Optimizing accuracy rate of Detection of COVID-19: A Machine Learning approach

K. Selvi*[1], K. Muthumanickam[2], P. Vijayalakshmi[3], S. Sakthivel[4]

[1]Professor, Department of IT, Paavai Engineering College, Namakkal, Tamil Nadu, India,
[2]Professor, Department of IT, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India,
[3]Associate Professor, Department of CSE, Knowledge Institute of Technology, Salem, Tamil Nadu, India,
[4]Assistant Professor, Department of CSE, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengode, Tamilnadu, India,

Emails: selvimidu@gmail.com; muthumanickam@kongunadu.ac.in; viji.vietw@gmail.com; ssakthivelaaasc@gmail.com

Abstract

COVID-19, one of the most highly transmissible diseases in the twenty-first century, has had a profound impact on global lifestyles. Recently, the medical industry has increasingly relied on machine learning, which shows promise in anticipating the presence of COVID-19. By using machine learning techniques, test result turnaround time can be accelerated, and medical personnel can promptly attend to patients' needs. These algorithms analyze various attributes to classify COVID patients and predict their likelihood of contracting the disease. This study aims to utilize X-ray images processed by machine learning algorithms to predict the occurrence of COVID-19 and enhance its detection rate. The paper outlines two strategies employing machine learning techniques: one for predicting the likelihood of infection and the other for identifying positive cases. Different machine learning algorithms, such as decision trees, logistic regression, support vector machines, naive Bayes, and artificial neural networks, were employed. The simulation results reveal that the artificial neural networks model outperforms other methods in terms of accuracy rate.

Keywords: Accuracy rate; COVID-19; Machine Learning; Optimization; Prediction

1. Introduction

Wuhan, a city in China, announced in December 2019 the first incidence of COVID-19, which then quickly spread to many other countries [1]. The revolutionary technology namely, artificial intelligence (AI), has found numerous uses in industries ranging from the medical industry. The utilization techniques like machine learning (MaL) and deep learning (DeL) improves scientists' ability to identify and analyze common variations that will result in disease. Machine learning (MaL) reliant techniques have the potential to accurately forecast the disease's progression and handle vast volumes of data effectively. These techniques offer valuable assistance in various areas, such as monitoring COVID cases, predicting outcomes, creating informative dashboards, diagnosing patients, prescribing appropriate medications, and even facilitating measures like social distancing to control the spread of the virus.

These techniques include common algorithms such as Convolutional Neural Network (CNN), in particular, are efficient methods for representation learning using multilayer neural networks [2], and they have produced outstanding performance solutions to a variety of problems in image cataloguing, decision creation, and object detection. Deep learning systems typically involve a number of processes, the most crucial of which is the classification of features and performance evaluation. the

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

general DeL reliant COVID-19 diagnostics pipeline. A specific model is developed using training data, and it is then tested against output data and validated against validation data.

2. Related Work

It is well known that to get an optimal solution for any linear programming problem using the direct simplex algorithm should be processed to be in standard form, the simplex method for solving an LP problem requires the problem to be expressed in the standard form. But not all LP problems appear in the standard form. In many cases, some of the constraints are expressed as inequalities rather than equations;

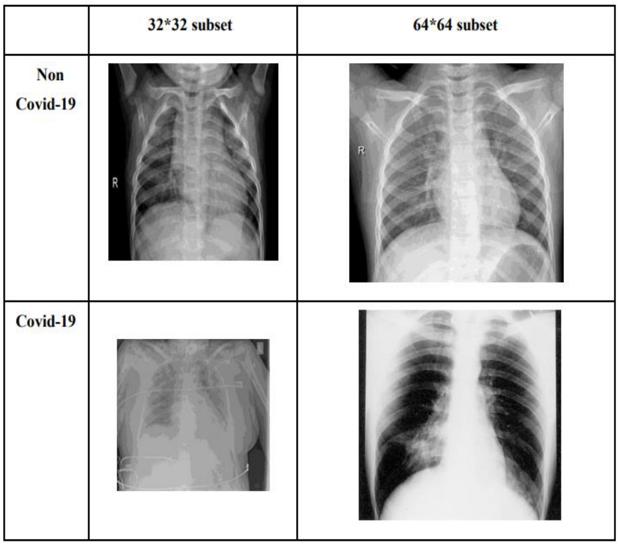

Figure 1: Covid and Non-Covid Subsets Sample Images.

