Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Artificial Intelligence based Covid19 detection using CTScan and Chest X-Ray Images: Machine Learning and Deep Learning techniques

[2]G. G. Rajput, [1]Ashvini Alashetty

[1] [2]Department of Computer Science, Karnataka State Akkamahadevi Women's University, Vijayapura, Karnataka-586108

E-mail: [2]ggrajput@yahoo.co.in, [1]ashwinialashetty@gmail.com

Abstract: COVID-19, also known as the coronavirus disease 2019 is a highly contagious respiratory illness caused by SARS-COV-2 virus which spread form china globally resulting in pandemic. With advances in diagnostic tools, radiologic imaging is widely used for COVID-19 pneumonia diagnosis other than usual clinical and laboratory testing. In this paper, several deep learning and machine learning enhanced techniques are applied to X-Ray and CT-Scan medical images for the detection of covid-19 and also a clinical data for the prediction of the covid19. The images are preprocessed and trained using U-Net model, a popular architecture for image segmentation tasks. The accuracy and F1-score were found to be above 98% in the diagnosis of COVID-19 using CT-scan images. Further, transfer learning techniques were applied to overcome the insufficient data and to improve the training time. The binary and multi-class classification of X-ray images tasks were performed by utilizing enhanced CNN deep transfer learning architecture. An accuracy of 99% was achieved by enhanced CNN in the detection of X-ray images from COVID-19 and pneumonia.Further,using clinical dataset we compared the performance of supervised machine learning algorithms: logistic regression, random forest classifier, and XGBoost classifier. These algorithms were trained and evaluated on the preprocessed and feature-selected dataset to predict COVID-19 cases. The results showed that XGBoost classifier outperformed logistic regression and random forest classifier in predicting COVID-19 cases based on symptoms, age, gender, and test indications.

1. Introduction

COVID-19, an illness caused by the SARS-CoV-2 virus, was officially designated by the World Health Organization (WHO) on February 11, 2020 (World Health Organization, 2020)[1]. Since the initial case was identified, COVID-19 has rapidly disseminated across numerous nations, resulting in the loss of more than 4 million lives and affecting nearly 180 million individuals, according to statistics provided by the World Health Organization as of June 2021[2]. The initial step in managing COVID-19 involves screening patients at primary healthcare centers or hospitals. While the definitive diagnosis primarily relies on transcription-polymerase chain reaction (PCR) tests, medical imaging has emerged as a preferred protocol in hospitals for individuals displaying severe respiratory symptoms. This approach offers a rapid and straightforward means for doctors to promptly identify diseases and their associated impacts3 Following this established protocol, individuals suspected of COVID-19 undergo an initial X-ray examination, and if further details are required, a subsequent CT-scan session is conducted. Consequently, computed tomography scan (CT scan) and X-ray images have gained significant usage in clinical settings as alternative diagnostic tools for detecting COVID-19 and assessing the virus's effects[4].In the diagnostic process, healthcare professionals rely on X-ray or CT-scan images to examine the lungs and identify signs of COVID-19-related abnormalities. The rapid transmission rate of the virus has led to a surge in hospital admissions within a short timeframe, creating a significant burden for imaging physicians and exacerbating the shortage of healthcare providers in the fight against the disease. Deep learning methods offer a potential solution to this challenge, as they have made remarkable advancements in recent years. These advancements can be attributed to the increasing computational power, the growing availability of data, and the continuous improvement of deep learning models and algorithms. This progress has been demonstrated through challenge competitions, where deep learning models have achieved record-breaking performances[5]. The essence of deep learning lies in constructing multi-layered machine learning models that utilize hidden layers to learn more accurate features. These models are trained with a substantial amount of sample data, ultimately enhancing the accuracy of classification or prediction tasks[6,7].The side effects of COVID-19 being like viral pneumonia can at times lead to wrong determination in the present circumstance, where medical clinics are over-burden and working nonstop[8,9].

