pp.2606-2614.

[17]. Khattar, A., Jain, P.R. and Quadri, S.M.K., 2020, May. Effects of the disastrous pandemic COVID 19 on learning styles, activities and mental health of young Indian students-a machine learning approach. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1190-1195). IEEE.

[18]. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J. and Shao, L., 2020. Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Medical Imaging, 39(8), pp.2626-2637.

[19]. Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R., Rouf, N. and Mohi Ud Din, M., 2020. Machine learning based approaches for detecting COVID-19 using clinical text data. International Journal of Information Technology, 12(3), pp.731-739.

[20]. Zoabi, Y., Deri-Rozov, S. and Shomron, N., 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj digital medicine, 4(1), pp.1-5.

[21]. El-Kenawy, E.S.M., Ibrahim, A., Mirjalili, S., Eid, M.M. and Hussein, S.E., 2020. Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. IEEE access, 8, pp.179317-179335.

[22]. Flesia, L., Monaro, M., Mazza, C., Fietta, V., Colicino, E., Segatto, B. and Roma, P., 2020. Predicting perceived stress related to the Covid-19 outbreak through stable psychological traits and machine learning models. Journal of clinical medicine, 9(10), p.3350.

[23]. Tuli, S., Tuli, S., Tuli, R. and Gill, S.S., 2020. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things, 11, p.100222.

[24]. Muhammad, L.J., Algehyne, E.A., Usman, S.S., Ahmad, A., Chakraborty, C. and Mohammed, I.A., 2021. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. SN

computer science, 2(1), pp.1-13.
[25]. Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y. and Zhu, T., 2020. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. Journal of medical Internet research, 22(11), p.e20550.
[26]. Assaf, D., Gutman, Y.A., Neuman, Y., Segal, G., Amit, S., Gefen-Halevi, S., Shilo, N., Epstein, A., Mor-Cohen, R., Biber, A. and Rahav, G., 2020. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Internal and emergency medicine, 15(8), pp.1435-1443.