COVID OMICRON PRIDCATION USING RANDOM FOREST CLASSIFICATION ALGORITHM

Mrs.V.Ramya, Research scholar
Department of Computer Science
VISTAS Pallavaram, Chennai, India.
ramyars89@gmail.com.
Dr.T.S.Thirunavukkarasu, Assistant Professor
Department of Computer Science
VISTAS Pallavaram, Chennai, India.
thirukst@gmail.com.

Abstract:

COVID-19 is a contagious disease and the most extensive health disaster, with daily increases in infected people and fatalities. Coronavirus variants such as Omicron and Delta are spreading rapidly, which are more dangerous for people of all ages. Due to its prevalence in humans, the prediction and discovery of the Omicron variation of COVID-19 raised significant concerns among the medical community.The frequency and severity of chronic diseases are rising, and population aging results in an increased risk of COVID-19 related to hospitalization or mortality. Finding potentially high-risk patients is complex, and the diagnostic procedures for this COVID-19 variation have become more intricate. Researchers and experts are involved in establishing quicker and less expensive diagnostic approaches. Due to the high population growth and spread of disease, automatic disease diagnosis and prediction have become essential topics in the medical industry. The machine learning (ML) technique, such as an improved random forest model, is proposed to predict COVID-19 and its variants in this work.It is a two-phase process; the bootstrap resampling method is employed for obtaining the original samples and applying a decision tree voting approach for prediction. The obtained accuracy by the proposed RF model is compared with CNN [12] and ECNN-ERNN [13]. The proposed RF model archives 98.8% of accuracy, which defines its enhanced efficiency more than the others.

Keywords: COVID-19, Data Mining, Omicron, Machine Learning, Classification, Prediction

Introduction:

Coronavirus Disease (COVID-19) is a dangerous disease that was first identified in Wuhan, China, in December 2019 [1]. The Coronavirus Disease (COVID-19) virus is a member of a long-lived family of viruses that can cause an outbreak or pandemic [2,3]. A coronavirus (COVID-19) is a respiratory illness with common symptoms like stuffy nose, dry cough, tiredness, fever,sore throat, headache, loss of taste and smell, aching and pains, etc. It also leads to a disease called parosmia, in which once-pleasant smells may now be intolerable. The olfactory loss brought on by the virus is one of the causes of parosmia symptoms. Pneumonia brought on by COVID-19 can harm the lungs [4].

In the most extreme situations of COVID-19, the lungs are affected, which can cause Pneumonia. Radiographic examination of the lungs using techniques such as a CT scan, chest X-ray (CX-Ray), ultrasound, and others can be performed to determine whether the lung is infected by bacterial, viral, or COVID-19 Pneumonia [5,6]. The haziness patterns like linear opacities and ground-glass opacities (GGO) can be detected in a lung affected by COVID-19 [7]. A battery of reverse transcriptase-polymerase chain reaction examinations can also be used in traditional laboratory methods to diagnose COVID-19 (RT-PCR).The RT-PCR requires more time and occasionally results in false negatives [8]. However, using radiographic pictures of the lungs, COVID-19 disease can be effectively recognized. This might result in the sick person being quarantined before it spreads further.

SARS-CoV-2 is the virus that produces the COVID-19 epidemic sickness (severe acute respiratory syndrome coronavirus-2). SARS (severe acute respiratory syndrome) and Middle East Respiratory Syndrome (MERS-CoV) are two diseases caused by the huge virus family known as coronaviruses (CoV) (SARS-CoV). The World Health Organization recognized B.1.1.529, a novel SARS-CoV-2 variant, as a kind of concern and named Omicron [9].The most harmful COVID-19 variations are Omicron and Delta, which was prevalent in Europe and spread to western Asia, where they caused a record-breaking rise in incidence and fatality rates. The Omicron variation takes over as the predominant variety globally in January 2022 because it spreads twice as quickly as the Alpha variant and is numerous times more highly dangerous than the preceding versions, according to the CDC. If it is not identified promptly, it can damage the entire lung, at which point patients may experience hypoxia brought on by pulmonary fibrosis, which can be fatal [10,11].

Several COVID-19 diagnostic kits now available in hospitals is noticeably less than what is required to handle the rising number of cases. Additionally, because of ignorance and fear, people are forced to do tests to determine if the results are positive or negative. In order to employ analytical kits successfully and stop the disease's development by providing early treatment, it is necessary to develop an automatic prediction system. Use of machine learning (ML) is a very effective weapon in the war against COVID-19. It may be used to successfully predict the progress of the disease and manage enormous amounts of data. Both diagnosis and COVID-19 prediction are aided by it. ML techniques are helpful for various purposes, including tracking COVID instances, forecasting, developing dashboards, diagnosing patients and prescribing the right drugs, providing notifications to improve social distance, and other possible control mechanisms of viral propagation.

