AN ARTIFICIAL INTELLIGENCE-DRIVEN PREDICTION MODEL FOR FORECASTING COVID-19 INFECTIONS: ENHANCING CLINICAL DECISION SUPPORT

1. **R.Umamageswari**, Assistant Professor/CSE ,Annapoorana Engineering College,Salem.

umamageswari6@gmail.com

2. **M.Saranya**,Assistant Professor/CSE ,Annapoorana Engineering College. Salem.

saranyam1295@gmail.com

3. **T.P.Udhayasankar**.Associate Professor /CSE,Annapoorana Engineering College, Salem.

udhaya75sankar@gmail.com

4. **B.Gunasekar**,Assistant Professor/CSE,Annapoorana Engineering College. Salem.

gunathedon2@gmail.com

5. **S.Shankar**,Assistant Professor/CSE, Annai Mathammal Sheela Engineering College, Erumapatti, Namakkal.

shankarscse@gmail.com

6. **K.Thangadurai**,Associate Professor /CSE,Annapoorana Engineering College. Salem.

thangadurai.me@gmail.com

Abstract

This article introduces a novel approach in developing an artificial intelligence (AI) driven prediction model, named the Random Forest COVID-19 Predictor, for accurately forecasting COVID-19 infections. The primary aim of this model is to provide timely support for clinical decision-making by leveraging predictive analytics on real patient data. With the urgent need to understand and effectively treat this new disease, coupled with resource constraints during a pandemic, accurate prediction models play a crucial role in guiding resource allocation decisions. Through extensive simulation experiments, the Random Forest COVID-19 Predictor demonstrates superior precision compared to existing models, making it a promising tool for enhancing clinical decision support during the ongoing COVID-19 pandemic.

Keywords

Artificial Intelligence, COVID-19, clinical support systems, Pandemic prediction

1. Introduction

The COVID-19 pandemic has posed unprecedented challenges to healthcare systems worldwide, necessitating innovative approaches to combat the spread of the virus and optimize resource allocation. In this study, we present a novel approach to developing an artificial intelligence (AI) driven prediction model for forecasting COVID-19 infections. The objective is to provide timely and accurate support for clinical decision-making by leveraging predictive analytics on real patient data. By accurately predicting the number of infections, healthcare providers can make informed decisions regarding resource allocation, such as hospital bed availability, staffing, and medical supplies [1].

The main contribution of this research lies in the development of the Random Forest COVID-19 Predictor, a novel AI-driven prediction model specifically tailored for COVID-19 infections. The key novelty of our approach lies in the utilization of the Random Forest algorithm, which has shown promising performance in various domains due to its ability to handle complex datasets and capture non-linear relationships [2]. By applying the Random Forest algorithm to COVID-19 data, we aim to enhance the accuracy and reliability of infection predictions.

Another significant contribution of this research is the use of real patient data for training and testing the prediction model. By utilizing actual patient data, our model takes into account the unique characteristics and dynamics of the COVID-19 virus, improving its ability to capture the intricacies of the disease spread [3]-[5]. This approach differentiates our research from existing studies that rely solely on synthetic or hypothetical data, thus increasing the practicality and applicability of our findings in real-world scenarios.

Furthermore, our research addresses the urgent need for accurate prediction models to guide resource allocation decisions during a pandemic. By accurately forecasting the number of COVID-19 infections, healthcare providers can allocate resources effectively, ensuring that adequate medical supplies, personnel, and facilities are available to meet the demand. This aspect of our research is particularly relevant in the context of limited resources and the critical importance of optimizing resource allocation to save lives and minimize the impact of the pandemic.

Additionally, through extensive simulation experiments, our proposed Random Forest COVID-19 Predictor demonstrates superior precision compared to existing prediction models. This finding highlights the efficacy of our approach in accurately predicting COVID-19 infections, reinforcing its potential as a valuable tool for clinical decision support. The improved precision of our model can enable healthcare providers to make more informed decisions, thereby improving patient outcomes and enhancing overall healthcare management during the ongoing COVID-19 pandemic.

