International Journal of Industrial Engineering
2021;5(1):1-6.
ISSN: 2456-8449

Diagnosis of Covid-19 Using Hybrid Ensembled Convolutional Neural Networks

Durai Allwin[1*], Dhaniyasravani M[2], Dinesh Babu S[3]

[1]*Stavropol State Medical College, Stavropol, Stavropol krai, Russia.*
[2]*Progress Medical University, Gyumri, Armenia.*
[3]*Independent Researcher, Madurai, Tamil Nadu, India*

**Corresponding author: duraiallwindavid25@gmail.com*

Abstract. COVID-19 has arisen as a worldwide pandemic and has the potential to cause social, economic, and political problems that are catastrophic. Corona virus disease 2019 (COVID-19) has been a problem for the world, including India, notably during its second wave ever since it first appeared in December 2019. Disease caused by the Corona Virus in 2019, which was triggered by an initial case of severe acute respiratory syndrome caused by the Corona Virus 2. (SARS-CoV-2). An effective screening for this infection may facilitate the rapid and effective detection of COVID-19, which in turn can relieve some of the strain on the medical care system. An in-depth analysis of the provided dataset may enable the construction of one-of-a-kind and distinct types of AI computations, which, after being shown, can be subjected to additional processing and evaluation. A hybrid ensemble classifier was suggested in this research by coordinating Random Forest with SVM (Support Vector Machines), and CNN. In the accompanying case study, the suggested model was successful against a broad variety of machine learning methods, including SVM, Decision Tree, KNN, and Logistic Regression.

Keywords: SARS-Cov-2, Machine Learning (ML), SVM, KNN, CNN Decision Tree, Random Forest, Logistic Regression

INTRODUCTION

The Corona Virus Disease 2019 (COVID-19), which was caused by SARS-CoV-2 contamination [1-4], has spread all over the world and was declared a global pandemic by World Health on March 11, 2020[5-6]. As of April 2021, the general number of cases affirmed to convey this illness has reached up to 150 million of every 219 countries and regions, and there have been very nearly 3 million passing's brought about by this infection [7].The pandemic actually keeps on testing the clinical framework from one side of the planet to the other and this raised an abrupt interest for clinical gear, while the entire country was in lockdown, the clinical area was intensely tested by this infection and many courageous wellbeing laborers needed to lose their lives [8]. The transcriptase polymerase chain reaction, often known as RT-PCR, has recently been the diagnostic test of choice for SARS-Cov-2; nevertheless, this test has been in low availability in emerging and developing nations [9]. This will result in an increase in cases and may delay fundamental preventative actions for those who are unlucky [10]. The viable determination of COVID-19 can decrease the weight on the clinical/medical care framework [11].

Some models of expectations include several aspects to evaluate, including PC tomography (CT) filters, medical side effects, tests conducted at research centers, and a reconciliation of the aforementioned factors [12]. In this research, we provide a machine learning model that, by inquiring about a person's medical history, can determine whether they are a carrier of the SARS-CoV-2 illness [13]. This model was developed with the help of the data obtained from the website kaggle.com [14]. Subsequently, this methodology is capable of being used worldwide for the purpose of successfully screening and prioritizing testing for such infection for all individuals [15].

SVM gives an incredible yield while in contrast with various procedures for an extraordinary issue forecast inside calculations [16]. This is a Machine Learning (ML) model that utilized a characterization calculation for two-bunch grouping issues [17]. What it does is fundamentally assemble a learning model that appoints new guides to some gathering [18]. Measurement digging is valuable for getting critical data. The information was gathered from kaggle.com [19] At times, ML was utilized in Bibliometric examination of Coronavirus,

Received: 15.10.2020 Revised: 16.11.2020 Accepted: 07.12.2020

grouping and recognition of Covid-19 utilizing X-Ray Images, and reconnaissance of the illness by utilizing genomically - far reaching ML design and some forecast models join a few provisions to gauge like PC tomography (CT) filters, clinical side effects, research facility tests, and a combination of these given components and further in indications there are papers that have utilized gradient boosting machine model, LightGBM, auROC, SHAP which gives an exact output and furthermore ML is been utilized in forecast for different sicknesses like Diabetes, Heart infections, and so on [20]. Different neural network-based image processing applications are discussed in [21-22].

METHODOLOGY

The Random Forest, Support Vector Machines (SVM), and Convolutional Neural Networks were all part of the hybrid ensemble classifier. In the following case study, the suggested model was successful when other Machine Learning models, such as SVM, Decision Tree, KNN, and Logistic Regression, had failed. Figure 1 is a representation of the planned architecture and may be seen below.

