Performance Evaluation of Ensemble Deep Learning Algorithms for Prediction of Pandemic Disease

Mhd Saeed Sharif
Computer Science and DT, ACE, UEL
University Way, London, UK
S.Sharif@uel.ac.uk

Abin Zorto
Computer Science and DT, ACE, UEL
University Way, London, UK
u2091940@uel.ac.uk

Ibukun David Aluko
Computer Science and DT, ACE, UEL
University Way, London, UK
u2350733@uel.ac.uk

Abstract—For optimal healthcare management and counter-measures, it is essential to monitor and predict severe disease at the right time, before it becomes pandemic. In this research work, the most recent pandemic is considered as an example, as the viral coronavirus disease (COVID-19) prognosis is crucial to learn from. The severe COVID-19 threat has had a substantial influence on the global health security scene, forcing the creation of cutting-edge computer models to improve monitoring, control, and mitigation measures. The research study aims to develop a generalized model assessing the healthcare parameters at a personalized and community dimensions and predicting the severity of the disease before becoming pandemic. To achieve this aim, this paper has systematically evaluated the outcomes of different experiments utilizing the ResNet, DenseNet, and ensemble models using a variety of performance criteria. The ensemble model consistently demonstrated superior performance across all metrics, exhibiting an accuracy and f1-score of 97%. In comparison, the DenseNet model earned an accuracy and f1-score of 93%, while the ResNet model achieved an accuracy of 89% and a f1-score of 88%. All models in this paper demonstrated promising accuracy and the potential to aid in COVID-19 prediction. Chest x-ray images were employed to experiment the computational models of accurately predicting the disease. Such experiment allows us to have a better understanding of the advantages and disadvantages of various computer models for predicting sever disease, which will help create more precise and effective prediction systems for any medical condition. The achieve result highlights the efficacy of ensemble techniques for exploiting the synergistic benefits of multiple models. The knowledge gained from this study aims to go beyond the theoretical sphere and expand its influence into the real world of hospital administration.

Index Terms—Healthcare Management, CNN, Deep Learning, Ensemble, AI.

I. INTRODUCTION

One of the most contagious illnesses in human history, and more significantly in the twenty-first century, is COVID-19. The World Health Organization (WHO) formally labeled this lethal illness a pandemic after it had worldwide economic, social, and environmental effects on everyone in the world in recent years [1]. To be able to implement healthcare and epidemical treatment plans, accurate disease prediction is now more important than ever thanks to COVID-19. This research's goal is to provide thoughtful, well-reasoned responses to pressing problems by thoroughly examining the computational models now in use for the precise forecasting of COVID-19. This research will shed light on the benefits and drawbacks of these carefully chosen models by contrasting their perfor-

mances over a wide variety of criteria and indicators. This study will offer a convincing body of work supported by facts gathered during this journey, so to speak, in finishing this study by carefully examining their approaches, architectures, structures, layers, and datasets. By integrating the strengths of diverse algorithms and datasets, deep learning approaches have recently demonstrated considerable promise in improving the performance of disease prediction models. The number of computer models for COVID-19 prediction has increased, but this has also resulted in a lack of comprehensive, in-depth examination and comparison of the models' performance. Many of the models were developed independently using various techniques, datasets, and assumptions. It is now more difficult to assess the usefulness, accuracy, and applicability of these models in the real world due to this unexpected segregation. Additionally, this field necessitates clear precepts on the benefits and limitations of these computational models, which makes it challenging for researchers and healthcare practitioners to select the best model for each circumstance [2]. The rapid evolution of the COVID-19 also necessitates ongoing evaluation and improvement of the current computational models. To enhance the models' predictive power, it is vital to identify areas where they excel and those where they fall short. Therefore, the issue is typically the lack of a thorough knowledge of computational models for COVID-19 prediction. Effective model application and design are hampered by this information gap. This problem will be resolved and helpful insights for enhancing prediction accuracy

will be produced by a thorough review of the benefits and shortcomings of various computational models.