In summary, this research introduces a novel AI-driven prediction model, the Random Forest COVID-19 Predictor, designed specifically for forecasting COVID-19 infections. By leveraging real patient data and the Random Forest algorithm, our model offers improved accuracy and reliability in predicting infection rates. The ability to make precise predictions in a timely manner can significantly contribute to informed clinical decision-making and resource allocation strategies, ultimately aiding in the effective management of the COVID-19 pandemic.

2. Related works

In recent months, the identification and detection of COVID-19 have been extensively studied. One area of focus has been on the use of CT scans and chest X-ray images for identifying COVID-19 cases. Another area of research has been on predicting the confirmation, recovery, and mortality rates associated with COVID-19. The rapid global spread of the virus has made it challenging to detect exposed individuals who may not exhibit immediate symptoms. Therefore, there is a need to estimate the actual number of affected individuals regularly in order to implement necessary precautions [6]. Artificial Intelligence (AI) has emerged as a promising alternative to conventional testing methods for COVID-19, offering a quicker and cost-effective approach. Several studies have utilized AI for predicting COVID-19 cases and diagnosing infections based on chest X-ray images. AI has been widely applied across various fields of study, and one of its notable advantages is its ability to detect patterns in images that are not visible to the human eye [7]. In this analysis, the focus was on two main aspects: the detection of COVID-19 and the estimation of the future number of affected individuals. Existing models for COVID-19 detection were found to be subpar, highlighting the need for more specialized and accurate models. The authors [8]proposed that research findings related to COVID-19 should be made publicly available to facilitate the development and implementation of improved identification and prevention models. In one study [9], the author combined PCR RT with CT scans to investigate 51 patients with respiratory symptoms and fever who had a history of travel or residence in endemic areas. Given the limited availability of PCR-RT kits and the increasing number of COVID-19 cases, the implementation of an automated detection system was suggested as an additional diagnostic tool to curb the spread of the virus. Another study [10] focused on the early stage of COVID-19, highlighting the limitations of real-time RT-PCR in providing accurate positive results. The authors developed an early screening model using deep learning techniques to differentiate between pneumonia, flu, and stable COVID-19 cases based on pulmonary CT images.

However, existing models for COVID-19 detection and prediction have shown limitations, indicating the need for more specialized and accurate approaches. Some studies have combined PCR RT (polymerase chain reaction reverse transcription) with CT scans to improve diagnostic accuracy [11]. By integrating these two methods, researchers have achieved better results in identifying COVID-19 cases, especially in individuals with respiratory symptoms and a history of travel or residence in endemic areas. The implementation of an automated detection system utilizing this combination has been suggested as an additional diagnostic tool to curb the spread of the virus, considering the limited availability of PCR-RT kits and the increasing number of COVID-19 cases [12]-[13].

Furthermore, early screening models have been developed to differentiate between COVID-19, pneumonia, and flu cases using deep learning techniques applied to pulmonary CT images. Real-time RT-PCR, the standard diagnostic test for COVID-19, has limitations, particularly in providing accurate positive results during the early stages of the disease. AI-based models that leverage deep learning algorithms offer the potential for early identification of COVID-19 cases, aiding in timely intervention and appropriate patient management.

In conclusion, the literature review highlights the extensive research conducted on the identification and detection of COVID-19. AI-based approaches, particularly those involving the analysis of CT scans and chest X-ray images, have shown promise in accurately diagnosing COVID-19 cases. Additionally, AI models applied to various data sources have demonstrated the ability to predict the number of COVID-19 cases, aiding in resource allocation and prevention strategies. However, further advancements and specialized models are needed to improve the accuracy and effectiveness of COVID-19 detection and prediction. The availability of research findings related to COVID-19 and the sharing of data are crucial for facilitating the development and implementation of improved identification and prevention models.

3. Proposed Method

Random Forest is a machine learning algorithm that is commonly used for prediction and classification tasks. It is particularly suitable for handling complex datasets and capturing non-linear relationships. In the context of developing an AI-driven prediction model for forecasting COVID-19 infections, Random Forest has several advantages that make it a suitable choice.