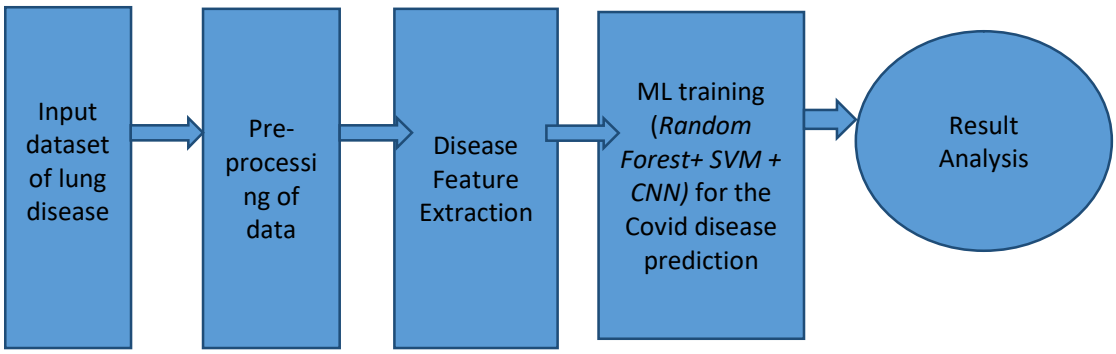

FIGURE 1. Proposed hybrid ensemble classifier for Covid-19 disease

The information of 316831 cases was utilized in this review, which was gathered from kaggle.com. Male, Female and Trans-sexual orientation sex were remembered for the review. The age bunch was arranged into 1-classifies and it's addressed in Figure 2, each having a distinction of 9 years, and a missing (data not given by the client) classification. The manifestations of COVID-19 are influenza like side effects and its subtleties are organized in Table 1.

TABLE 1. Symptoms of COVID-19

Most Common Symptom	Moderate Symptoms	Severe Symptoms
Fever, Tiredness and Dry cough	Flu symptoms include conjunctivitis, diarrhea, headaches, aches, severe pains, sore throat, loss of taste or smell, rashes, and chills.	Shortness of breath, chest pain and loss of speech

Random Forest

Random forest is a directed framework, which is utilized for order, relapse and different undertakings. In this, the calculation comprises of tree structures where the quantity of tree structures is straightforwardly relative to the exactness of the yield or result, and where each inner hub inside that specific tree compares to a trait and every single leaf hub addresses a class name.

SVM

SVM represents Support Vector Machine, it is a standard directed learning calculation, and this was presented by Vapnik in 1995. This is an AI model that utilizes characterization calculation for two-bunch arrangement issues. What it does is essentially construct a learning model that appoints new guides to some

gathering. This model can be utilized for all grouping and relapse difficulties. The model makes a hyperplane and partitions the information into classes bringing about all examples having a place with one specific class which could be ordered on one side and staying on the opposite side. So, we need to choose the class which has a higher Margin (distance between the planes).

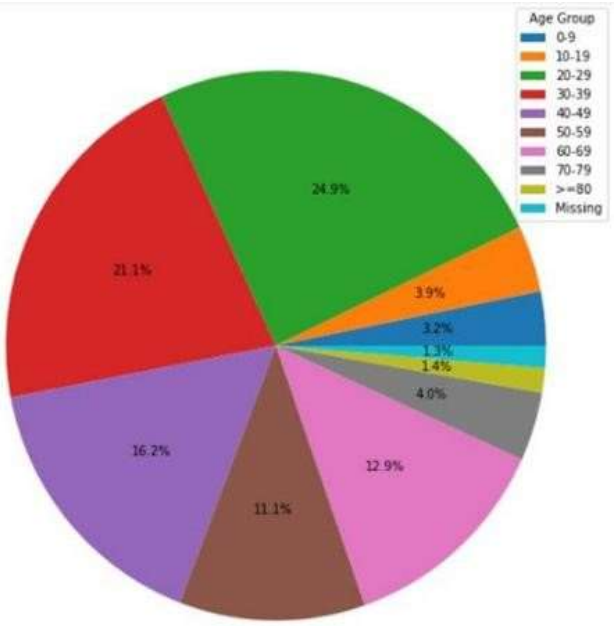

FIGURE 2. Percentage of COVID-19 cases per age group

Convolution Neural Network

A variety of grouping tasks, including those involving images, sounds, and words, make use of artificial neural networks. Neural Networks come in a wide variety of subtypes, each of which is optimized for a specific set of tasks. For instance, to predict how groups of words will be arranged, we make use of recurrent neural networks, more specifically LSTMs. Convolutional neural networks are utilized to organize images. In the next section, we will develop the fundamental building blocks for CNN's infrastructure. There are three different types of layers included in a typical neural network. The filters of a matrix and a convolution neural network are seen in Figure 3 respectively.