The mining and cataloguing of characteristics are a crucial step during diagnosing COVID-19. When applied the DeL technique, it is possible to extract numerous operations, and classification is carried out using class labels. The exploration of MaL and DeL techniques improves scientists' ability to identify and analyze common changes that will result in disease. Figure 1 displays the detection components from the COVID-19 and normal pictures, respectively. These images indicate that the segmented chest X-ray images of COVID-19 patients are comparatively smaller than the original patient X-ray pictures. The discovery of the COVID19 coronavirus is now a critical undertaking for medical professionals and researchers.

In Wang et al. [2] study, a DeL system relied on radiographic alterations of COVID-19 images obtained as output of CT scan. This system aids in inferring the graphical features of COVID-19 well before infective testing, which significantly speeds up the verdict of the illness. The chest X-ray and CT can reveal symptoms similar to pneumonia, according to Hamimi's [3] research of MERS-CoV. According to Huang et al. [4], who established the clinical features on 41 COVID-19 patients, cough, heaviness, myalgia, or exhaustion were typical start symptoms. The chest CT scan for the 41 patients showed an abnormality and both had pneumonia.

A DeL system was created to inevitably slice all utilizing chest computed tomography namely CT to detect lung infections in a Shan et al. study [5]. In order to build an early prototype for the diagnosing COVID-19 viral contaminations in a steady setting, Xu et al. [6] used CT scans and in-depth education methodologies. The first proof of COVID-19 spreading through infected persons was made public by the Kok-KH squad at the University of Hong Kong [7]. Zhao et al. [8] developed a mathematical prototypical to ascertain the actual recorded COVID-19 infected cases during the first part of January

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

2020. It was inferred that 469 occurrences that were not documented between January 1, 2020 and January 15, 2020. Instances doubled in number after January 17, 2020, they added.

A mathematical approach was suggested by Tang et al. [9] to find out how likely it is that COVID-19 will spread. They found that 6.47 straightforward reproductions were feasible. A 7-day estimate of confirmed cases was also provided. In [10], the authors suggested a prototype for evaluating the COVID-19 death risk. The estimations are 5.1 percent and 8.4 percent for two different examples. Both scenarios' reproductive numbers were calculated to be 2.1 and 3.2 respectively. COVID-19 pandemic has been predicted to observe the entire body for identifying fractures, pneumonia, bone dislocations, lung poisons, and polyps using X-ray image.

Based on 565 Japanese collected from Wuhan residents who were displaced, for forecasting the COVID-19 afflicted cases in Wuhan on January 29–31, 2020, Nishiura et al. [11] recommended a model forecast. They assume that the real rate will be 9.5% and that between 0.3 and 0.6 percent of people would die. However, the count of Japanese residents who were uprooted from Wuhan is inadequate to compute infection and fatality rates due to its small size. Additionally, they anticipated the finest result after 23 January 2020. The authors in [12] utilized 47 patients' data to estimate COVID-19 sustainable human-to-human transfer. The author inferred that although 0.4 is communicated, 0.012 percent is spread if hospitalization effects take half as long to manifest as test findings.

Modern X-ray technology called CT scanning examines the internal soft tissues and organs of the body, which are very transparent and soft in nature [13]. Compared to CT, radiation is simpler, more powerful, more efficient, and less dangerous. Mortality may rise if infection of COVID-19 is not identified and treated right away. By examining a chest X-ray image, a pre-trained transmission model, and a deep convolution network, we have demonstrated an automatic COVID-19 prevision. The J48 procedure was utilized to classify the deep features in these deep-sensed models. Deep-learning models affect the parameter tweaking. A multi-impartial algorithm [14] is effectively employed to change CNN model parameters in order to solve this issue. The authors in [15] explored DeL and multi-objective optimization to identify people with the coronavirus by means of X-ray pictures. To effectively identify filthy patients, the J48 decision-tree technique categorizes the in-depth properties of infected X-ray pictures. Eleven different CNN-based replicas were created in this learning to recognize individuals with coronavirus pneumonia from X-ray images. Additionally, the CNN deep learning representation's parameters are specified using a sovereign penguin based optimization through a different goals. The authors in [16] exploited the application of DeL system to recognize the presence of COVID-19 infections using chest X-ray pictures. This review took advantage of the CNN model's ability to excerpt features through input images and foresee the presence of COVID-19 infected P-cases. In this article, we seek to recognize the existence of the COVID-19 disease and use machine learning systems to improve detection rate of COVID-19.