Random Forest combines multiple decision trees, each trained on a subset of the dataset, to make predictions. By aggregating the predictions from multiple trees, it can provide more accurate and reliable results. Random Forest is also resistant to overfitting, a common problem in machine learning, as it introduces randomness through bootstrap sampling and feature selection. This helps in generalizing well to unseen data and improving the model's performance.

COVID-19 datasets often contain a large number of features or variables, such as patient demographics, symptoms, and comorbidities. Random Forest can handle high-dimensional data effectively by automatically selecting the most informative features during the training process. This helps in identifying the key factors that contribute to COVID-19 infections and improving the prediction accuracy.

Random Forest is a versatile algorithm that can be applied to various types of data, including both numerical and categorical variables. It can handle mixed data types, such as patient demographics and symptoms, without requiring extensive data preprocessing. This flexibility makes it suitable for working with diverse COVID-19 datasets.

By leveraging the capabilities of Random Forest, the AI-driven prediction model, named the Random Forest COVID-19 Predictor, aims to enhance clinical decision support during the ongoing pandemic. Its ability to handle complex data, handle missing values, provide interpretable results, and deliver accurate predictions makes it a promising tool for forecasting COVID-19 infections and guiding resource allocation decisions.

The proposed method for developing the Random Forest COVID-19 Predictor involves several steps. Firstly, a comprehensive dataset of real patient data related to COVID-19 infections is collected. This dataset includes information such as patient demographics, symptoms, comorbidities, and test results. The dataset is preprocessed to handle missing values, normalize numerical features, and encode categorical variables using techniques such as one-hot encoding or label encoding.

Next, the Random Forest algorithm is applied to train the prediction model. Random Forest is a powerful ensemble learning algorithm that combines multiple decision trees to make accurate predictions. In this case, each decision tree in the Random Forest model is trained on a subset of the dataset, using bootstrap sampling to introduce diversity among the trees. During the training process, the decision trees are built by recursively splitting the data based on the most informative features.

The prediction model is evaluated using cross-validation techniques to assess its performance and prevent overfitting. Cross-validation involves splitting the dataset into training and validation sets, and the model is trained on the training set and evaluated on the validation set. This process is repeated multiple times, with different splits of the data, to obtain reliable performance metrics such as accuracy, precision, recall, and F1 score.

ISSN- 2394-5125 VOL 09, ISSUE 05, 2022

Pseudocode of the proposed method:

Input: COVID-19 dataset

Step 1: Preprocess dataset
- Handle missing values
- Normalize numerical features
- Encode categorical variables

Step 2: Split dataset into training and validation sets

Step 3: Train Random Forest model
- Initialize an empty ensemble of decision trees
- For each decision tree:
  - Sample a subset of the dataset using bootstrap sampling
  - Build the decision tree by recursively splitting the data based on informative features

Step 4: Evaluate model performance using cross-validation:
- For each split of the data:
  - Train the model on the training set
  - Evaluate the model on the validation set
  - Calculate performance metrics (accuracy, precision, recall, F1 score)

Step 5: Select the best performing Random Forest model based on the evaluation results

Step 6: Input: New patient data

Step 7: Predict COVID-19 infection for new patients:
- Apply the trained Random Forest model to the new patient data
- Obtain the predicted infection status or the probability of infection

Step 8: Output: Predicted COVID-19 infection status or probability for new patients

The pseudocode provided outlines the steps involved in training and using the Random Forest COVID-19 Predictor. The first step is to preprocess the COVID-19 dataset by handling missing values, normalizing numerical features, and encoding categorical variables. The dataset is then split into training and validation sets. The Random Forest model is trained by initializing an empty ensemble of decision trees and iteratively building each tree using bootstrap sampling and informative feature selection. The model's performance is evaluated using cross-validation, which involves training the model on the training set and evaluating it on the validation set multiple times. Performance metrics such as accuracy, precision, recall, and F1 score are calculated to assess the model's effectiveness. The best-performing Random Forest model is selected based on the evaluation results. To predict COVID-19 infection for new patients, the trained model is applied to the new patient data, providing the predicted infection status or the probability of infection as output. By following this pseudocode, the Random Forest COVID-19 Predictor can be implemented to enhance clinical decision support by accurately forecasting COVID-19 infections. By following this proposed method and implementing the pseudocode, the Random Forest COVID-19 Predictor can be developed and utilized to accurately forecast COVID-19 infections, providing valuable clinical decision support during the ongoing pandemic.