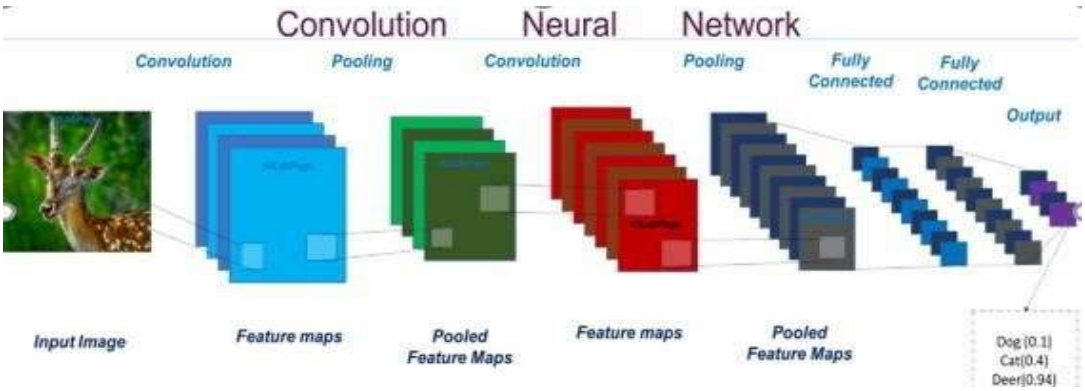

FIGURE 3. Filters of a matrix

Informational Layer: This is the layer in which our contributions to the model are made. The total number of items in our data is equal to the sum of all the neuronal connections in this layer (number of pixels if there

should arise an occurrence of a picture).

Secret Layer: The contribution made by the input layer is then sent into the secret layer once it has been processed. It's possible that there are a lot of hidden levels, but it all depends on our model and how much data there is. It's possible for each hidden layer to have a unique number of neurons, which are often easier to count than the total number of objects. The yield from each layer is processed by framework duplication of the yield from the layer below it with learnable loads from the layer below that. This is followed by expansion of learnable predispositions and initiation work, which ultimately results in a nonlinear organizational structure.

Yield Layer: After that, the yield from of the secret layer is processed via a calculated capacity like sigmoid or softmax, which turns the yield of each class into a probability score for each class. This score is then output. After that, the data are inputted into the model, as well as the yield from each layer is then determined. The process in question is referred to as feed forward. Following that, we use a blunder capacity to ascertain where the mistake lies. Cross entropy and square tragic mistake are two examples of the many different types of frequent errors that might occur. After that, we start to make our way back into the model by figuring out which companies are subsidiaries. This process is traced back to engendering, which is essentially used to limit loss. The execution was done utilizing Python programming language with Scikit-learn, Pandas, NumPy, and Matplotlib libraries utilizing CNN calculation.

The principle advantage is their precision in picture acknowledgment issues. The principle benefit of CNN contrasting with its archetypes is that it naturally distinguishes the significant elements with no human management. For instance, giving many images of benign and malignant, it learns provisions for each class without help from anyone else. CNN is likewise computationally productive.

RESULTS AND DISCUSSIONS

The measurement set is taken straightforwardly from Kaggle. The measurements set has many credits like Age, Symptoms, any contact with a transporter (Covid), and so on Records set is prepared to get the precise outcome and further it is tried. The accuracy is measured by the formula given by C5.0 algorithm which is,

ACCURACY = (TP+TN) / (TP+FP+TN+FN) Where the variables are, TP: True Positive TN: True Negative FP: False Positive FN: False Negative

The sensitivity or the true positive rate (TPR) is defined by TP / (TP + FN); while the specificity or the true negative rate (TNR) is defined by TN / (TN + FP)

The Result of the hybrid Machine Learning Models are as follows:

TABLE 2: Result Description

Different Algorithms	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
SVM	90	90	90
KNN	91	91	91
Decision Tree	92	93	92
Random Forest	91	91	91
Logistic Regression	90	90	91
Random Forest + SVM + CNN	99	99	98

Above Table 2 explains the different algorithms accuracy, sensitivity, and specificity. Initially the SVM algorithms accuracy, sensitivity and specificity were low percent. And following KNN, Decision Tree, Random Forest Logistic regression algorithm results also were not high.