The problem statement discussed in this paper is the need for a reliable COVID-19 pandemic prediction mechanism. The virus's widespread transmission capacity is potential harm. The number of COVID-19-related data is always growing and changing, which makes it challenging to find the best solutions. As previously stated, no effective or advanced methods for detecting omicron variations exist. The time requirements of the current techniques need to be lowered to achieve accurate predictions.

This paper addresses the abovementioned problems by implementing the proposed advanced Random Forest Model.The proposed model effectively delivers accurate prediction results by removing the abnormal and noise values.

The rest of this paper is organized as follows: Section I describes the introduction. In section II, various works related to COVID-19 prediction are described. Section III explains the proposed design and its workflow in detail. Section IV shows the observations obtained from the experimental work. Section V describes the conclusion of this work.

Related Works:

Mustafa Ghaderzadeh et al. [12] proposed a convolutional neural network (CNN) using transfer learning and parameter optimization to identify the COVID-19 Delta and Omicron Variants. It is a two-stage process; initially, data preparation is processed in which dataset images are collected from the diagnostic center and Kaggle website. The obtained images are in DICOM format, which is the output from the PAC system. Next, each image is replicated three times to create dummy RGB channels after the collected photos have been transformed into a single grey channel space (grayscale). Finally, the proposed mechanism is implemented; as a result, the entire system could distinguish between COVID-19 patients and non-COVID-19 instances from the X-ray and CT Scan images.

Anand Kumar Gupta et al. [13] proposed the combined architecture of Extended Convolutional Neural Networks (ECNN) and Extended Recurrent Neural Networks (ERNN) for predicting omicron viruses. The main motto of the proposed system is to establish an automatic prediction of omicron virus from chest CT Scan images. Initially, the datasets are extracted from the Kaggle public repository and undergo preprocessing. Next, extended CNN techniques such as ECNN and ERNN are applied. Accordingly, the ECNN extracts the deep features from the dataset, and the ERNN is responsible for exploring the extracted features. The experimental results show the prediction efficiency through the values obtained with various validation parameters.

Asifullah Khan et al. [14] analyzed multiple deep learning (DL) approaches for detecting COVID -19. The author conducted an in-depth survey on various deep learning techniques, their learning approaches, and diagnostic strategies in this work. This study includes various categories, and each category is provided with a custom-made Convolutional Neural Network technique for predicting COVID-19 disease from X-Ray, Computer Tomography (CT), and radiographic images. As a result, this study paves the way for developing individualized DL-based diagnostic tools for efficiently addressing COVID-19 variants and new difficulties.

L.J. Muhammad et al. [15] developed a machine-learning approach for detecting positive cases of COVID-19. Deep learning approaches were used in this study for performing classification, regression, and decision-making tasks. The generated categorization models are employed to track the start of human disease—extensively dispersed clinical records, which included COVID-19 patients that were both infected and normal. When compared to naive algorithms, this approach achieves 95% of accuracy.

Li et al. [16] examined five COVID-19 patients aged between ten months and six years with lung CT scans. According to the study, three patients had significant abnormalities, whereas two showed no signs of disease on lung CT scans.

Karami [17] analyzed several COVID-19 research publications between May 5, 2020, and May 5, 2021. They used frequency analysis to determine the most typical symptoms and treatment options, highlighting the importance of the two biomedical concepts. This study also uses topic modeling, which led to 25 categories that show connections between the two main categories.

Alsunaidi et al. [18] state that the big data process is crucial for gaining the knowledge required to decide what to do and how to do it. Given the enormous amount of COVID-19 data available from numerous sources, it is crucial to review the contribution of big data analysis to controlling COVID-19's spread and to outline the main challenges and prospects in COVID-19 data analysis. In this study, the challenges associated with analyzing COVID-19 data are explored. The findings of this study highlight critical areas that require additional investigation and implementation.

Valdiviezo-Diaz [19] implemented data mining models for identifying COVID-19 recovery cases from the china day-level information dataset. Data mining algorithms are used to forecast the recovery of COVID-19 patients. The COVID-19 patient dataset is subjected to the neural network, decision trees, and logistic regression techniques using the R programming language. Experimental results showed that the neural network greatly improves classification accuracy and has a lower error rate when compared to earlier methods.

Narin et al. [20] proposed an automatic COVID-19 detection framework by implementing five pre-trained existing CNN models. They are ResNet-101, ResNet-152, ResNet-50, Inception-V3 and Inception-ResNet-V2. The proposed framework progresses with various binary classifications, such as COVID-19 vs. bacterial Pneumonia, COVID-19 Vs. Normal and COVID-19 vs. viral Pneumonia. Among the five models, ResNet-50 attained maximum accuracy in computing with three datasets extracted from the Kaggle repository, ChestX ray8 database, and open-source GitHub repository.