II. LITERATURE REVIEW

Using an unsupervised machine learning approach called K-means clustering on the attributes or features that were highly associated, Mujumdar et al. [3] conducted a study on the prediction of diabetes. When it came to properties like blood pressure and glucose level, which could not have empty or zero values, certain entries had some missing and inconsistent data. The data was also normalized using a scaling approach. They then built the models and assessed their accuracy using a variety of machine learning algorithms. Support Vector Classifier achieved an accuracy rate of 60%, followed by Random Forest Classifier (91%), Decision Tree Classifier (86%), K-

Nearest Neighbor (90%), Gaussian Nave Bayes (93%), and Logistic Regression (96%), which had the best accuracy. To ensure evaluation at several stages to promote consistency, they adopted a pipeline technique, which is relatively prevalent in machine learning. Finding the elements that influence non-diabetics to develop diabetes is the next goal for their research.

The study by Kablan et al. [4] evaluated the performance of base and meta learner models in stacked ensembles for predicting clinical outcomes, specifically severe cardiac events with COVID-19, from in-hospital patient data. Machine learning is used to analyze clinical data and predict patient outcomes. Stacked ensemble models are created by combining different base learners and meta learners. The results show that the Generalized Linear Model, Multi-Layer Perceptron, and Partial Least Squares meta learners have the highest performance, while K-Nearest Neighbors has the lowest. Numeric data is imputed after removing missing features. The methodology is designed to accurately assess the performance of base and meta learner models in stacked ensembles for clinical outcomes.

Su et al. [5] found that the WOCLSA model, a deep learning model combining Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), and Artificial Neural Network (ANN) outperforms other models. The model uses the Whale Optimization Algorithm to optimize parameters, achieving high accuracy, precision, and recall in identifying COVID-19 patients. The research involved 18 patient indicators and the team suggests that future research should focus on increasing the sample size of hospital patients to validate the predictive performance of WOCLSA.

Convolutional neural network architecture, a subset of deep learning known as CovXNet, was introduced by Mahmud et al. [6]. The CovXNet technique divides the normal, Covid-19, viral pneumonia, and bacterial pneumonia groups based on the varied dilation rates applied to the chest X-ray pictures. The training, transfer-learning, and testing parts were the three key elements that made up the CovXNet methodology's cogent structure. In the transfer-learning phase, the CovXNet model was updated by being trained once more on a smaller dataset that included both positive and negative patient Covid images. The training phase employed a dataset of negative patient Covid images. The model was applied to predict the class of X-ray images during the testing phase. The employment of two fundamental architectural components known as the shifter and residual units, which were required to distinguish between the different classes, was a significant component of this paradigm. The dataset used in this study was compiled from medical facilities in China and Bangladesh and included 305 COVID-19 X-ray pictures as well as 1583 normal, 1493 viral, and 2780 bacterial pneumonia images. 90% of the time, the model was accurate.

In their study, Badawi et al. [7] employed DenseNet201, VGG16, and VGG19 as the foundational models for conducting two distinct classifications: distinguishing between COVID-19 and normal cases, as well as performing a three-class classification including COVID-19, normal, and pneumo-

nia cases. The pre-existing models underwent modifications through the incorporation of additional components, including a flatten layer, average pooling layers, a dropout layer, and a fully linked layers which replaced the original final layers. Their dataset, consisting of 15,000 photos divided into COVID-19, normal, and pneumonia categories, was assembled from 11 distinct sources. The model displayed outstanding performance in binary classification using DenseNet201, with training and validation accuracies of 98.02% and 94.66%, respectively. The VGG16 model, on the other hand, performed better than expected with a training accuracy of 99.3% and a validation accuracy of 98.75%. Additionally, with training and validation accuracies of 99.02% and 98.59%, respectively, the VGG19 model produced impressive results. DenseNet201, VGG16, and VGG19 each achieved accuracy levels of 91.97%, 95.48%, and 94.96% for the multi-class classification challenge.