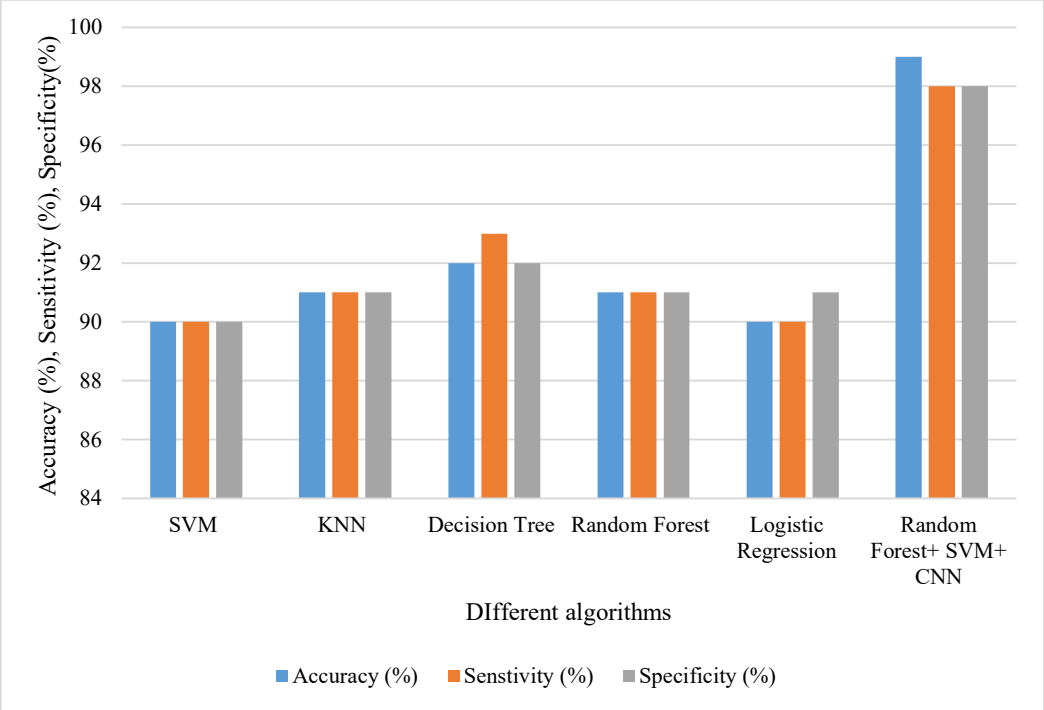

FIGURE 4. Bar Graph on Result Description

Figure 4 infers the proposed hybrid work that predicts the COVID-19 effectively when contrasted with other algorithms. Then combining the random forest, SVM, and CNN algorithms are giving the results effectively. The above table values show the high-performance metrics from the Random Forest + SVM + CNN. Then below figure displays the same as high performance of proposed method.

CONCLUSION AND FUTURE SCOPE

In our research, a large percentage of participants were aware of the prevalent symptoms of COVID-19, which is consistent with the results of other studies carried out in other regions of India. This is an important discovery that indicates a large gap in the practice of cleansing hands, which is a fundamental preventative measure for the management of illness. In this study, we employed four ML algorithms that performed very well to forecast the pace of COVID-19's dissemination. The efficacy of these ML models was examined by utilizing the data collected from patients who had COVID-19 sickness. These individuals had the infection. In this hybrid architecture, the machine learning algorithms that are used include SVM, KNN, Decision tree, random forest, and logistic regression. The C5.0 algorithm is used to determine the level of accuracy. The accuracy, sensitivity, and specificity of different algorithms are examined above. With the use of these comparison studies, we were able to test a variety of factors included within the dataset, and the results showed that the suggested hybrid framework was superior in terms of accuracy, sensitivity, and specificity. In the near future, machine learning classifiers will be necessary to accommodate the growing COVID-19 dataset.

REFERENCES

[1]. M. M. Lamers, J. Beumer, J. Van Der Vaart, K. Knoops, J. Puschhof, T. I. Breugem, R. B. Ravelli, J. Paul van Schayck, A. Z. Mykytyn, H. Q. Duimel and E. Van Donselaar, 2020, "SARS-CoV-2 productively infects human gut enterocytes," *Science*, **369(6499)**, pp. 50-54.