2. Literature Review

Researchers have developed a variety of machine learning and deep learning models specifically tailored for the detection of COVID-19 using X-ray datasets. While machine learning methods are valuable for critical tasks, they typically require manual feature extraction from images. In contrast, deep neural network models leverage computer vision techniques to automatically extract features, making them well-suited for COVID-19 detection.Convolutional Neural Network (CNN) has emerged as one of the most commonly used and effective methods for diagnosing COVID-19 from digitized chest X-ray images. Numerous reviews have been conducted to summarize the recent advancements and contributions in the field of COVID-19 detection. These reviews provide valuable insights into the performance and capabilities of various deep learning approaches, including CNNs, in accurately identifying COVID-19 cases based on X-ray images.In recent years, deep learning-based medical image processing systems have demonstrated effectiveness in the classification and diagnosis of Chest X-ray images for COVID-19. Keidar et al. achieved an accuracy of 90.3% by employing augmentation and normalization techniques on CXR images using deep learning models such as ResNet50, ResNet[15,2], and VGG[16,17]. Shelke et al. designed an automated COVID-19 screening system using deep learning, which can further classify COVID-19 cases into mild, medium, and severe categories18. Various convolutional neural network architectures, including ResNet101, Xception, InceptionV3, InceptionResNetV2, VGG16, and VGG19, have been utilized for COVID-19 classification from chest X-ray images19.Hammoudi et al. developed a hierarchical classification approach for COVID-19 detection, integrating it with pneumonia viral classification. Their study achieved an average accuracy exceeding 84% 20. Abbas et al. developed a deep Convolutional Neural Network called DeTraC (Decompose, Transfer, and Compose) for the classification of COVID-19 and SARS from CXR and CT images, achieving a high accuracy of 93% 21. These studies highlight the effectiveness of deep learning techniques in COVID-19 classification and demonstrate the potential of different models in achieving accurate and reliable results.In the classification of chest X-ray images into healthy and COVID-19 categories, Rekha et al. employed a transfer learning model that utilized XGBoost with a convolutional neural network (CNN) for feature extraction22. Guefrechi et al. designed an effective detection system for COVID-19 using deep learning and transfer learning techniques with ResNet50, InceptionV3, and VGG16. They achieved high performance in classifying COVID-19 and pneumonia cases, even with a small image dataset, by fine-tuning the models 23. Bozkurt et al. developed a framework for COVID-19 diagnosis from chest X-ray images, incorporating both deep learning-based features and handcrafted features. They employed k-nearest neighbors (kNN), Naive Bayes (NB), and MLPClassifier algorithms for classification 24. Elgendi et al. focused on geometric transformation as a perspective of image augmentation in deep learning networks for COVID-19 detection[25]. Sousa et al. utilized CNN-COVID for the classification of COVID-19 from chest X-ray images, highlighting the cost-effectiveness and quick results obtained compared to computed tomography[26].

The proposed framework is compared with previous works in terms of several performance metrics such as accuracy, f1-score, precision, and recall. With an extensive evaluation to validate the proposed methods, we find the proposed UNet and CNN deep transfer learning model shows excellent performance on binary and three-class classification tasks, the accuracy of the model is as high as 99%.

3. Methodology

In the majority of research studies, Convolutional Neural Networks (CNNs) have been widely employed for COVID-19 classification using X-ray image datasets. In this paper, we present a COVID-19 classification system using X-ray image dataset, incorporating Data Augmentation and Hyperparameter Tuning techniques.The proposed work follows a four-phase approach (Figure 1).