In work [21], the author reviewed the clinical decision support (CDS) and IA relationship, highlighting the valuable and trustworthy datasets that can be accessed for better COVID-19 solutions. To determine the dataset quality, it is necessary to offer specifics regarding the technical features of the structure of the accessible datasets.

Rapid disease diagnosis and case monitoring are the first and most essential steps in controlling the most recent COVID-19 variations. The Omicron and Delta forms of COVID-19 had rapid illness development and deterioration due to the lengthy turnaround time for PCR findings. In order to overcome this, the Random Forest model is proposed. The main contribution of this work is designing and implementing the improved RF model. The proposed approach can be utilized for identifying high-risk individuals in pandemics like COVID-19. It is very effective in identifying the severe, life-threatening COVID-19 variant Omicron. In the experimental work, the proposed improved RF model is computed with the COVID-19 dataset containing omicron variants obtained from European Centre for Disease Prevention and Control (ECDC) repository. Next, a comparison of various evaluation parameters is carried out with the existing approaches for examining the proposed RF's overall performances.

Proposed System:

An improved Random Forest model has been developed to recognize COVID-19 patients in the initial stages of the disease. The primary goal of the proposed model is to improve the classification accuracy in predicting the COVID-19 Omicronvariants from huge datasets.

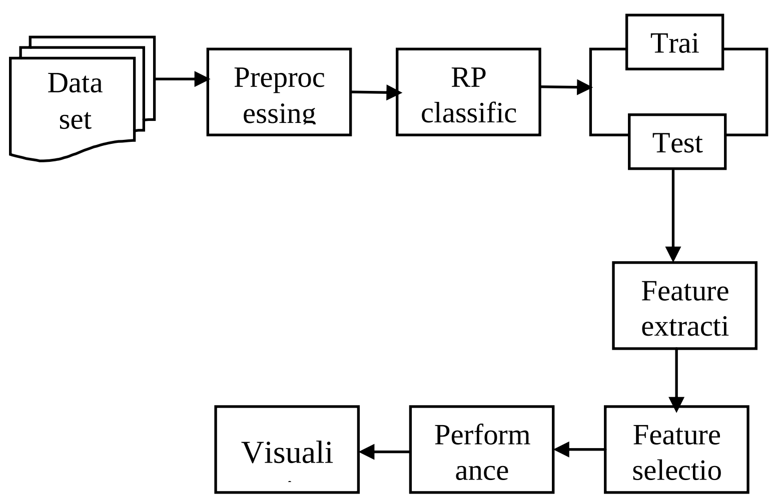

Figure 1 Proposed architecture

The proposed improved Random Forest model enhances the random forest (RF) algorithm. In 2001, Breiman proposed the RF algorithm as a combination method that includes random forest classification, regression, and continuous value. Most of the research demonstrated that the outcomes of the random forest algorithm had greater accuracy predictions that effectively avoided the noise and aberrant values. The proposed RF classification model is a two-phase work in which, at the initial stagebootstrap resampling approach is applied to obtain the subsamples from the original samples. Next, a decision tree is created for each sample based on which the classified decision tree enables the voting system. The decision tree that receives the most votes is used as the outcome of the final prediction.The proposed RF model's working process is defined below;

(1) Select the training set: The K training sets are retrieved from the original dataset using the bootstrap random sampling approach (M properties). Much like the original training set, the training set is utilized with its size.

(2) Build the RF model: The bootstrap training setscategorization regression trees to produce the K decision trees build as a "forest"with the trees in the forest remaining unchanged.Due to internal nodes, no tree form resembles the best qualities. In this case, branching is accomplished by randomly selecting all features m⩽Mof.

(3) Create simple voting: The training procedure for the decision tree is independent of one another. Random forest training helps to increase its effectiveness.The same combination of trained K decision trees is used to build the RF. Once the input samples have been classified, each decision tree is accomplished with a simple voting system for determining the output.The RF algorithm determines the final classification of each decision tree, independent sample determination, and distributed decision trees.

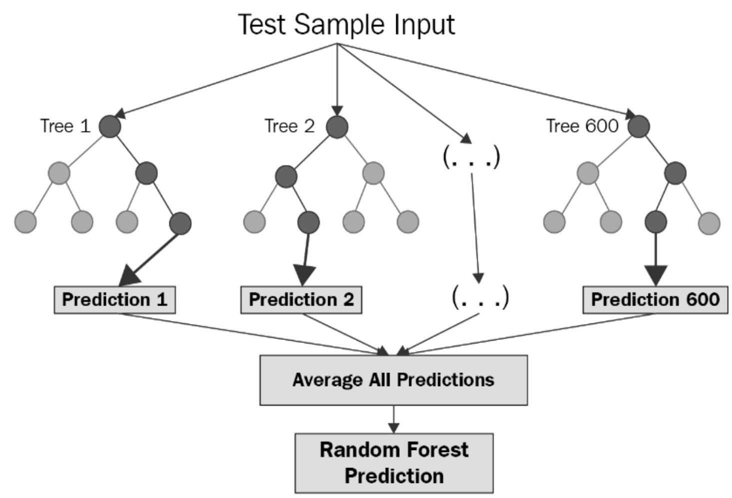

Fig-2 random forest structure

The proposed RF model has a two-stage functioning system. The N decision tree and the random forest are combined in the first stage. The second step addresses predictions for each tree constructed in the first phase.