III. METHODOLOGY

A. Data Description

A diverse range of data sources were combined to create a robust dataset that proved useful for COVID-19 prediction. For a precise and efficient e xamination o f C OVID-19 prediction models, the collecting of chest X-ray pictures from several data sources was essential [8]. A sample of the images used can be seen in Figure 1 and 2. To ensure variation and a sense of representativeness in the dataset, various data sources and medical constituencies were used [9].

1) National Institutes of Health - Clinical Center: Some chest X-ray images were retrieved via vetted resources such as the NIH chest X-ray collection [10]. The dataset consists of two distinct classes, encompassing a total of 1,327 CT images and 263 Chest X-Rays text-mined samples that have been labelled as positive for COVID-19.

2) Kaggle: This well-known data science and analysis site gave users access to datasets with chest X-ray pictures pertinent to COVID-19 cases, which were thoroughly investigated [11]. The dataset has three unique categories, including 137 photos of COVID-19 and a total of 317 images comprising Viral Pneumonia and Normal Chest X-Rays. These images are organised into separate directories for testing and training purposes.

A range of chest X-ray pictures particularly collected for COVID-19 prediction make up the dataset used for the models. To facilitate later comparison, all the models were trained on the same set of data. A total of 2000 chest X-ray pictures collected from various sources make up the collection. To improve the robustness and generalizability of models, a wide variety of COVID-19 examples were offered by the images chosen. A subfolder of normal and COVID-19 instances from the dataset was fed into the models for training before being randomly assigned to test and validate them.
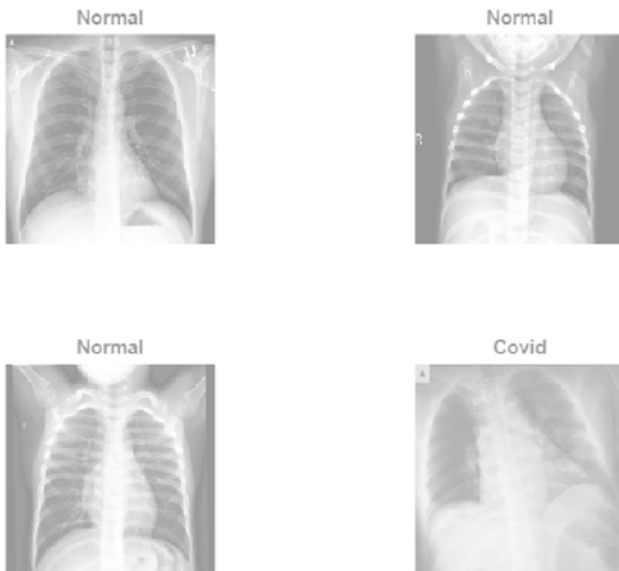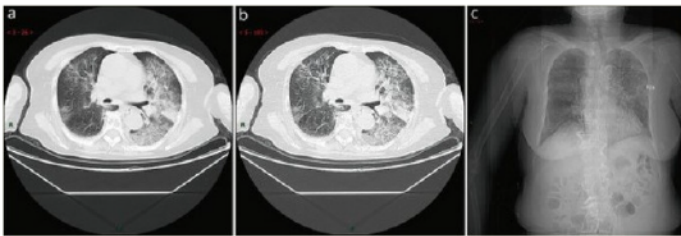
Fig. 1. Sample chest X-ray images Datset 1 [11]

Fig. 2. Sample CT and chest X-ray images Dataset 2 [10]

B. Data Preprocessing

Preprocessing measures were conducted to produce industrially approved standard chest X-ray pictures, which was more critical given the utilization of various data sources, to ensure data quality and integrity, which was required for an impactful and accurate analysis.

- Image Cleaning and Enhancement: To preserve the integrity and uniformity of the dataset across the board. A rigorous denoising, image-enhancing, and sharpening process was necessary for some photos.
- Image Standardization: Having a constant resolution of the photos to speed up model training is a typical approach in data science and analysis, which we also used. Grayscale conversion was required due to this uniformity requirement, which increased the computing effort.
- Augmentation: Data augmentation, which includes rotation, flipping, and zooming to mention a few, is another popular and useful strategy to prevent overfitting of the models on the dataset [12].
- Normalization and Anomaly Detection: The pixel intensity normalization method was required to provide uniform pixel values across all photos. Additionally, anomaly

detection was required to remove potentially false data.
- Data Balancing: It has been demonstrated that a balanced dataset is advantageous for model training, which helps the model avoid biases during the testing phase.
- Resizing: To establish a consistent dimension scaling and eliminate computational model mistakes caused by varying image resolutions, the dataset's photos needed to be resized.