Senthil Kumar T et al0. (2022), utilised an imageries deepening and related thinning procedure on the imageries to extract the performance evaluation standards and rise precision [20]. Kumarganesh et al. (2018), recommended an ANFIS classifier approach may be utilised to find faulty portions from basis images with a precision of 96.0% [21]. A GA-based CNN categorization procedure was planned by Elayaraja et al. (2022) to segment the recognised area in the images, and it reached 99.37% Sensitivity (Se), 98.9 % Specificity (Sp), and 95.21 % Accuracy (Ac) [22]. For the determination of recognizing and segmenting image imperfections, Thiyaneswaran B et al. (2020) suggested the k-mean clustering method and attained a mean accuracy of 90.0% [23]. An ANFIS (Adaptive Neuro Fuzzy Inference System) categorizer process was suggested by Kumarganesh et al. (2016) for the categorization of imperfect portions from the basis pictures, and it attained 93.07% Sensitivity (Se), 98.79% Specificity (Sp), and 97.63% tumour segmentated Accurateness (Ac) [24]. Thiyaneswaran B et.al (2022), projected an AlexNet with ADAM solver reached a system accurateness of 98.21% [25].

A. Motivations

- ✓ This virus has caused us a great deal of suffering over the last two years, therefore that is why it developed.
- ✓ The findings will be extremely sensitive when using ML.
- ✓ This will make it easier for people to identify those who are impacted.

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

- ✓ A trusted party should formulate, rearrange and clean the information to be appropriate for perfect training. Nevertheless, the process of handling data to build the model can potentially compromise data privacy.
- ✓ Conventional machine learning models usually require a significant amount of time to be developed with satisfactory accuracy, causing potential delays for businesses, particularly newly established ones.
- ✓ For traditional machine learning to produce a model with sufficient accuracy, a significant amount of historical data must also exist.

3. Proposed System

The system architecture displays the user viewing a picture, analyzing an image, and producing output, and it also includes a module for training and prediction. The dataset was acquired from github and Kaggle for the proposed COVID-19 detection model's training process and testing process. The procedure for training has been provided the go-ahead to patch images using pre-trained CNN model during testing. The person's covid-19 affected X-ray image will be evaluated by integrating the pre-trained model with the evaluation model as depicted in Figure 2.

Datasets can be trained using ML in the training module, and photographs will be input according to the training in the prediction module. These images will display results such as Covid-19 infected Positive cases (P-case) and Negative cases (N-case). The prediction model receives the image as input once the user selects it from the local disc. Then, the block of analysis determines if the image is P-case or not.