Figure 2 illustrates the structure of Random Forest, and its working process is defined as below;

Step-1: From the training dataset, Random K data points are chosen

Step 2: The associatedselected data points, referred to as subsets, are used for building the decision trees.

Step 3: N is chosen as the required number to build decision trees.

Step 4: Repeat Steps 1 & 2.

Step 5: Next, a tree prediction is made for each decision. A standard voting method is used to classify the new data points; the data point receiving the most votes is considered the most accurate. The simple example below explains how the RF algorithm is functioning;

Example: Consider a dataset with a collection of fruit photos fed into a Random Forest classifier. The dataset is initially divided into subgroups, and many decision trees are built. Each decision tree has a prediction output when a new data point appears during the training phase.The Random Forest model's final prediction is based on the data point receiving the most votes.

Data preprocessing

Data preprocessing involves transforming unusable input into knowledge that a machine learning algorithm can utilize.In order to create a machine-learning model, data preprocessing is the first and most crucial step. While working on a machine learning model, clean accessand properly prepared data are mandatory.Before using data for any purpose, it must constantly be cleansed and formatted.Therefore, data preprocessing is significant in machine learning.

Data set

The proposed model is executed in this work with the dataset extracted from the European Centre for Disease Prevention and Control (ECDC) repository. Since 2005, ECDC has been the European Union (EU) public health organization that collects more than 500infectious datasets, most of which belong to germs and parasites. ECDS collects, examines, and shares more than 50 infectious disease-related issues, such asCOVID-19 omicron, influenza, TB, hepatitis, antibiotic resistance, measles, HIV/AIDS, and vaccinations. ECDCProfessionalsassess the threats to Europe and provide suggestions to help countries stop outbreaks and address public health risks. Beginning on March 11, 2021, the ECDC started reporting the data about any new cases or fatalities of COVID-19 omicron in any EU/EEA nations. Worldwide, on each Thursday, they provide the reports in various formats such as JSON, CSV, XLSX, and XML. From there, data on COVID-19 omicron instances can be obtained weekly and daily.

Tools used

The Classification Learner software is employed for training the data classification models. It discovers supervised machine learning using various classification models. It enables performance data analysis, feature selection, outcome evaluation, stated validation processes, and model training. The ML algorithms such as naive Bayes, neural network classification, Support vector machines, decision trees, kernel approximation, nearest neighbors, ensemble, logistic regression, and discriminant analysis are trained automatically for determining the best classification model type. The supervised machine learning outcomes are expressed as labels or classes, described by providing defined input data and known data responses. Next, the data is used to train the predicting model. The trained model can now be exported to the workspace or used with new data by creating MATLAB code, or it can be used to explore programmatic categorization further.

Experimental results:

The proposed approach experimental work is carried out with the COVID-19 dataset. The dataset is extracted from the ECDC repository, including numerous COVID-19-related datasets. The ECDC repository and its dataset are widely known for research executions. In this section, the experimental work is conducted with a normal and ICU-admitted cases dataset. These data are obtained daily and weekly around European countries. The main contribution is obtaining the daily and weekly admitted cases and classification of positive omicron variant cases. MATLAB is a numeric computing and simulation platform used to describe the obtained results in Precision, F-measure, and Recall.

=== Summary ===

Correlation coefficient	1
Mean absolute error	0.0085
Root mean squared error	0.055
Relative absolute error	0.0871 %
Root relative squared error	0.3103 %
Total Number of Instances	552

Table 1: Evaluation factors vs. Obtained Values

Table 1 shows the summary of the evaluation factors obtained during the execution. The proposed Random Forest model is executed with the COVID dataset in which the total number of instances is 552 with a relative absolute error of 0.0871%. The other considered parameters are shown in the table with their obtained values.

Scatter plot:

A scatter plot is a mathematical representation that applies Cartesian coordinates for the resulting two variables. The scatter plots are progressed from the data collection and are represented in plots or figures. The representation contains the horizontal and vertical axes, in which the horizontal axis denotes the value of one variable. In contrast, the vertical axis denotes the value of the other variable. The obtained points, such as shape, color, or size, are denoted as plots.