The photos were treated to standardized preprocessing processes before the data was utilized as input for the computational models to ensure compatibility and uniformity.

C. Model Design

1) Model Selection: For accurate and thorough examination of a variety of computational models for the reliable forecasting of COVID-19, the choice of appropriate computational models is of the utmost importance. We looked at several computational model types, but we ultimately chose to employ pretrained deep learning models as they can be used as a starting point because they have previously learned and retrieved effective and useful features. The ResNet and DenseNet models are the base models used in this paper for the following reasons.

- CNNs, or convolutional neural networks: Convolutional neural networks will be a crucial component of this study because they are appropriate for picture classification and analysis projects [13]. Diverse models and architectures, including ResNet and Inception architectures, will also be assessed. DenseNet is a deep convolutional network with several layers.
- Transfer Learning: Convolutional neural networks that have already been trained, especially those that have been pretrained on ImageNet, will be fine-tuned on chest X-ray pictures [14], allowing models to use significant features discovered from datasets to enhance performance in prediction.
- Ensemble Models: The performance of models in the prediction of COVID-19 can be improved by combining numerous convolutional neural network models, as shown by an analysis of ensemble methodologies.
- Pretrained Models: Traditional pretrained learning models like ResNet and DenseNet will be used to create a baseline for comparison because they are extremely relevant for image classification tasks.

In the context of transfer learning, the entire procedure includes utilizing a model that has already been trained for one segment to tackle a related issue. To do this, the pretrained models' architecture must be modified to fit the new task at hand.

Once the data cleaning and model selection procedures have been accomplished, the training and testing of the models can be viewed. The complete process of creating the model is depicted in its entirety in Figure 3.
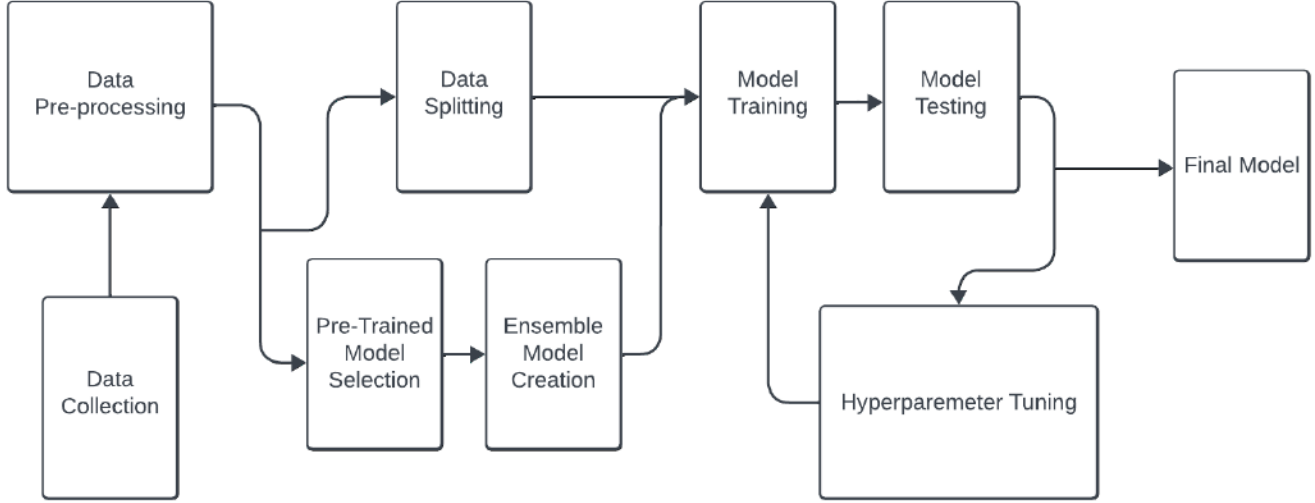
Fig. 3. Model Flowchart

D. Performance Evaluation

1) Evaluation Metrics: The effectiveness of the implemented computational models for COVID-19 illness prediction will be evaluated utilizing a comprehensive and detailed set of assessment measures and the dataset of chest X-ray images. These metrics will facilitate the creation of analyses and insights into many facets of model performance and will also be applied subsequently to carry out a comparative analysis.