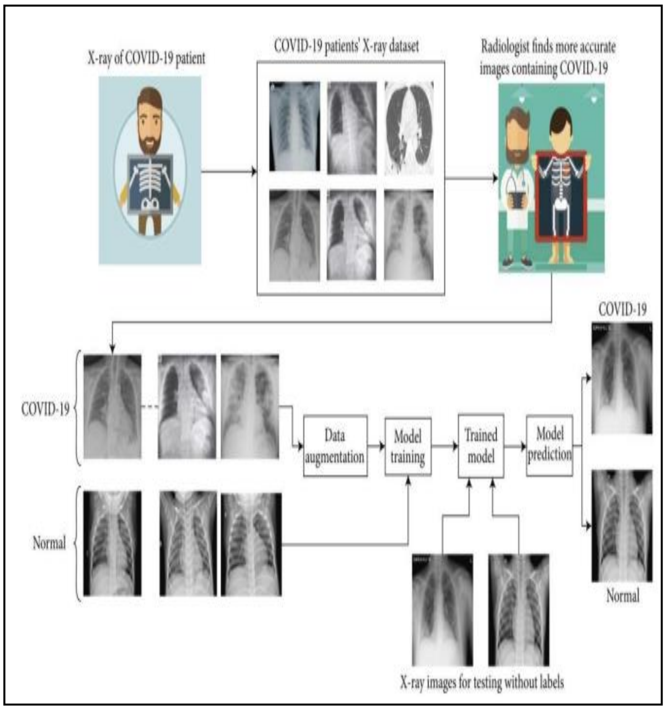

Figure 2: Vital components of the proposed method

The workflow of the suggested investigation is publicized in Figure 3. It starts with the collection of the primary dataset, which consists of two picture classes: one class of images belongs to covid-19 infected P-cases, and the other class of images belongs to healthy individuals. The concerned medical experts examined the dataset in the next immediate stage of the study and eliminated some of the X-ray images that were unclear based on diagnostic constraints and quality. Because each X-ray image was of high quality and was clear based on important diagnostic characteristics according to their experience, the resulting dataset was highly clean

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

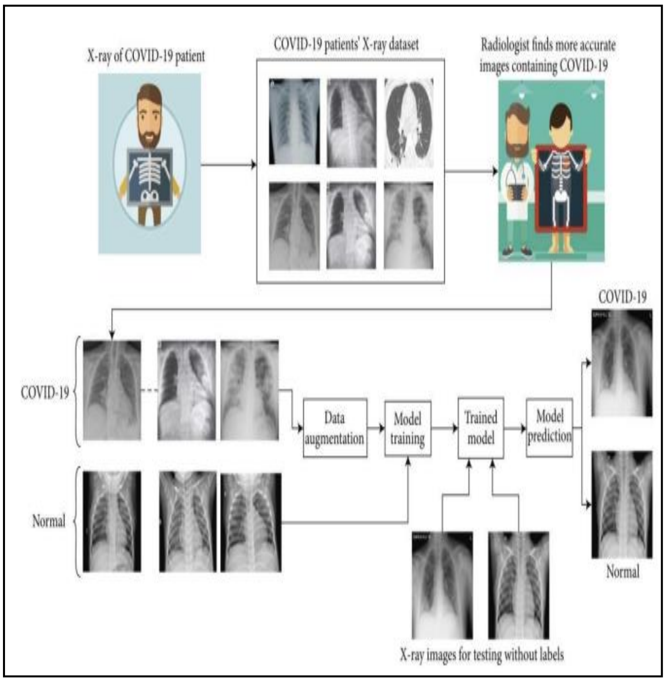

Figure 3: Detailed model of the Proposed method

During the third stage, the dataset was expanded by exposing conventional augmentation methods. The realtime dataset which was collected was utilized to train the model. Following training, the model's capability to identify diseases was evaluated. Both the validation dataset and the test dataset from the main dataset was used to test the proposed CNN [26]. The data flow model of the proposed model is elaborated as dataflow diagram in Figure 4**.**