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

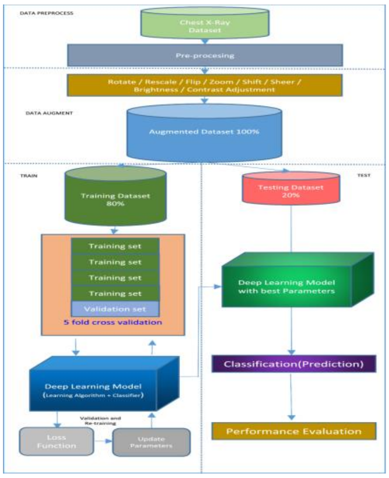

Fig 1: A deep learning-based COVID-19 classification from chest X-ray image: case study

3.1 COVID-19 X-ray image dataset

The researchers of Qatar University have compiled the COVID-QU-Ex dataset, which consists of 21165 chest X-ray (CXR) images including: 3616 are COVID19, Viral Pneumonia are 1345, lung opacity patients are 6012 and normal patients are 10192. Collected the data from the kaggle website30. The dataset containing COVID-19 radiography images are acquired from popular Cohen/IEEE8023 dataset which is the real data collected recently. Sample images are shown in Figure 2.

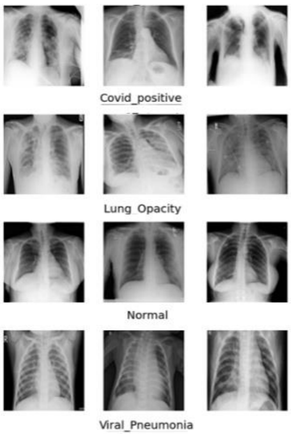

Fig 2: Chest X-ray augmented images of COVID-19, Lung Opacity, Normal and Viral_Pneumonia

3.2 Data augmentation

COVID-19 chest X-ray datasets are obtained from public donations; however they are insufficient for complex CNN training to get more accuracy To increase the performance of deep learning model we need more data, so augmentation technique is used to increase the dataset and performance accuracy.Converting an image to the "b" channel, represented by the equation B(r, g) = I(r, g, b), provides a valuable technique for exploratory data analysis in image processing. The mean blue intensity (μ_b), standard deviation (σ_b), and histogram analysis (h_b(k)) contribute to the understanding of the blue color distribution within the image dataset. By considering the specific objectives and requirements of the analysis, the conversion to the "b" channel enhances the utilization of image data for further analysis and machine learning tasks.Ben Graham's method is a valuable approach for exploratory data analysis in image datasets. By calculating the mean and standard deviation of pixel values and visualizing them through scatter plots, this method provides a comprehensive understanding of the distribution, patterns, and potential preprocessing requirements of the dataset. It helps detect outliers and anomalies, uncover correlations between mean and standard deviation, and guide the selection of appropriate preprocessing techniques. Overall, Ben Graham's method enables researchers to gain valuable insights and make informed decisions during the analysis of image datasets. The mathematical equations of Ben Graham's method, involving the calculation of mean and standard deviation, coupled with scatter plot visualization, provide a robust approach for exploratory data analysis in image datasets. By leveraging these equations, researchers can detect outliers, uncover correlations, and determine suitable preprocessing techniques, leading to valuable insights and informed decision-making during the analysis of image datasets. Figure 3 shows different types of augmentations.