- Accuracy: One of the measures used for evaluation will be the models' accuracy, which will show the proportion of instances in the dataset that were properly predicted out of all instances. This explains how we should evaluate and understand the model's overall accuracy.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1)$$

- Specificity and Sensitivity: These metrics are frequently combined because they exhibit tradeoffs that can be seen in the model and are associated. Sensitivity, which is sometimes referred to as recall, tells us how well the model can identify COVID-19 positive cases among the overall positive cases. On the other hand, specificity accomplishes the opposite because it tells us whether the model accurately distinguishes COVID-19 negative cases from the overall negative cases.

$$Senitivity = \frac{(TP)}{(TP+FN)} \quad (2)$$

$$Specificity = \frac{(TN)}{(TN+FP)} \quad (3)$$

- Precision and F1 Score: The precision evaluation metric measures the percentage of correctly predicted instances among all the postulated cases. This is particularly essential when there are a lot of false positives. The F1 score

where beta is 1 depends on recall and precision in that it is the harmonic mean of both and offers an accurate evaluation of the model's performance.

$$F_\beta = (1+\beta^2) \times \frac{precision \times recall}{(\beta^2 \times precision) + recall} \quad (4)$$

- Confusion Matrix: This provides a visual summary of the model's performance and displays the number of true positives, true negatives, false positives, and false negatives. It is more visually appealing but still contains numerical analytical content. This makes the types of errors the model makes easier to see and comprehend [15].

2) Transfer Learning Approach: In the context of transfer learning, the entire procedure includes utilizing a model that has already been trained for one segment to tackle a related issue. To do this, the pre-trained models' architecture must be modified to fit the new task at hand. As an illustration, we may use an image classification model that was initially trained on the massive ImageNet dataset (which consists of millions of labeled photos) and use it as a solid foundation for the specialized job of recognizing COVID-19 using a smaller dataset. When training data is limited, transfer learning is very helpful, especially in specialist fields like medical picture categorization for developing disorders. With deep neural network-based models, which have many parameters, this problem is extremely severe. A pre-trained model can be used for a particular task in one of two ways. The internal weights of the pre-trained model stay unchanged for the new task in one method where the pre-trained model functions as a feature extractor. To accomplish classification, a classifier is then trained on top of this feature extractor. An alternative strategy involves fine-tuning the entire network or a portion of it for the new purpose. We chose to just make minor changes to the convolutional neural network's final layer. The pre-trained

models were essentially employed as feature extractors. We compared ResNet50 [14] and DenseNet-201 [16], two popular pre-trained models. The architectures of these models and how they were used to predict COVID-19 are briefly described in the section that follows.

IV. RESULTS

In this section, we present the results of the ResNet model, the DenseNet model, and the ensemble model, along with an analysis of their predictions for COVID-19 using the chest X-ray dataset. The dataset was randomly divided into training (60%), testing (20%), and validation (20%) sets using the pandas and scikitlearn libraries in python.

