iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44680

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Covid-19 Prediction Using Artificial Intelligence Techniques

Surekha Samsani[1], Sai Sandeep Ankadi[2]
[1]*Assistant Professor,* [2]*MCA. Student, Department of Computer Science & Engineering, University College of Engineering Kakinada (Autonomous), Jawaharlal Nehru Technological University Kakinada, Kakinada, India*

Abstract: ***The Covid-19 epidemic is one of the most serious worldwide health crises in recent years, having far-reaching repercussions for all civilizations throughout the world. Because of the virus's high contagiousness and fatality rate, the cost in terms of human lives lost is enormous. Millions of individuals have been infected, and many of them require ongoing care and monitoring. Tests such as RTPCR, Rapid Antigen, Chest Scan etc., which are being used to detect Covid-19 are painful and time consuming. Several investigations have found that the existence of Covid-19 virus can be predicted or analyzed by using Artificial Intelligence Techniques. Covid-19 patients have symptoms similar to other common illnesses. This project develops a system which analyzes an individual data relating to health status, Coughing, breathing, gender, age, certain pre-existing health conditions (i.e, asthma, diabetes) and geographical location to predict the infected one in a lesser time. Machine Learning approaches such as SVM, NB, SGD, ADABoost, ANN are used to discern between healthy and infected person. The Coswara database, accessible from GitHub, is taken to test the prediction ability of the classifier.***

Keywords: **Artificial Intelligence, Machine Learning, Support Vector Machine, Naive Bayes, Artificial Neural Network**

I. INTRODUCTION

The emergence of coronavirus has been considered a major threat to public health in almost all countries around the world during the last few years. This pandemic has impacted the lives of millions of people and continues to do so. Since the epidemic began, more than 80 million confirmed Covid-19 positive cases have been reported worldwide, with more than one million deaths, numbers that are, tragically, constantly rising [1]. The world's health services and network resources have been put to the test. Many patients require prompt treatment, as well as early diagnosis and monitoring, with healthcare personnel attempting to make the most of the limited resources available. Because of its high infection incidence, it's critical to develop procedures that can quickly and accurately detect Covid-19 and distinguish it from other types of influenza and pneumonia. The Internet of Things (IoT) and Artificial Intelligence (AI) can play an important role in healthcare by providing better insight into data and enabling affordable individualised therapy, often through the use of opportune wereable sensors. It is conceivable to improve not just the processing and storage capabilities of large IoT data streams (big data), but also to provide better patient care using AI-based diagnosis systems that are faster and more reliable. The Internet of Things (IoT), cloud and edge computing, wireless communication, mobile health systems, and trustworthy AI algorithms have all helped to improve disease diagnosis and treatment. In recent years, several monitoring systems that are effective in controlling chronic illnesses and emergencies have been presented. These systems have a variety of features, such as the ability to gather and analyse health data in real time for real-time monitoring, as well as accurate and faster patient data processing. In this study, look into the prospect of utilising machine learning (ML) techniques to aid in the early detection and assessment of the existence of the Covid-19 infection. The goal is to find the most reliable machine learning technique for detecting Covid-19 and integrate it into a smart mobile healthcare solution that can accurately differentiate between diseased and healthy people. Although the World Health Organization (WHO) currently recommends diagnosing Covid-19 via molecular tests in laboratories, this method could be very useful in tracking the virus internationally and diagnosing the pathology at an early stage. As a result, the key idea has been to discover the best ML technique in terms of detecting the presence of Covid-19, especially in asymptomatic persons, who are thought to be the pandemic's "silent drivers. "In reality, while symptomatic persons have been identified as the predominant source of SARS-CoV-2 transmission, asymptomatic people have a significant chance of being infected as well. Such subjects are the most difficult to track due to the lack of symptoms. As a result, meta data from the Coswara database, a publicly available crowd-sourcedRef database, was processed for use as inputs to the major machine learning algorithms in our research. Therefore, in our study, meta data selected from the Coswara database, an available crow sourced database were processed to be used as inputs of the main machine learning algorithms.