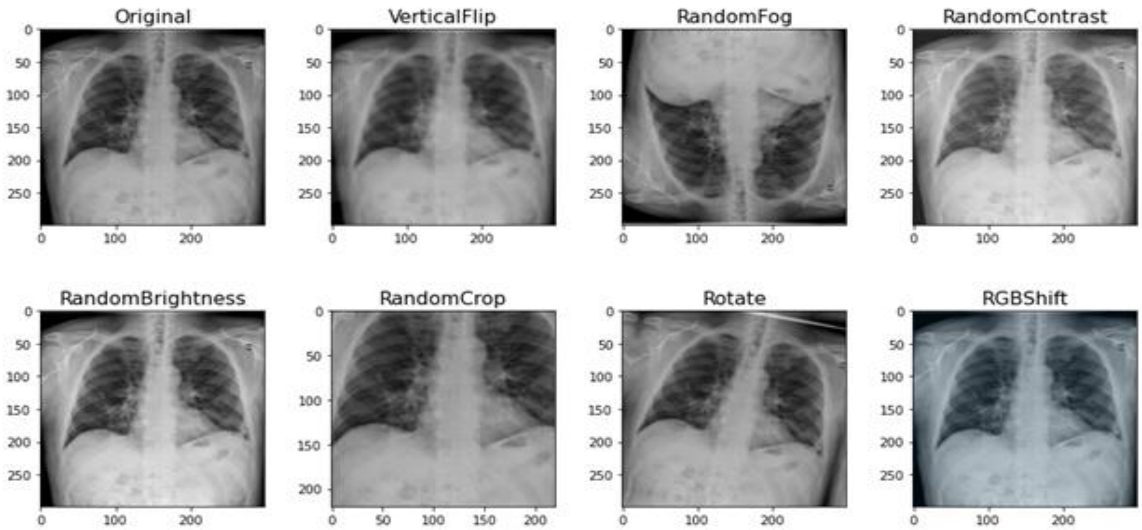

Fig 3: Different Types of Augmentations

The detail description of number of images utilized for train, test and validation is shown in table

Table 1:

Dataset	Original images	Train images	Test images	Validation
COVID-19	3616	2892	723	362
Viral Pneumonia are	1345	1,076	269	134
lung opacity	6012	4,809	1,202	601
Normal	10192	8,153	2038	1019
Total	21165	15238	4233	1694

Chest X-ray scan images were used to build and test the model. The training dataset was split with 80% for constructing the model 20% for Testing. The performance of the model is measured using 10-fold cross-validation technique.

3.4 CT-Scan Lung Images dataset

CT scans are important for diagnosing COVID-19 and assessing its severity.We used a combination of two datasets to conduct the experiments with the CNN model. They include scanned Lung CT Scan images with Lung_Mask, infection_mask and Lung_and_infection_mask features. These images are pre-processedand used for training the models of the UNet model. Models that can detect COVID-19 evidence and characterize findings are valuable, particularly where radiologists are scarce. A dataset with 20 CT scans of COVID-19 patients and expert-made segmentations is available for research and development purposes.To assess the effectiveness of our proposed model trained on different scales of images, we encountered an inconsistent number of X-ray and CT-scan images. As a result, we focused on the CT-scan dataset, which consisted of a total of 20 images. Out of these, 80% images were allocated for training, while the remaining 20% were used for validation. The dataset included both CT scan Images of lungs, infected lungs cases of mask.

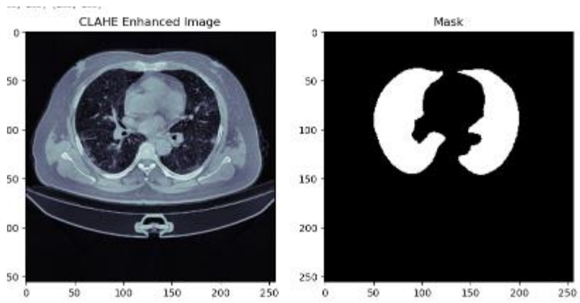

Fig 4: CLAHE Enhanced images with Mask Image

The code processes the lung images and CT scans to prepare them for further analysis. It converts the lung images into binary masks by setting all values greater than 0 to 1. This step is likely performed to segment the lung areas from the images. Additionally, the CT scans are enhanced using the CLAHE technique to improve contrast (Fig. 4 and 5).