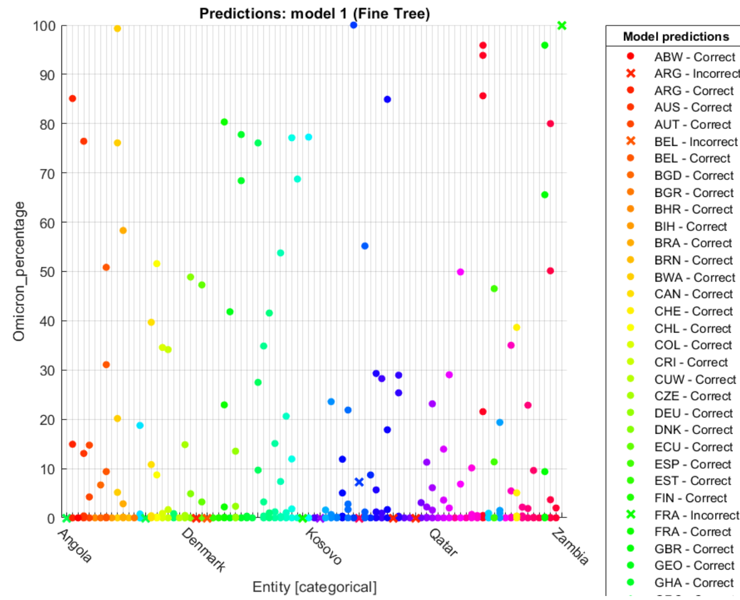

Figure 3 scatter plot data.

Figure 3 represents the omicron prediction model concerning the dataset. The dataset contains COVID-19 records from Angola, Denmark, Kosovo, Qatar, and Zambia. The countries are represented in the x-axis and the y -axis represents the classified omicron percentage concerning the countries. The plots are mapped with Omicron-affected cases for the cities of the defined countries.

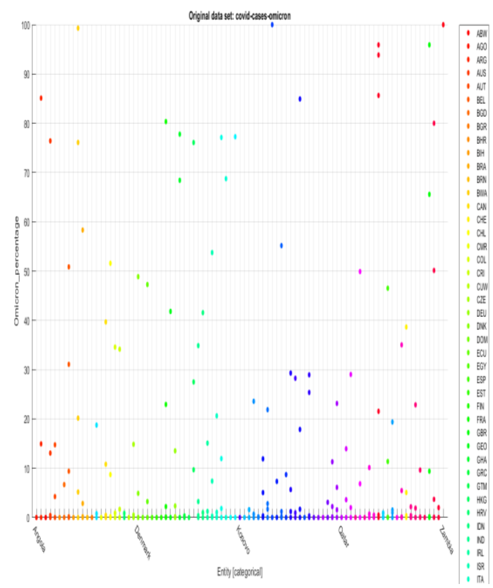

Figure 4 scatter plot prediction.

Figure 4 represents the original omicron dataset concerning the Omicron affected percentage. The dataset contains COVID-19 records from Angola, Denmark, Kosovo, Qatar, and Zambia. The countries are represented in the x-axis and the y -axis represents the classified omicron percentage. The plots are mapped with Omicron-affected cases for the cities of the defined countries. In comparison to figure 1 obtained prediction tree result with the original prediction, our proposed Random Forest model has achieved about 98.8 % accuracy, such as similarities.

ROC

Machine learning is the best technique for determining the classification, but it is more important to guarantee whether the applied ML model performance is sufficient. Once the ML model is built, the procedure evaluates the model in the aspect of various evaluation parameters. Several evaluations are there, but the most important metric needed for fulfilling the performance of the classification model is the AUC-ROC curve. The AUC-ROC curve is the most popular metric in defining classification accuracy, and the upcoming results show obtained AUC-ROC curve with the proposed model.

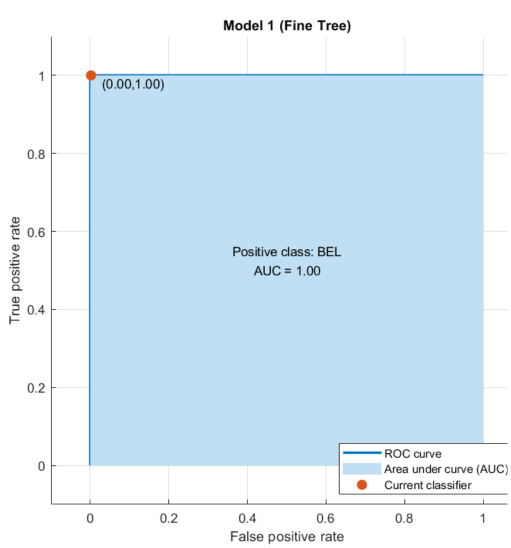

Figure 5 ROC for Belgium

Figure 5 represents the obtained ROC for Belgium, in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Belgium is 1.00.