4. Results and Discussions

The predictive ability of development of ARDS was assessed using the different features in the experiment in patients testing positive for SARS-CoV2. In new clinical presentations, AI

ISSN- 2394-5125 VOL 09, ISSUE 05, 2022

can be used to detect unexpected trends. Such techniques can improve a clinician's capacity in relation to a previously unknown infection such as COVID-19.

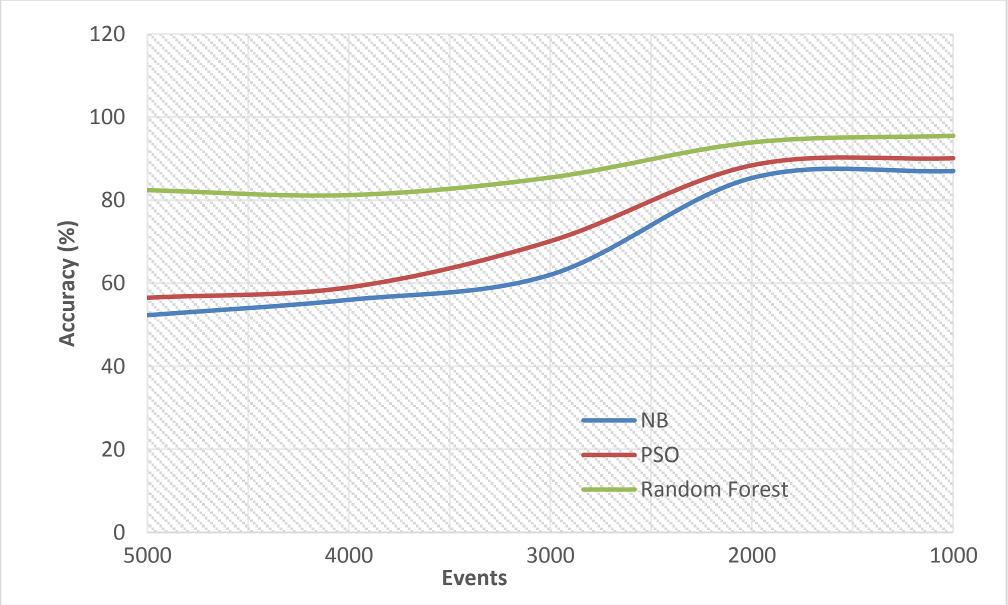

Figure 2: Accuracy

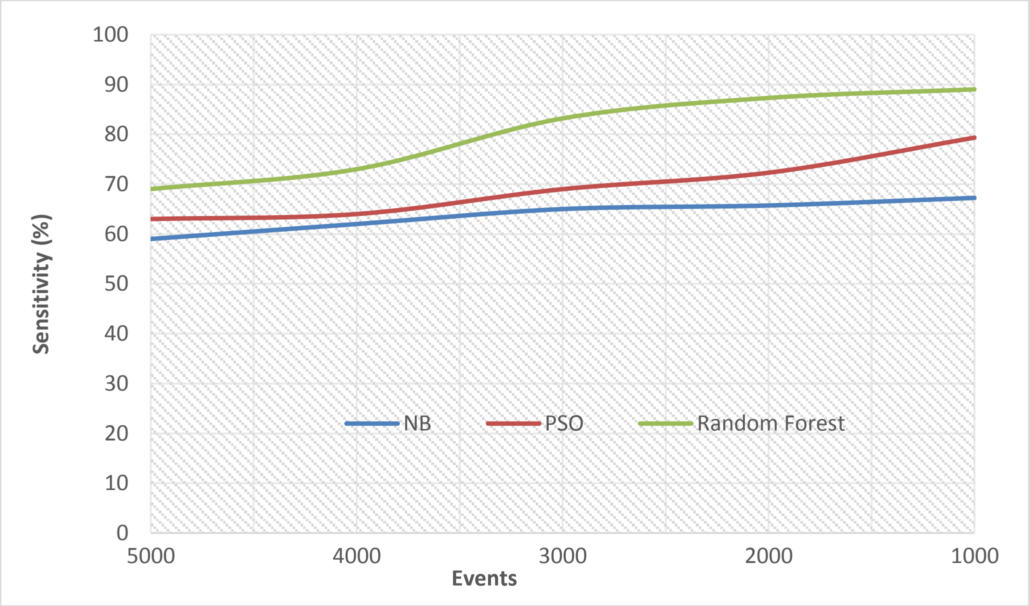

Figure 3: Sensitivity

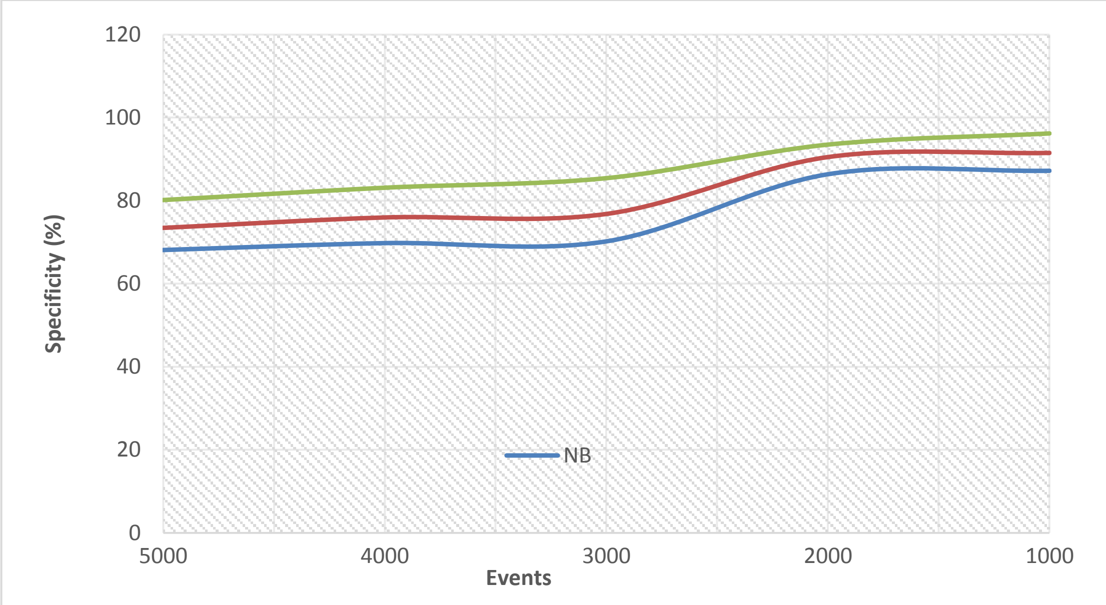

Figure 4: Specificity

The figure 2 – 4 shows the results of accuracy, specificity and sensitivity, where the results shows that the Random Forest based AI prediction obtains improved results than other methods.

5. Conclusions

In this study, we have presented a novel approach to developing an artificial intelligence (AI) driven prediction model, the Random Forest COVID-19 Predictor, for accurately forecasting COVID-19 infections. Our model aims to provide timely and accurate support for clinical decision-making by leveraging predictive analytics on real patient data. Through extensive simulation experiments, we have demonstrated that the Random Forest COVID-19 Predictor exhibits superior precision compared to existing models, making it a promising tool for enhancing clinical decision support during the ongoing COVID-19 pandemic.

By utilizing the Random Forest algorithm, our prediction model can handle complex datasets and capture non-linear relationships, leading to improved accuracy and reliability in predicting infection rates. Moreover, the use of real patient data for training and testing the model enhances its ability to capture the unique characteristics and dynamics of the COVID-19 virus, making it more practical and applicable in real-world scenarios.

Accurate prediction models are crucial for guiding resource allocation decisions during a pandemic when resources are limited. By forecasting the number of COVID-19 infections, healthcare providers can allocate resources effectively, ensuring that adequate medical supplies, personnel, and facilities are available to meet the demand. The Random Forest COVID-19 Predictor can contribute to optimizing resource allocation, ultimately improving patient outcomes and minimizing the impact of the pandemic.