©IJRASET: All Rights are Reserved

II. RELATED WORK

Reference [5] shows how clinicians need fast and accurate artificial intelligence (AI) solutions to help them anticipate the severity of covid19 and the risk of patient mortality. By offering early pharmaceutical interventions, early prediction of patient severity could assist save hospital resources and reduce the number of patients who die. Using machine learning approaches, a prediction model was constructed to forecast different levels of severity risks for the COVID-19 patient based on X-ray pictures in this project. A total of six machine learning techniques were used. With 97 percent accuracy, the XGBoost classifier, SVM, had the best performance. Reference [6] provides a detailed examination of recently developed forecasting models and forecasts the number of confirmed, recovered, and death cases caused by COVID-19 in India. For prediction, correlation coefficients and multiple linear regression were used, and autocorrelation and autoregression were used to improve accuracy. The predicted number of cases agrees well with the actual values, with a 0.9992 R-squared score. The discovery suggests that lockdown and social distancing are two important factors that can help to slow the spread of COVID-19.

Reference [13] The goal of this research is to determine how one can easily predict if he or she has been infected with COVID-19. Another consideration is determining which COVID-19 symptom is more likely to indicate virus contamination. The required objective is obtained through the use of complex machine learning algorithms that can predict, to some extent, whether a person has covid-19 and whether the related covid-19 symptoms provided are relevant to the condition or not. The algorithm used is a LOGISTIC REGRESSION algorithm, which is used as a classification tool to separate data into binary results, which in our case is whether or not the person has covid-19 (YES OR NO). The dataset used to train this machine learning algorithm is obtained from online resources and public surveys. So far, the machine learning model has been able to predict the probability of infection with the virus by 66.89% accuracy further the model is able to relate if a given symptom is valid or not. The result was then improved using ANN but as it is computational expensive and due to lack of resources the expected performance cannot be met. The maximum accuracy achieved with ANNs was about 71%. Reference [4] Machine learning (ML)-based forecasting techniques have demonstrated their use in predicting perioperative outcomes and improving decision-making about future actions. Various forecasting methodologies are frequently utilised to address forecasting issues. To forecast the dangerous elements of COVID-19, four conventional forecasting models were applied in this project: linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and Exponential Smoothing (ES). Each of the models makes three types of projections in the next 10 days: the number of newly infected cases, the number of fatalities, and the number of recoveries. In [4] the results show that ES outperforms all other models, followed by LR and LASSO, which excel in predicting new confirmed cases, death rates, and recovery rates, respectively, while SVM performs badly in all prediction scenarios given the provided information.

III. MATERIALS AND METHODS

A. Dataset

In this project using Coswara Dataset, Coswara is a crowdsourced database developed at the Indian Institute of Science (IISc) in Bangalore and available on the GitHub platform [2]. The goal of the project, entitled Coswara after a combination of the words Covid-19 and Swara (sound in Sanskrit), is to gather sound samples in order to create a database that can be used to assess the efficiency of technologies for Covid-19 diagnosis. Each subject's coughing, breathing, and voice sounds were recorded, as well as information on their health status, gender, age, and any pre-existing health issues (such as asthma or diabetes) and geographic region.

B. Machine Learning Techniques

Various machine learning algorithms are used to predict covid-19. There are many other methods for covid-19 prediction but following Machine learning Techniques are used in this work. All the work was done on Jupyter Notebook.