A. ResNet Model

The ResNet model was trained, tested, and validated using the chest X-ray dataset. It was implemented in the MATLAB software environment with appropriate adjustments made to the dataset for uniform dimensions. Preprocessing was applied to meet the model's input size requirement, and modifications were made to the final classification and fully connected layers. The model's performance was evaluated using relevant criteria and produced the results the shown in Figure 4.
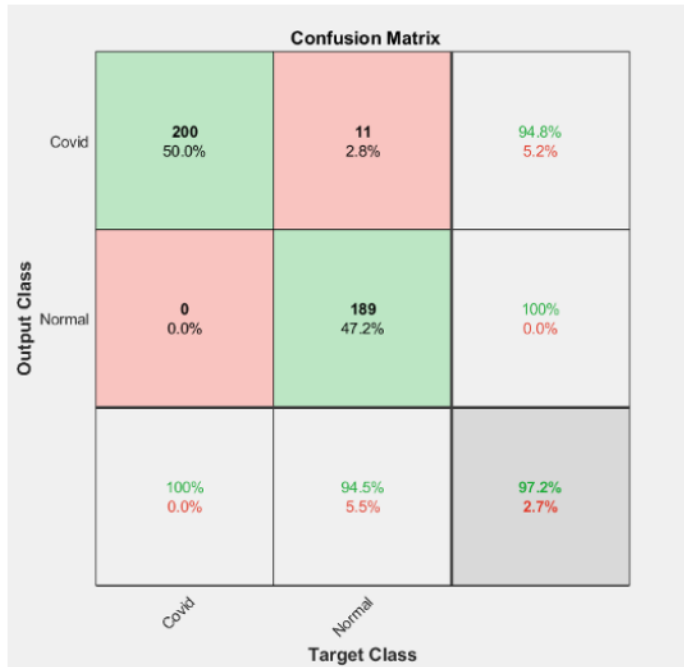

Fig. 4. ResNet Confusion matrix

The depth of the ResNet architecture distinguishes it; for example, ResNet-50 and ResNet-101 have 50 and 101 layers, respectively. Residual blocks, consisting of multiple convolutional layers, batch normalization, and ReLU activation functions, are the fundamental units of the network. Each stage is composed of these residual blocks, and each stage typically reduces spatial dimensions while increasing the number of filters. This can be seen in Figure 5.
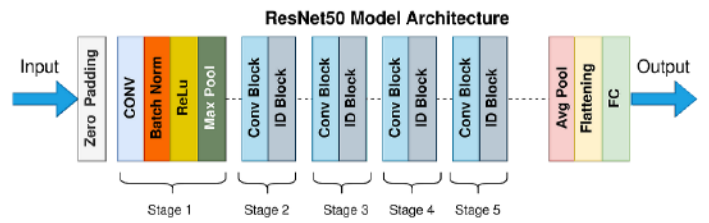

Fig. 5. ResNet Architecture [17].

B. DenseNet Model

The DenseNet Model followed a similar procedure to the ResNet Model to produce its results. Dataset expansion was performed to ensure error-free training, and modifications were made to the final fully connected and classification layers. Training and testing phases were executed, and the model's performance was assessed. The results can be seen in Figure 6.
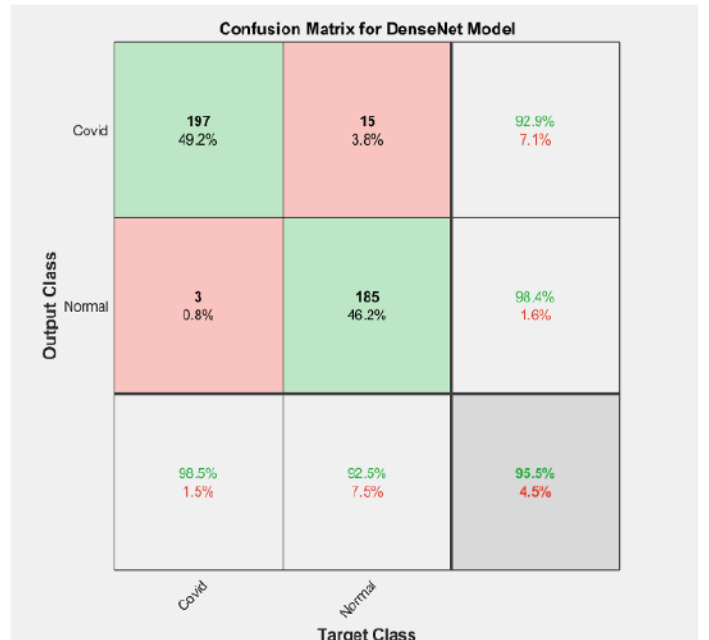

Fig. 6. DenseNet Confusion matrix

DenseNet consists of dense blocks, each containing a series of convolutional layers with batch normalization and ReLU activation functions. Transition layers are incorporated between dense blocks to reduce spatial dimensions and restrict the growth of feature maps. Convolution and pooling procedures are typically used in transition layers. Hyperparameters like the growth rate and the number of dense blocks can be adjusted to modify the network's architecture. This can be seen in Figure 7.