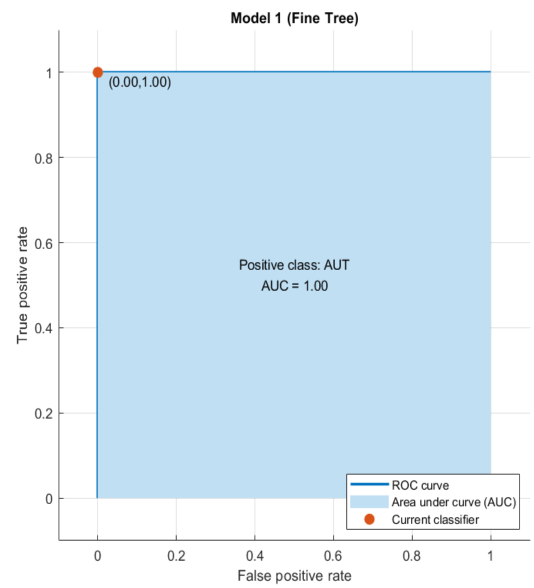

Figure 6 ROC for Australia

Figure 6 represents the obtained ROC for Australia in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Australia is 1.00.

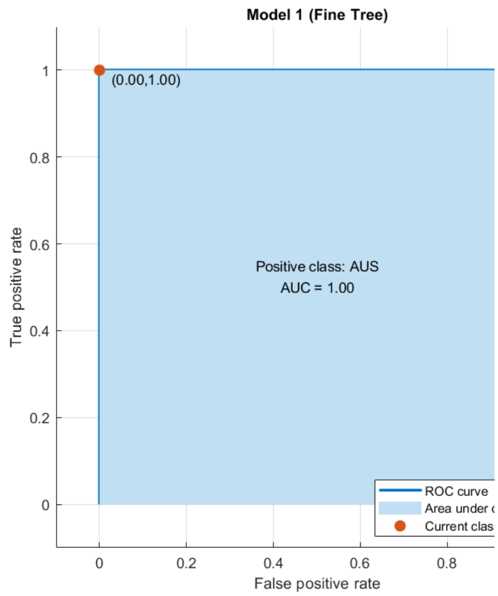

Figure 7 Austria

Figure 7 represents the obtained ROC for Austria, in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Austria is 1.00.

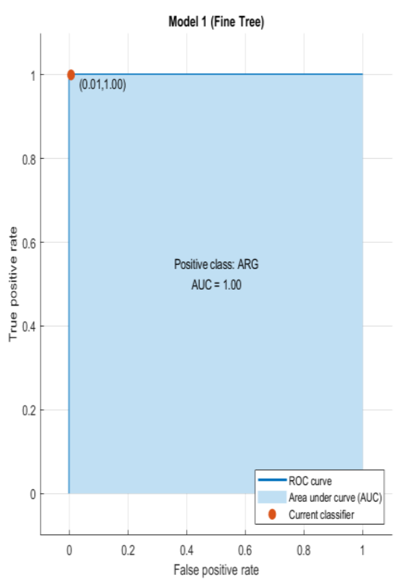

Figure 8 ROC for Argentina

Figure 8 represents the obtained ROC for Argentina, in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Argentina is 1.00.

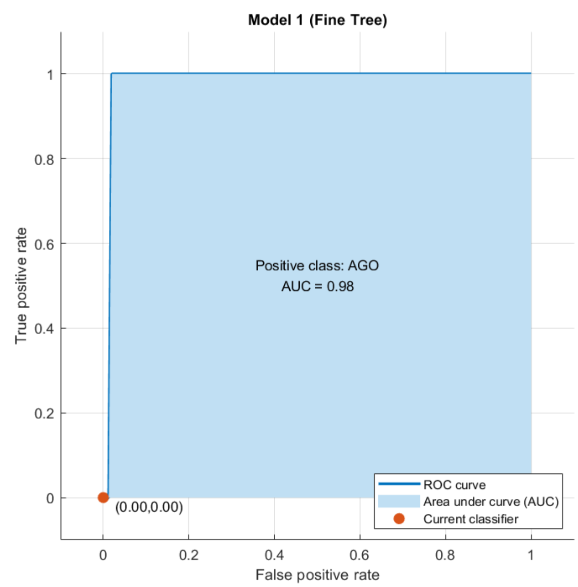

Figure 9 ROC for Angola

Figure 9 represents the obtained ROC for Angola, in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Angola is 0.98.