1) *Support Vector Machine*: Support Vector Machine (SVM) is a popular Supervised Learning algorithm for Classification and Regression problems. It is, however, mostly used in Machine Learning for Classification Problems. The SVM algorithm's goal is to find the best line or decision boundary for categorising n-dimensional space into classes so that future data points can be easily placed in the correct category. The optimal choice boundary is called a hyperplane. SVM is used to select the extreme points/vectors that will aid in the creation of the hyperplane. The algorithm is known as a Support Vector Machine, and support vectors are the extreme examples.
2) *Naive Bayes:* The Naive Bayes algorithm is a supervised learning technique that solves classification issues and is based on the Bayes theorem. With a high-dimensional training dataset, it is mostly utilised in text categorization. The Nave Bayes Classifier is a basic and powerful classification method that aids in the creation of quick machine learning models that can make correct predictions. It's a probabilistic classifier, meaning it makes predictions based on the likelihood of an object. The Nave Bayes Algorithm is used in a variety of applications, including spam filtration, sentiment analysis, and article classification.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) *SGD Classifier:* Stochastic Gradient Descent is abreviated as SGD. Gradient Descent is a common optimization approach for machine learning and deep learning. Gradient descent is divided into three categories: batch gradient descent, stochastic gradient descent, and mini-batch gradient descent. Instead of using the full dataset, some samples are taken at random for each iteration. In SGD, each iteration is done with a single sample. The sample is scrambled and chosen at random for the iteration.
4) *ADABoost*: AdaBoost, or Adaptive Boosting, is a machine learning approach used as an ensemble method. Decision trees with one level, or decision trees with only one split, are the most frequent AdaBoost algorithm. Decision Stumps is another name for these trees. This approach generates a model by assigning the same weight to all data points. It therefore gives more weight to poorly classified points. All points with higher weights are given more weight. It will continue to train models until a low error is received.
5) *Artificial Neural Network:* The phrase "artificial neural network" is derived from biological neural networks, which are in charge of the development of the human brain's structure. Artificial neural networks, like the real brain, have neurons in multiple layers that are interconnected. These neurons are known as nodes. In artificial neural networks, dendrites of biological neural networks represent inputs, cell nuclei represent nodes, synapses represent weights, and axon represents output.

IV. RESULTS

Table 1 shows a comprehensive summary of ML approaches for covid-19 prediction. All these results shown in table 1 are obtained by training the model using different ML algorithms by applying Coswara dataset. 80% of the samples constituted the training set, while the remaining 20% constituted the testing set. Using Support Vector Machine, accuracy achieved on test set is 90%. Using Naïve Bayes, accuracy achieved on test set is 90%. Using SGD classifier, accuracy achieved on test set is 89%. Using ADABoost, accuracy achieved on test set 90%. And using Artificial Neural Network, accuracy achieved on test set is 93%. Among all the approaches highest accuracy is obtained by using Artificial Neural Network Algorithm.

Table 1
Accuracy of ML Algorithms

Algorithm	Accuracy
Support Vector Machine	90%
Naïve Bayes	90%
SGD Classifier	89%
ADABoost	90%
Artificial Neural network	93%

The Train/Test Accuracy of algorithms is depicted in the graph below. It says that the Artificial Neural Network outperforms all other algorithms in terms of accuracy. Artificial Neural Network having **93%** accuracy which is shown in below graph Fig 1 Comparison Graph.