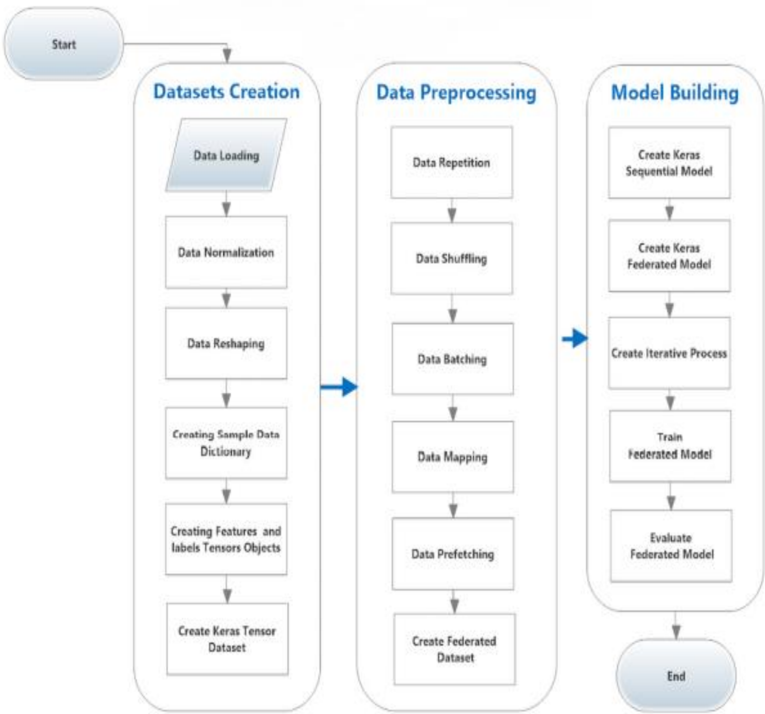

Figure 4: Dataflow diagram of the Proposed method

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

4. Results and Discussion

The system being suggested allows us to capture real-time images and use them as input for pre-processing. From these pre-processed video images, crucial facial expression features will be extracted.

A. Dataset Description

We integrated two diverse data collection to increase the sample size for our research. Kaggle was used to collect the initial dataset. We retrieved the database of viral pneumonia, normal, and chest X-ray pictures for COVID19 positive cases. During realtime data collection, 250 COVID-19 P-case images and 1,450 images that were normal. As it is possible to lead the system astray, downloading viral pneumonia photos allowed us to evaluate our model's ability to distinguish COVID-19 P-cases other infections like illnesses. The total count of X-rays used during training, testing, validation is listed in below Table 1.

Table 1: Dataset used for validating the proposed model

Dataset	Covid-19 infected data	Uninfected data	Total Images
Total	250	1450	1700
Training	150	1300	1450
Testing	100	150	250
Validation	95	575	670

B. Pre-Processing Phase

The preparation of test data plays a vital role in constructing a ML model. Often, the acquired data contains missing values, out-of-range numbers, and other uncontrolled data, which can potentially bias the results of an experiment.

- ✓ To manage missing values in our data, we employed the simple imputer function from the Sklearn Python library. The mean strategy is used to replace the missing values.
- ✓ One Hot Encoder, a Python programme that process categorical data and produce encode categorical data.
- ✓ Data Repetition - Repetition of data to simulate a client base.
- ✓ Data Shuffle - Data is shuffled to prevent repeating results.
- ✓ Data batching - To improve performance, data are divided into groups.
- ✓ Flattening array datasets to 1D array datasets is data mapping.
- ✓ Pre-fetching of data - Memory caching of data to improve performance.

C. Training and Testing

Positive cases, viral pneumonia, and ordinary cases of X-ray pictures are collected as the dataset is constructed using Kaggle. The prediction model is trained and tested using the dataset. The testing module uses the last 30% of the dataset after the training module has used 70% of it. The CNN model is a particular kind of ANN mainly exploited for image processing and recognition that is made specifically to handle pixel data. Technically, a sequence of convolutional layers, pooling, and filters are given as input X-ray image. Using probabilistic numerical values between 0 and 1, classifying an object by utilizing the function namely, Softmax. This is done in order to training and also testing the deep learning relied CNN models. Convolutional layers were added and the outcome was examined incrementally after starting with a single convolutional layer. The simulation of the suggested prototypical is given in Figure 5.

Various performance metrics are exploited to validate the effectiveness of various machine-learning schems as listed in Equation 1, Equation 2 and Equation 3.

$$\text{Accuracy} = (TN + TP) / (TN + TP + FN + FP) \quad (1)$$
$$\text{Sensitivity} = TP / (TP + FN) \quad (2)$$

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

$$\text{Specificity} = TN / (TN + FP) \qquad (3)$$

Where,

True Positive namely, (TP) corresponds to the total number of tests accurately identified as COVID-infected patients.
False Positive namely, FP) indicates the total number of tests mistakenly identified as COVID-infected patients.
True Negative namely, (TN) refers to the total number of tests correctly recognized as COVID-uninfected patients.
False Negative namely, (FN) signifies the total number of tests inaccurately identified as COVID-uninfected patients.