C. Ensemble Model

The ensemble model was formed by combining the ResNet and DenseNet models. The majority voting technique was
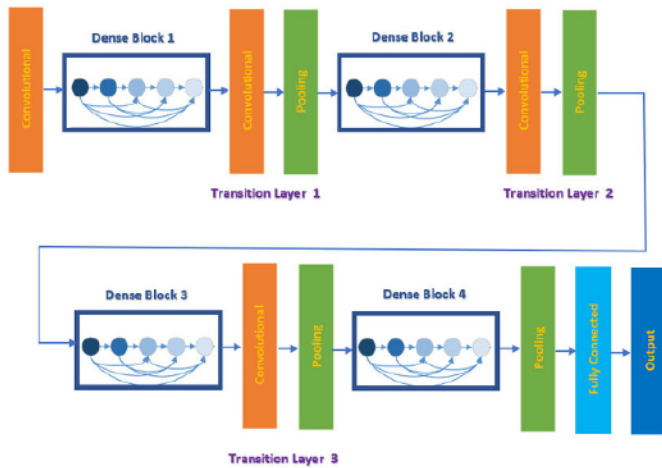
Fig. 7. DenseNet Architecture [18].

employed to create the ensemble model, where the final prediction of the ensemble is determined by the class predicted by most of the individual models. The individual models were saved locally and loaded into the environment for ensemble creation. The dataset was loaded, resized, and enhanced before using the individual models to generate predictions. These predictions were then combined to form the ensemble model (voting), which was evaluated using the required assessment metrics, as seen in Figure 8. The final results are shown in Table I.
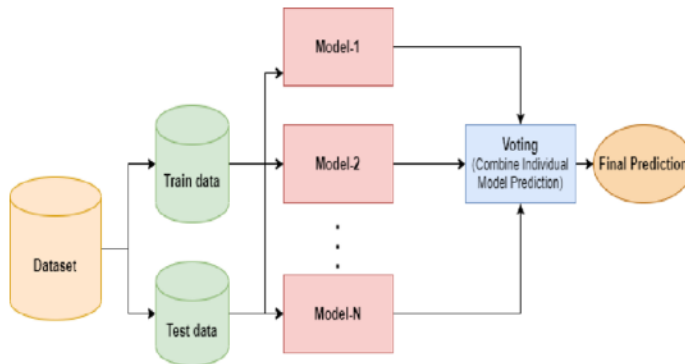

Fig. 8. Ensemble Model Architecture [19].

TABLE I
FINAL MODEL PERFOMANCE RESULTS

Metric	ResNet	DenseNet	Ensemble Model	GWOCLSA [5]
Accuracy	89%	93%	97%	93%
Precision	90%	91%	99%	93%
Recall	87%	95%	95%	93%
F1-Score	88%	93%	97%	93%