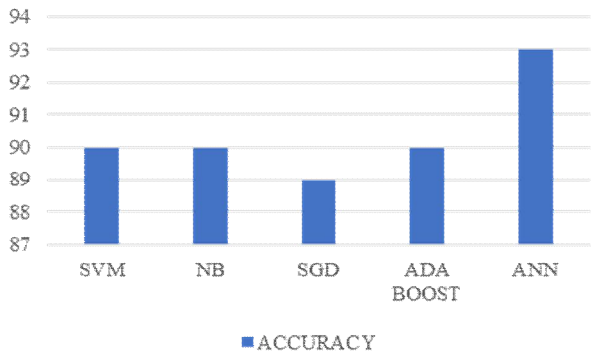


Fig 1. Comparison of accuracies of SVM, NB, SGD, ADABoost and ANN techniques

V. CONCLUSION

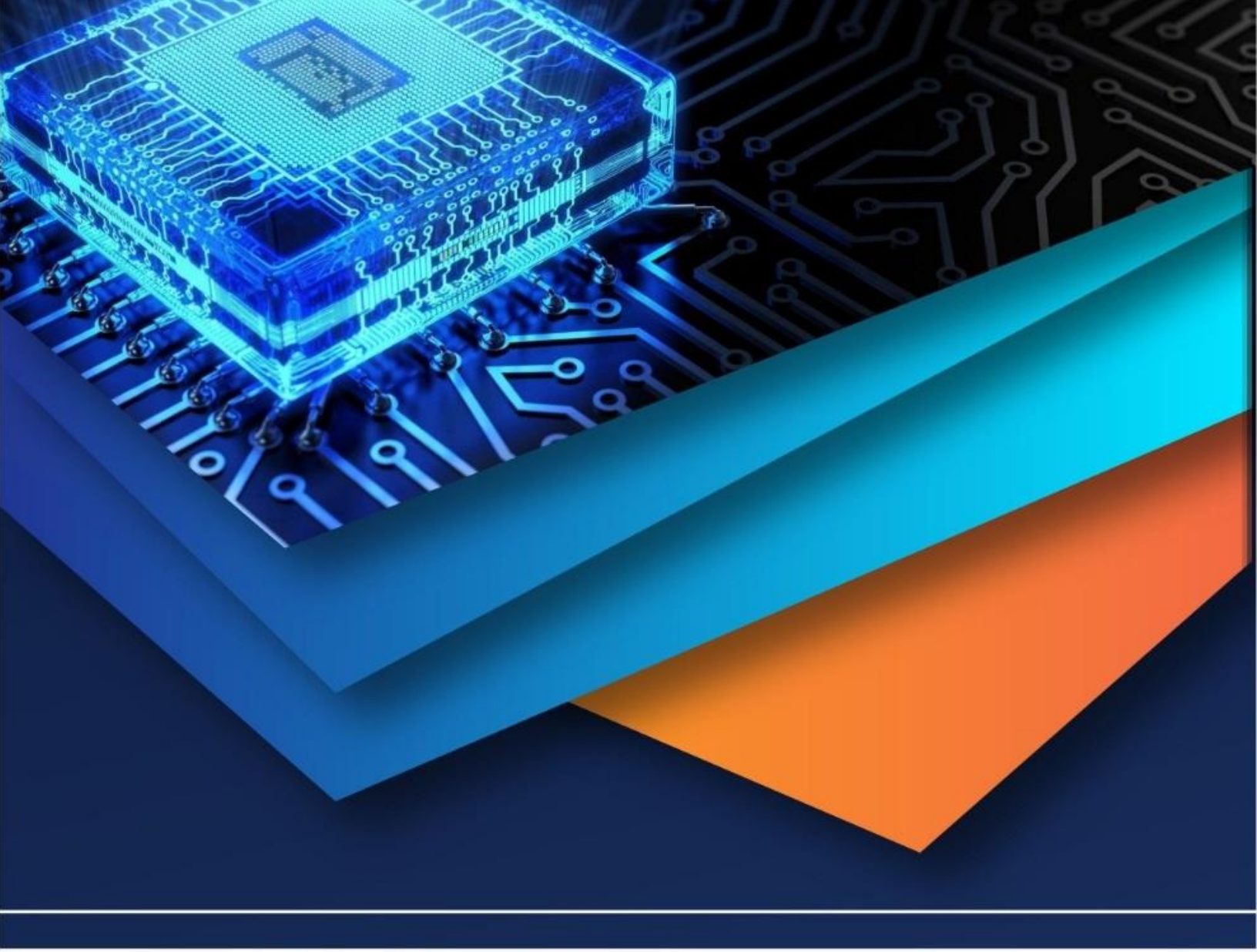
In terms of controlling the spread of the virus, IoT technologies and big data analysis using artificial intelligence algorithms can make a significant contribution. The value of these approaches to providing population-based treatment by limiting physical contact and achieving early diagnosis and monitoring of Covid-19 patients' health in a faster, more reliable, and less expensive manner is undeniable.