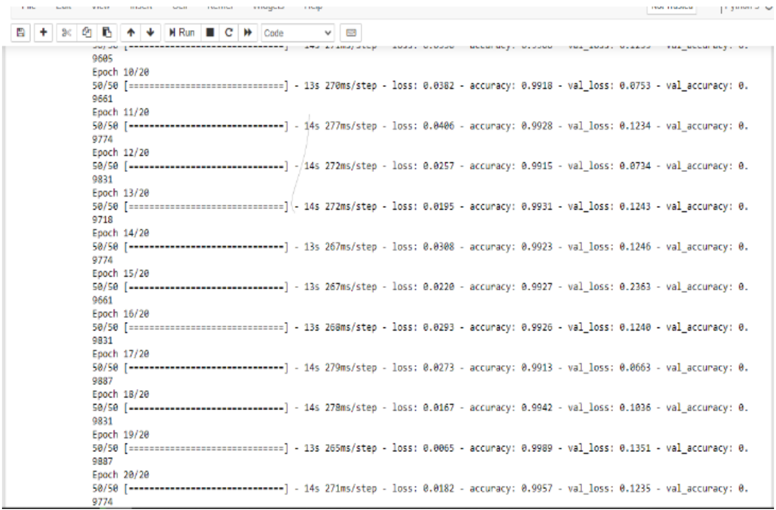

Figure 5: Simulation of the proposed model through google co-lab

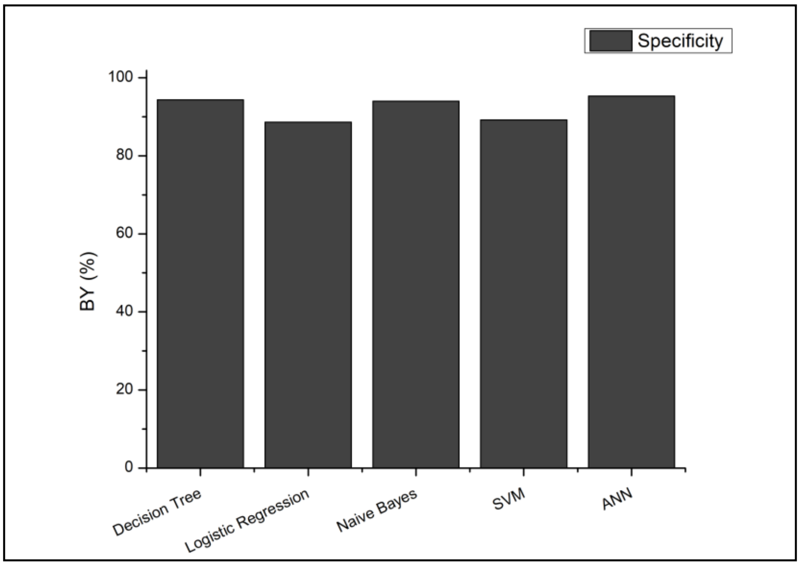

Figure 6: Performance Analysis based on Specificity

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

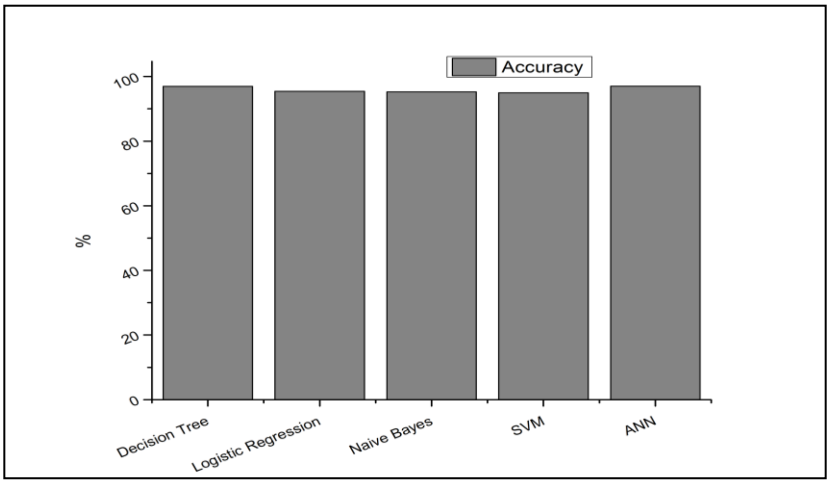

Figure 7: Performance Analysis based on Accuracy rate

The simulation values for a number of models are shown in Figures 6, 7, and 8. The findings indicate that the artificial neural network (ANN) model exhibits superior accuracy of 97.01% when compared to other classification models concerning the performance criterion, which is accuracy.