V. CONCLUSION AND FUTURE WORK

This work built the first step towards developing a generalized model to assess the healthcare parameters at a personalized and community dimensions and predict the severity of the disease before becoming pandemic. It focused on offering a thorough evaluation of the effectiveness of different computer models for predicting severe disease utilizing the dataset of chest x-ray images. Using a variety of performance criteria, the outcomes of trials using the ResNet, DenseNet, and ensemble models were meticulously assessed. The ensemble model consistently demonstrated superior performance across all metrics, exhibiting an accuracy and f1-score of 97%. In comparison, the DenseNet model earned an accuracy and f1-score of 93%, while the ResNet model achieved an accuracy of 89% and a f1-score of 88%. They showed encouraging accuracies and potential to help with severe disease prediction such as COVID-19. Notably, the Ensemble Model achieved a higher performance in all metrics outperforming the GWOCLSA model by su et al. [5]. This study adds to the body of knowledge in the field of medical image analysis by demonstrating how well computational models predict COVID-19 from chest x-ray pictures. On the other hand, it emphasizes how effective ensemble techniques are for utilizing the synergistic benefits of various models. Exploring more sophisticated designs, particularly the area of transformer-based models, is one of the interesting areas for future research to achieve the general model. These architectures, which were first created for challenges involving natural language processing, have demonstrated astounding potential in picture analysis. The capacity to capture long-range dependencies in images could be used to improve disease prediction accuracy even more. Investigating these sophisticated architectures may lead to the discovery of fresh information that completely alters the way disease is predicted. We are working on improving the model further by using more sophisticated structures, such as transformer-based models, as these architectures are made to identify enduring dependencies in images and other clinical parameters. In conclusion, this study acts as a steppingstone into a future that is rife with opportunities and difficulties. It is our common responsibility to use innovation to advance healthcare as we forge new paths in the field of disease prediction utilizing computer models. We can create a future in which computational models play a key role in enhancing patient outcomes and global health through continuous research, collaboration, and ethical concerns.

REFERENCES

[1] Domenico Cucinotta and Maurizio Vanelli. Who declares covid-19 a pandemic. *Acta bio medica: Atenei parmensis*, 91(1):157, 2020.

[2] Mohammad Pourhomayoun and Mahdi Shakibi. Predicting mortality risk in patients with covid-19 using machine learning to help medical decision-making. *Smart health*, 20:100178, 2021.

[3] Aishwarya Mujumdar and V Vaidehi. Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, 165:292–299, 2019.

[4] Rianne Kablan, Hunter A Miller, Sally Suliman, and Hermann B Frieboes. Evaluation of stacked ensemble model performance to predict clinical outcomes: A covid-19 study. *International Journal of Medical Informatics*, 175:105090, 2023.

[5] Xiaoying Su, Yanfeng Sun, Hongxi Liu, Qiuling Lang, Yichen Zhang, Jiquan Zhang, Chaoyong Wang, and Yanan Chen. An innovative ensemble model based on deep learning for predicting covid-19 infection. *Scientific Reports*, 13(1):12322, 2023.

[6] Tanvir Mahmud, Md Awsafur Rahman, and Shaikh Anowarul Fattah. Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with

transferable multi-receptive feature optimization. *Computers in biology and medicine*, 122:103869, 2020.

[7] Abeer Badawi and Khalid Elgazzar. Detecting coronavirus from chest x-rays using transfer learning. *Covid*, 1(1):403–415, 2021.

[8] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. *arXiv preprint arXiv:2003.11597*, 2020.

[9] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*, 296(2):E32–E40, 2020.

[10] Yifan Peng, Yuxing Tang, Sungwon Lee, Yingying Zhu, Ronald M Summers, and Zhiyong Lu. Covid-19-ct-cxr: a freely accessible and weakly labeled chest x-ray and ct image collection on covid-19 from biomedical literature. *IEEE transactions on big data*, 7(1):3–12, 2020.

[11] Pranav Raikote. Covid-19 image dataset, April 2020.

[12] Weifang Kong and Prachi P Agarwal. Chest imaging appearance of covid-19 infection. *Radiology: Cardiothoracic Imaging*, 2(1):e200028, 2020.

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[15] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer, 2014.

[16] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[17] Suvaditya Mukherjee. The annotated ResNet-50. https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758, August 2022. Accessed: 2023-9-16.

[18] Omneya Attallah. Mb-ai-his: histopathological diagnosis of pediatric medulloblastoma and its subtypes via ai. *Diagnostics*, 11(2):359, 2021.

[19] Md Akbar Hossain, Nawrin Zaman, Mahmuda Begum, Rakibul Hasan, and Emadul Islam Shovon. Explainable artificial intelligence to improve human decision support in heart disease. November 2022.