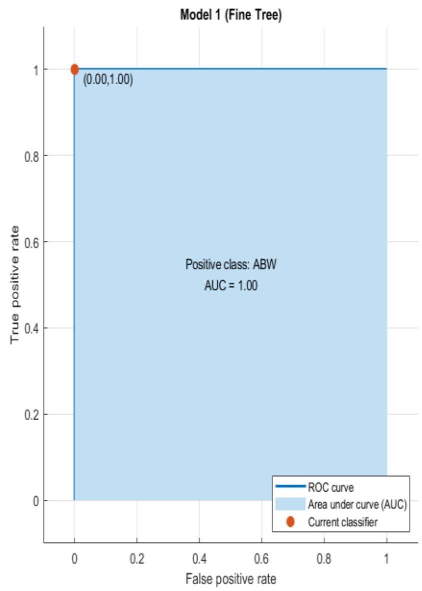

Figure 10 ROC for Aruba

Figure 10 represents the obtained ROC for Aruba in which the x-axis shows the false positive rate and the y-axis shows the true positive rate. The area under the Curve (AUC) or Receiver operating characteristic (ROC) is the indicator for measuring the classification performance. High AUC defines the maximum accuracy in distinguishing omicron cases and normal cases. From the dataset, the AUC achieved for patients from Aruba is 1.00.

Comparison Result:

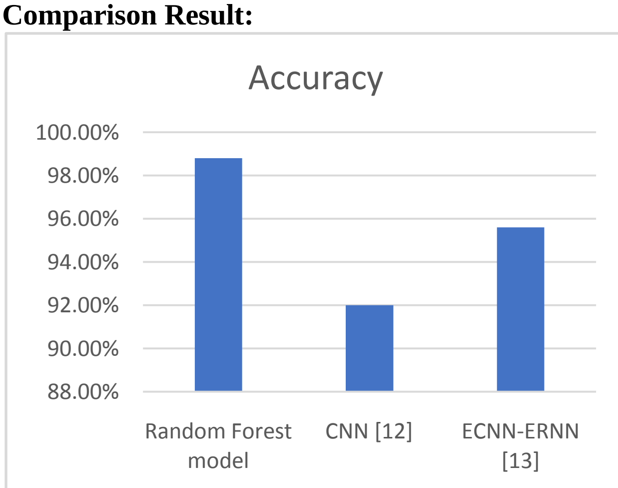

Figure 11 Accuracy comparison level

Algorithms	Accuracy in %
Random Forest model	98.8%
CNN [12]	92%
ECNN-ERNN [13]	95.6%

Table 2: Comparison Table

As discussed above table 2 and figure 11, the classification performance of the proposed Random Forest (RF) model with evaluation metrics AUC is defined. In order to determine the performance efficiency of the proposed model in defining the omicron variants a comparison work is conducted with the existing convolutional neural network (CNN) [12] and Extended Convolutional Neural Networks and Extended Recurrent Neural Networks (ECNN - ERNN) [13]. The comparison is based on the obtained accuracy in which the proposed RF model achieves 98.8% accuracy, whereas CNN achieves 92% and ECNN – ERNN with 95.6%. The comparison result proves that the accuracy obtained by the proposed RF model is far better than the others.

Conclusion:

In order to give the public a clear strategy for combating the COVID-19 pandemic, which is spreading globally, an epidemiological examination must be conducted in real-time. Because the Omicron COVID-19 variants can harm the entire lung in a few days, it is essential to develop faster and more accurate diagnostic tools with the aid of medical professionals and healthcare systems.Day by day, the total number of affected cases are at its peak, and additionally, the multiple symptoms with various source make omicron prediction complex. The existing approaches need enhancement because the patient data volume and disease sources vary from country to country. Hence, an advanced random forest model is proposed, derived from the traditional random forest (RF) algorithm. The proposed model uses the updated COVID dataset to classify the Omicron-affected cases. The level of accuracy is defined by scatter point representation and ROC curve. The proposed accuracy is compared with the existing CNN and ECNN – ERNN. The proposed RF model is more significant in achieving accuracy than the others.

Reference:

1) Albarello, F., Pianura, E., di Stefano, F., Cristofaro, M., Petrone, A., Marchioni, L., Palazzolo, C., Schininà, V., Nicastri, E., Petrosillo, N., Campioni, P., Eskild, P., Zumla, A., Ippolito, G., Abbonizio, M. A., Agrati, C., Albarello, F., Amadei, G., Amendola, A., Valli, M. B. (2020). 2019- novel Coronavirus severe adult respiratory distress syndrome in two cases in Italy: An uncommon radiological presentation. International Journal of Infectious Diseases, 93, 192–197. https://doi.org/10.1016/j.ijid.2020.02.043
2) C. Orbann, L. Sattenspiel, E. Miller, and J. Dimka, “Defining epidemics in computer simulation models: How do definitions influence conclusions?,” Epidemics, vol. 19, pp. 24–32, Jun. 2017, doi: 10.1016/j.epidem.2016.12.001.
3) J. Rasheed et al., “A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic,” Chaos, Solitons & Fractals, vol. 141, p. 110337, Dec. 2020, doi: 10.1016/j.chaos.2020.110337
4) K. Nishitha , C. Shiny Gracy , B. Priyadharshini , M. Sowmiya , N. KadharBasha, 2021, COVID Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 05 (May 2021), DOI : 10.17577/IJERTV10IS050205
5) T. Ai et al., “Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases,” Radiology, p. 200642, Feb. 2020, doi: 10.1148/radiol.2020200642.
6) I. Katsamenis, E. Protopapadakis, A. Voulodimos, A. Doulamis, and N. Doulamis, “Transfer Learning for COVID-19 Pneumonia Detection and Classification in Chest X-ray Images,” in 24th Pan-Hellenic Conference on Informatics, Nov. 2020, pp. 170–174, doi: 10.1145/3437120.3437300.
7) P. Sachet et al., “Management of orofacial lesions with antimicrobial photodynamic therapy and photobiomodulation protocols in patients with COVID-19: A multicenter case series,” PhotodiagnosisPhotodyn. Ther., vol. 38, p. 102743, Jun. 2022, doi: 10.1016/j.pdpdt.2022.102743.
8) F. Shi et al., “Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19,” IEEE Reviews in Biomedical Engineering, vol. 14. pp. 4–15, 2021, doi: 10.1109/RBME.2020.2987975.
9) Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern, (n.d.), 2022, https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2- variant-of-concern.
10) D. Loconsole, F. Centrone, C. Morcavallo et al., “Changing features of COVID-19: characteristics of infections with the SARS-CoV-2 Delta (B. 1.617.2) and Alpha (B. 1.1.7) variants in

Southern Italy," Vaccines, vol. 9, no. 11, p. 1354, 2021.

11) D. A. Siegel, H. E. Reses, A. J. Cool et al., "Trends in COVID-19 cases, emergency department visits, and hospital admissions among children and adolescents aged 0-17 yearsUnited States, August 2020-August 2021," MMWR. Morbidity and Mortality Weekly Report, vol. 70, pp. 1249–1254, 2021.
12) Mustafa Ghaderzadeh , Mohammad Amir Eshraghi , FarkhondehAsadi , Azamossadat Hosseini , RamezanJafari , DavoodBashash , and Hassan Abolghasemi, "Efficient Framework for Detection of COVID-19 Omicron and Delta Variants Based on Two Intelligent Phases of CNN Models", Hindawi Computational and Mathematical Methods in Medicine Volume 2022, Article ID 4838009, 10 pages https://doi.org/10.1155/2022/4838009
13) Anand Kumar Gupta , AsadiSrinivasulu,KamalKantHiran,GoddindlaSreenivasulu,SivaramRajeyyagari , and MadhusudhanaSubramanyam, "Prediction of Omicron Virus Using Combined Extended Convolutional and Recurrent Neural Networks Technique on CT-Scan Images", Hindawi Interdisciplinary Perspectives on Infectious Diseases Volume 2022, Article ID 1525615, 11 pages https://doi.org/10.1155/2022/1525615.
14) Khan, A., Khan, S. H., Saif, M., Batool, A., Sohail, A., & Khan, M. W. (2022). A Survey of Deep Learning Techniques for the Analysis of COVID19 and their usability for Detecting Omicron. arXiv preprint arXiv:2202.0637
15) L. J. Muhammad, Ebrahem A. Algehyne, Sani Sharif Usman, Abdulkadir Ahmad, Chinmay Chakraborty and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset", Computer Science, 2020.
16) Li, M., Lei, P., Zeng, B., Li, Z., Yu, P., Fan, B., Wang, C., Li, Z., Zhou, J., Hu, S., & Liu, H. (2020). Coronavirus Disease (COVID-19): Spectrum of CT Findings and Temporal Progression of the Disease. Academic Radiology, 27(5), 603–608. https://doi.org/10.1016/j.acra.2020.03.003
17) Karami, A., Bookstaver, B., Nolan, M., &Bozorgi, P. (2021). Investigating diseases and chemicals in COVID-19 literature with text mining. International Journal of Information Management Data Insights, 1(2), 100016. https://doi.org/10.1016/j.jjimei.2021.100016
18) Alsunaidi, S. J., Almuhaideb, A. M., Ibrahim, N. M., Shaikh, F. S., Alqudaihi, K. S., Alhaidari, F. A., Khan, I. U., Aslam, N., &Alshahrani, M. S. (2021). Applications of Big Data Analytics to Control COVID-19 Pandemic. Sensors, 21(7), 2282. https://doi.org/10.3390/s21072282
19) Valdiviezo-Diaz, P. (2021). Data Mining to Predict COVID-19 Patients' Recovery on a Balanced Dataset. Communications in Computer and Information Science, 340–350. https://doi.org/10.1007/978-3-030-71503- 8_26
20) A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," Pattern Anal. Appl., vol. 24, no. 3, pp. 1207–1220, Aug. 2021, doi: 10.1007/s10044-021-00984-y.