In light of metrics like sensitivity and specificity, decision tree outperforms other methods. Also, the decision tree and ANN methods produce better classification for COVID-19 detection, with respect to sensitivity and specificity of 95.90 %, and 95.31 %. The X-ray chest imageries of healthy and coronavirus-affected individuals are used by CNN's DeL algorithm to precisely diagnose the presence of COVID-19.

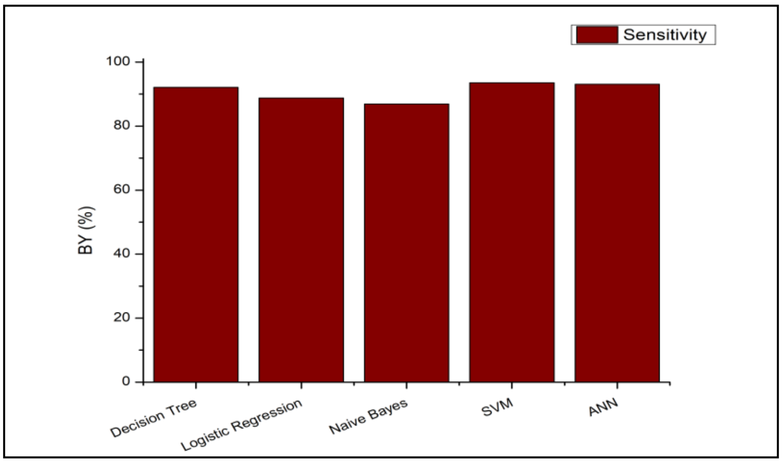

Fig. 1.

Figure 8: Sensitivity rate the various machine learning algorithms

5. Conclusion

Since its initial identification COVID-19 has presented a risk to human well-being and also influence the fiscal growth of a country. The virus parades comparable behaviors to any other virus-related pneumonias. The infection spread quickly as a result, making it impossible to maintain control of the situation. Clinical experiments show that the COVID-19 X-Ray imaging results show distinct findings. Thus, we suggested a MaL reliant covid-19 detection model to

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

enhance its accuracy rate which helps clinicians to offer better medical service to the infected people as quickly as possible. According to the simulation findings, the accuracy rate for the decision tree-based method, the logistic regression, SVM and the Nave Bayes methods is 96.95%, 95.43%, 95.27% and 94.98%. But ANN model performs better than other approaches with the accuracy rate of 97.01%.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Acter T, Uddin N, Das J, Akhter A, Choudhury TR, Kim S., “Evolution of severe acute respiratory syndrome coronavirus 2 (SARSCoV-2) as coronavirus disease (COVID-19) pandemic: A global health emergency”. Science of the Total Environment. 2020 Apr 30.

[2] Vijayalakshmi, P, Muthumanickam, K, Karthik, G, Sakthivel, S., “Diagnosis of infertility from adenomyosis and endometriosis through entroxon based intelligent water drop back propagation neural networks”. Journal of Intelligent & Fuzzy Systems, vol. 43, no. 3, pp. 2243-2251, 2022

[3] Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. “A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)”. MedRxiv Preprint 57:1451, 2020.

[4] Hamimi A, “MERS-CoV: Middle East respiratory syndrome corona virus: can radiology be of help? Initial single center experience”. Egypt J Radiol Nuclear Med, vol. 47, no. 1, pp. 95–106, 2016.

[5] Huang C, Wang Y, Li X, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan”. China, Lancet, 2020.

[6] Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., and Shi, Y., “Lung Infection Quantifcation of COVID-19 in CT Images with Deep Learning”. arXiv preprint arXiv:2003.04655, 1–19, 2020.

[7] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Chen Y, Su J, Lang G, Li Y, Zhao H, Xu K, Ruan L, Wu W, “Deep learning system to screen coronavirus disease 2019 pneumonia”. arXiv preprint arXiv:2002.09334, 1–29, 2020.

[8] Chan JF, Yuan S, Kok KH, “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster”. Lancet, 2020.

[9] Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, Lou Y, Yang L, Gao D, He D, “Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of january 2020: a data-driven modelling analysis of the early outbreak”. J Clin Med 9:388, 2020.

[10] Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, Wu J, “Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions”. J Clin Med 9:462, 2020.