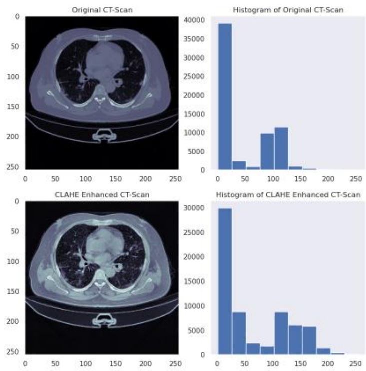

Fig 5: Original CT-Scan and its Histogram

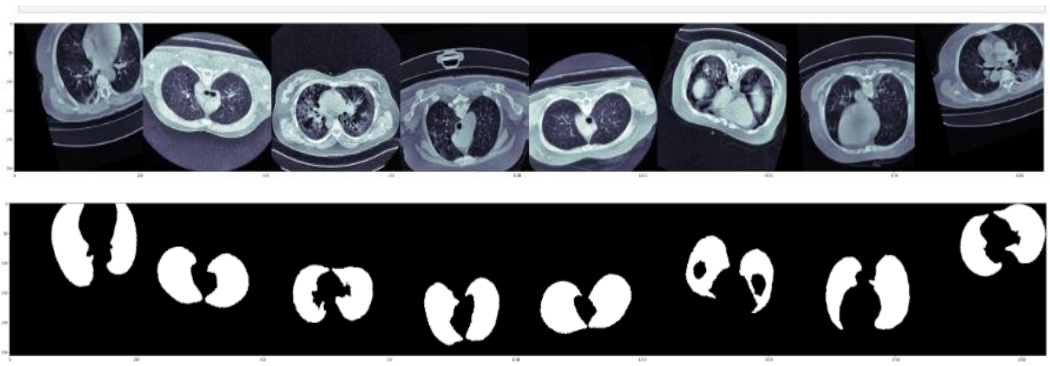

Fig 6: CT Scan Images and Mask Images in 60 degree angle visualization

The lung images and CT scans are preprocessed by converting the lung images into binary masks and enhancing the CT scans using the CLAHE technique.Data augmentation is applied including flipping, affine transformations, and random order sequencing.The augmented CT scan and lung mask images are reshaped to a desired dimension and stored in new arrays.A subset of the augmented images is randomly selected and displayed in a grid format for visualization.The augmented images are concatenated with the original images and normalized Samples of the CT scan images and their corresponding lung masks are plotted to illustrate the data characteristics (Fig. 6).

3.5 Clinical Data on Covid19

Clinical data on COVID-19 plays a crucial role in understanding the disease, its impact on individuals, and guiding effective healthcare interventions. This dataset includes various clinical parameters and information related to patients infected with COVID-19such as the date of the test, symptoms like cough, fever, sore throat, shortness of breath, and headache, along with demographic information like age group and gender. Additionally, the dataset includes the COVID-19 test results and the indication or reason for the test.The data set enables researchers and healthcare professionals to identify common symptoms and their prevalence, understand the relationship between symptoms and test results, and assess the impact of age and gender on disease severity. Furthermore, this data can aid in identifying high-risk groups, informing public health strategies, and evaluating the effectiveness of interventions and treatments.

Data Description: The data is described in terms of Cough, Fever, Sore Throat, Shortness of Breath & Headache [10]. There are a total of 5,861,480 records.

3.6 Data Pre-Processing:

Label Encoding:Categorical attributes, such as 'corona_result', 'gender', and 'age_60_and_above', are transformed into numerical representations using label encoding.Each category is assigned a unique numerical value (e.g., 'Negative' is mapped to 0 and 'Positive' is mapped to 1 for 'corona_result').Logistic Regression using GridSearchCV: GridSearchCV is applied to tune the hyperparameters of the Logistic Regression model. By specifying a parameter grid, the algorithm searches for the best combination of hyperparameters (such as regularization strength or penalty type) that optimizes the model's performance.M2 - Random Forest using GridSearchCV: Similarly, GridSearchCV is employed to optimize the hyperparameters of the Random Forest model. The parameter grid can include parameters like the number of trees, tree depth, or feature selection criteria, among others, to fine-tune the Random Forest model.M3 - XGBoost using GridSearchCV: GridSearchCV is utilized to find the optimal hyperparameters for the XGBoost model. Parameters such as learning rate, tree depth, regularization parameters, or subsampling ratio can be explored in the parameter grid to enhance the performance of the XGBoost model.By using GridSearchCV for all three models and tuning their hyperparameters, best configurations for each model is obtained resulting in improved model performance and predictive accuracy.