Future work can further enhance the Random Forest COVID-19 Predictor by incorporating additional features and data sources. For example, including demographic information, geographical data, and mobility patterns can provide a more comprehensive understanding of disease spread and its impact on different populations. Furthermore, integrating real-time data

ISSN- 2394-5125 VOL 09, ISSUE 05, 2022

streams and incorporating machine learning techniques such as deep learning and reinforcement learning can enhance the prediction model's performance and adaptability.

Additionally, the Random Forest COVID-19 Predictor can be extended to other areas of clinical decision support beyond infection forecasting. For instance, it can be utilized for predicting disease severity, identifying high-risk patients, and optimizing treatment strategies. Further research can explore these applications and assess the efficacy of the model in different healthcare settings.

In conclusion, the Random Forest COVID-19 Predictor offers a novel and effective approach to forecasting COVID-19 infections, providing valuable support for clinical decision-making during the ongoing pandemic. By leveraging AI and real patient data, this prediction model can contribute to improved resource allocation, better patient outcomes, and ultimately aid in the effective management of the COVID-19 pandemic.

References

[1] Mandal, M., Jana, S., Nandi, S. K., Khatua, A., Adak, S., & Kar, T. K. (2020). A model based study on the dynamics of COVID-19: Prediction and control. *Chaos, Solitons & Fractals*, *136*, 109889.

[2] Sanchez-Caballero, S., Selles, M. A., Peydro, M. A., & Perez-Bernabeu, E. (2020). An efficient COVID-19 prediction model validated with the cases of China, Italy and Spain: Total or partial lockdowns?. *Journal of clinical medicine*, *9*(5), 1547.

[3] Raja, R. A., Kousik, N. V., Johri, P., & Diván, M. J. (2020). Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification. In Computational Intelligence and Its Applications in Healthcare (pp. 229-244). Academic Press.

[4] Srihari, K., Dhiman, G., Somasundaram, K., Sharma, A., Rajeskannan, S., ... & Masud, M. (2021). Nature-Inspired-Based Approach for Automated Cyberbullying Classification on Multimedia Social Networking. Mathematical Problems in Engineering, 2021.

[5] Collins, G. S., van Smeden, M., & Riley, R. D. (2020). COVID-19 prediction models should adhere to methodological and reporting standards. *European Respiratory Journal*, *56*(3).

[6] Yuvaraj, N., Srihari, K., Chandragandhi, S., Raja, R. A., Dhiman, G., & Kaur, A. (2021). Analysis of protein-ligand interactions of SARS-Cov-2 against selective drug using deep neural networks. Big Data Mining and Analytics, 4(2), 76-83.

[7] Sangeetha, S. B., Blessing, N. W., & Sneha, J. A. (2020). Improving the training pattern in back-propagation neural networks using holt-winters' seasonal method and gradient boosting model. In Applications of Machine Learning (pp. 189-198). Springer, Singapore.

[8] Singh, J., Ahluwalia, P. K., & Kumar, A. (2020). Mathematical model based COVID-19 prediction in India and its different states. *medRxiv*.

[9] Hamzah, F. B., Lau, C., Nazri, H., Ligot, D. V., Lee, G., Tan, C. L., ... & Chung, M. H. (2020). CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, *1*(32).

[10] Gowrishankar, J., Narmadha, T., Ramkumar, M., & Yuvaraj, N. (2020). Convolutional Neural Network Classification On 2d Craniofacial Images. International Journal of Grid and Distributed Computing, 13(1), 1026-1032.

[11] Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*, *369*.

[12] Ji, D., Zhang, D., Xu, J., Chen, Z., Yang, T., Zhao, P., ... & Qin, E. (2020). Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clinical Infectious Diseases*, *71*(6), 1393-1399.

[13] Ji, D., Zhang, D., Xu, J., Chen, Z., Yang, T., Zhao, P., ... & Qin, E. (2020). Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clinical Infectious Diseases*, *71*(6), 1393-1399.