[11] Jung SM, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, Kobayashi T, Kinoshita R, Nishiura H, “Real time estimation of the risk of death from novel coronavirus (2019- nCoV) infection: Inference using exported cases”. J Clin Med 9:523, 2020.

[12] Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, Linton NM, Jung SM, Yuan B, Suzuki A, “The rate of under ascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation fights”. J Clin Med 9:419, 2020.

[13] Thompson RN, “Novel coronavirus outbreak in Wuhan, China, 2020: intense surveillance is vital for preventing sustained transmission in new locations”. J Clin Med 9:498, 2020.

[14] https://biodiferences.com/diference-between-x-ray-and-ctscan.html

[15] Harsimran K, Rai A, Bhatia SS, Dhiman G, “MOEPO: a novel multi-objective emperor penguin optimizer for global optimization: special application in ranking of cloud service providers”. Eng Appl Artif Intell 96:104008, 2020.

[16] Gaurav Dhiman, V. Vinoth Kumar, Amandeep Kaur, Ashutosh Sharma, “DON: Deep Learning and Optimization-Based Framework for Detection of Novel Coronavirus Disease Using X-ray Images”. Interdisciplinary Sciences: Computational Life Sciences, vol. 13, pp. 260–272, 2021.

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025

[17] Pandiaraja, P., Muthumanickam, K., “Convolutional Neural Network-Based Approach to Detect COVID-19 from Chest X-Ray Images,” Lecture Notes in Networks and Systems, vol 370, pp. 2022.

[18] Senthilkumar T, Kumarganesh S, Sivakumar P and Periyarselvam K, “Primitive detection of Alzheimer’s disease using neuroimaging: A progression model for Alzheimer’s disease: Their applications, benefits, and drawbacks” Journal of Intelligent & Fuzzy Systems, Vol. 43(4), pp. 4431–4444, 2022. DOI: 10.3233/JIFS-220628.

[19] Kumarganesh S, Suganthi M. An Enhanced Medical Diagnosis Sustainable System for Brain Tumor Detection and Segmentation using ANFIS Classifier. Current Medical Imaging Reviews 2018; 14(2): 271-279. DOI: 10.2174/1573405613666161216122938.

[20] P. Elayaraja, et al. “An Efficient Approach for Detection and Classification of Cancer Regions in Cervical Images Using Optimization Based CNN Classification Approach”, Journal of Intelligent & Fuzzy Systems, Vol. 43 (1), pp. 1023 - 1033, 2022. DOI: 10.3233/JIFS-212871.

[21] Thiyaneswaran B, et al. Early detection of melanoma images using gray level co-occurrence matrix features and machine learning techniques for effective clinical diagnosis. Int J of Imaging System and Technology 2020; 1-13. doi.org/10.1002/ima.22514.

[22] Kumarganesh S, Suganthi M. “An Efficient Approach for Brain Image (Tissue) Compression Based on the Position of the Brain Tumor” International Journal of Imaging System and Technology 2016; 26(4): 237-242. doi.org/10.1002/ima.22194.

[23] Thiyaneswaran, et al “An effective model for the iris regional characteristics and classification using deep learning alex network” IET Image Processing, Vol. 17(1) pp. 227-238, doi: 10.1049/ipr2.12630, 2022

[24] Kumarganesh S, Suganthi M, “Efficient Lossless Medical Image Compression Technique for Real World Applications Using JOSE-Encoding” International Journal of Applied Engineering Research 2014; 9 (24): 24625-24640.

[25] Lalitha, S.P, Murugan, A. “Workflow scheduling and optimization using evaluationary method and deep learning algorithm in cloud” Multimedia Tools and Applications, 2024; Vol.83, pp. 78879–78896. https://doi.org/10.1007/s11042-024-18556-7

[26] Lalitha, S.P, Murugan, A. “Performance Analysis of Priority Generation System for Multimedia Video using ANFIS Classifier” International Journal of Computational and Experimental Science and Engineering, 2024; Vol.10(4), pp. 1320-1328. https://doi.org/10.22399/ijcesen.707.

DOI: https://doi.org/10.54216/JAIM.090101
Received: October 14, 2024 Revised: November 05, 2024 Accepted: January 08, 2025