This work aims to study the classification accuracy of the main machine learning algorithms to identify the presence of Covid19. The goal was to find the most efficient method and incorporate it into a mobile health service. In other words, the goal was to create a system that could aid in the early detection of the Covid-19 condition, as well as serve as a rapid pre-screening test and a monitoring tool for patients' symptoms, decreasing the strain on national health systems. Various ML techniques were used in an analysis of samples chosen from a crowd-sourced database, the Coswara database, which is freely available. The results show that the ANN approach predicts Covid-19 with the highest accuracy. This algorithm distinguishes between infected and healthy person with an accuracy of around 93 %.

REFERENCES

[1] World Health Organization. (2020) WHO Coronavirus Disease (COVID-19) Dashboard. Accessed: Jan. 1, 2021. [Online]. Available: https://covid19.who.int/?gclidAIaIQobChMIht_qyL_K6gIVB7tCh2AIgwMEAAYASAAEgLyX_D_BwE/
[2] (2020). Coswara-Data. Accessed: Jan. 11, 2021. [Online]. Available: https://github.com/iiscleap/Coswara-Data/
[3] J. Laguarta, F. Hueto, and B. Subirana, ‘‘COVID-19 artificial intelligence diagnosis using only cough recordings,’’ IEEE Open J. Eng. Med. Biol., vol. 1, pp. 275–281, 2020
[4] F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in IEEE Access, vol. 8, pp. 101489-101499, 2020, doi: 10.1109/ACCESS.2020.2997311.
[5] S. A. -F. Sayed, A. M. Elkorany and S. Sayed Mohammad, "Applying Different Machine Learning Techniques for Prediction of COVID-19 Severity," in IEEE Access, vol. 9, pp. 135697-135707, 2021, doi: 10.1109/ACCESS.2021.3116067
[6] R. Kumari et al., "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 65-75, June 2021, doi: 10.26599/BDMA.2020.9020013.
[7] V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 116-123, June 2021, doi: 10.26599/BDMA.2020.9020016.
[8] M. Iqbal et al., "COVID-19 Patient Count Prediction Using LSTM," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 974-981, Aug. 2021, doi: 10.1109/TCSS.2021.3056769.
[9] E. Casiraghi et al., "Explainable Machine Learning for Early Assessment of COVID-19 Risk Prediction in Emergency Departments," in IEEE Access, vol. 8, pp. 196299-196325, 2020, doi: 10.1109/ACCESS.2020.3034032.
[10] T. Anwar and S. Zakir, "Deep learning based diagnosis of COVID-19 using chest CT-scan images," 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318212.
[11] S. Bhardwaj, H. Bhardwaj, J. Bhardwaj and P. Gupta, "Global Prediction of COVID-19 Cases and Deaths using Machine Learning," 2021 Sixth International Conference on Image Information Processing (ICIIP), 2021, pp. 422-426, doi: 10.1109/ICIIP53038.2021.9702560.
[12] P. Podder and M. R. H. Mondal, "Machine Learning to Predict COVID-19 and ICU Requirement," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), 2020, pp. 483-486, doi: 10.1109/ICECE51571.2020.9393123.
[13] D. N and K. K. G, "ANN based COVID -19 Prediction and Symptoms Relevance Survey and Analysis," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1805-1808, doi: 10.1109/ICCMC51019.2021.9418448.
[14] N. H. Hammadah, N. R. Das, M. Nayak and T. Swarnkar, "A hybrid approach of Deep Learning Algorithms for Identification of COVID-19 disease using Chest X-Ray Images," 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT), 2021, pp. 1-5, doi: 10.1109/APSIT52773.2021.9641398.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)