[2]. Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K. Y. Yuen and Q. Wang, 2020, "Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*,"

181(4), pp. 894-904.

[3]. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, 2020, "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2," *Science*, **367(6485)**, pp. 1444-8.

[4]. J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng and A. Auerbach, F. Li, 2020, "Structural basis of receptor recognition by SARS-CoV-2," *Nature*, **581(7807)**, pp. 221-4.

[5]. P. Vitiello, A. Sica, A. Ronchi, S. Caccavale, R. Franco and G. Argenziano, 2020, "Primary cutaneous B-cell lymphomas: an update," *Frontiers in Oncology*, **10**, pp. 1-7.

[6]. G. Qu, J. Chen, G. Huang, M. Zhang, H. Yu, H. Zhu, L. Chen, D. Wang and B. Pei, 2021, "A quantitative exploration of symptoms in COVID-19 patients: an observational cohort study," *Int. J. of medical sciences*.; **18(4)**, pp. 1082–1095.

[7]. C. D. Covid, R. Team, C. COVID, R. Team, C. COVID, R. Team, S. Bialek, E. Boundy, V. Bowen, N. Chow and A. Cohn, 2020, "Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States," *Morbidity and mortality weekly report*.; **69(12)**, pp. 343–346.

[8]. M. S. Rosário and I. C. Siqueira, 2021, "Concerns about COVID-19 and arboviral (chikungunya, dengue, zika) concurrent outbreaks," *Brazilian J. of Infectious Diseases*.; **24**, pp. 583-584.

[9]. J. M. Clarke, A. Majeed and T. Beaney, 2021, "Measuring the impact of covid-19," *bmj*., pp. 1-2

[10]. Gozes, M. Frid-Adar, H. Greenspan, P. G. Browning, H. Zhang, W. Ji, A. Bernheim and E. Siegel, 2020, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037,* pp. 1-22.

[11]. Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong and J. Shen, 2021, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Trans. on computational biology and bioinformatics*.; **18(6)**, pp. 2775-2780.

[12]. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng and B. Xu, 2021, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *European radiology*.; **31(8)**, pp. 6096-6104.

[13]. C. Jin, W. Chen, Y. Cao, Z. Xu, Z. Tan, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi and J. Feng, 2020, "Development and evaluation of an artificial intelligence system for COVID-19 diagnosis," *Nature communications*.; **11(1)**, pp. 1-4.

[14]. N. S. Punn and S. Agarwal, 2021, "Automated diagnosis of COVID-19 with limited postero anterior chest X-ray images using fine-tuned deep neural networks," *Applied Intelligence*; **51(5)**, pp. 2689-2702.

[15]. A. Pandey and G. Prakash, 2019, "Deduplication with attribute-based encryption in E-health care systems," *Int. J. of MC Square Sci. Res*., **11(4)**, pp. 16-24.

[16]. Feng, L. Wang, X. Chen, Y. Zhai, F. Zhu, H. Chen, Y. Wang, X. Su, S. Huang, L. Tian and W. Zhu, 2021, "A Novel Triage Tool of Artificial Intelligence-Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics," *MedRxiv*., pp. 1-47.

[17]. N. S. Punn, S. K. Sonbhadra and S. Agarwal, 2020, "COVID-19 epidemic analysis using machine learning and deep learning algorithms," *MedRxiv*, pp. 1-10

[18]. D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti and M. Ciccozzi, 2020, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data in brief*., **29**, pp. 1-4.

[19]. X. Mei, H. C. Lee, K. Y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. M. Robson, M. Chung and A. Bernheim, 2020, "Artificial intelligence–enabled rapid diagnosis of patients with COVID-19," *Nature medicine.*; **26(8)**, pp. 1224-1228.

[20]. M. A. Sarwar, N. Kamal, W. Hamid and M. A. Shah, 2018, "Prediction of diabetes using machine learning algorithms in healthcare," *In 2018 24th Int. Conf. on Automation and Computing (ICAC)* pp. 1-6.

[21]. S Murugan, S. Mohan Kumar, and T.R. Ganesh Babu, 2020, "CNN model Channel Separation for glaucoma Color Spectral Detection," *Int. J. of MC Square Sci. Res.* **12(2),** pp. 1-10.

[22]. F.J. Xavier, 2019, "Separation and Classification of Fetal ECG Signal by Enhanced Blind Source Separation Technique and Neural Network," *Int. J. Adv. Sig. Img. Sci*, **5(2)**, pp. 7–14.