4. Experimental results

Table 2: Results

Models	Recall	Specificity	Accuracy	Precision	F1score
Logistic Regression	0.999	0.964	0.976	0.943	0.970
Random Forest	0.999	0.963	0.977	0.943	0.970
XgBosst	0.999	0.963	0.977	0.943	0.970

5. CT Scan Images Results

Constructs a UNet model architecture for segmentation and trains the model using the training data. The results would include evaluation metrics such as loss and accuracy, as well as the predicted segmentation masks for the CT scans.Segmentation of COVID-19 CT scans can play a crucial role in assisting medical professionals in the accurate identification and localization of affected regions.

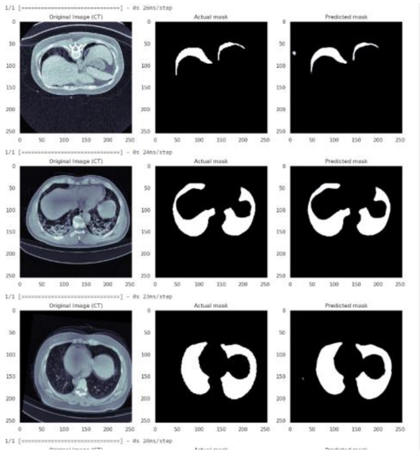

Fig 12: Comparative results with Actual Mask and Predicted mask of lung images

6. X-Ray Images Results:

The CNN model will be trained on the training data and evaluated on the testing data. The evaluation metrics such as accuracy, precision, recall, and F1-score can be calculated to assess the performance of the model in classifying the corona result categories.

Accuracy.

Table 3: of Accuracy results

	Mini	Max	current
Training	0.981	1.00	1.00
Validation	1.00	1.00	1.00

7. Conclusion

Combining X-ray images, CT scan images, and clinical data has several advantages and provides valuable insights in diagnosing and understanding medical conditions.The combination of X-ray images, CT scan images, and clinical data enhances diagnostic accuracy by leveraging multiple sources of information.The integration of imaging data and clinical data enables a comprehensive evaluation of patients, considering both anatomical and clinical features.This multimodal approach improves the detection and characterization of abnormalities, leading to more precise diagnoses.The combined analysis of imaging data and clinical data provides a holistic view of the patient's condition, aiding in treatment planning and monitoring.By merging different data modalities, healthcare professionals can gain a deeper understanding of the disease progression and response to therapies.The fusion of imaging and clinical data promotes personalized medicine, allowing tailored treatment strategies based on individual patient characteristics.This comprehensive approach facilitates interdisciplinary collaboration among radiologists, clinicians, and other healthcare specialists, fostering a team-based approach to patient care.Overall, the integration of X-ray images, CT scan images, and clinical data synergistically improves diagnostic accuracy, treatment planning, and patient outcomes in a wide range of medical conditions.

References

[1] Sohrabi, C. *et al.* World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int. J. Surg.* **76**, 71–76 (2020).

[2] World Health Organization. Weekly epidemiological update on COVID-19—29 June 2021, 46th edn. (2021). https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-june-2021.

[3] Sandri, T. L. *et al.* Complementary methods for SARS-CoV-2 diagnosis in times of material shortage. *Sci. Rep.* **11**, 1–8 (2021).

[4] Elsharkawy, M. *et al.* Early assessment of lung function in coronavirus patients using invariant markers from chest X-rays images. *Sci. Rep.* **11**, 1–11 (2021).

[5] Bejnordi, B. E. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **318**, 2199–2210 (2017).

[6] Amin, J., Sharif, M., Yasmin, M. & Fernandes, S. L. Big data analysis for brain tumor detection: Deep convolutional neural networks. *Future Gener. Comput. Syst.* **87**, 290–297 (2018).

[7] Pastur-Romay, L. A., Cedrón, F., Pazos, A. & Porto-Pazos, A. B. Deep artificial neural networks and neuromorphic chips for big data analysis: Pharmaceutical and bioinformatics applications. *Int. J. Mol. Sci.* **17**, 1313 (2016).

[8] Goceri, E. & Goceri, N. Deep learning in medical image analysis: Recent advances and future trends. IADIS Digital. *Library* (2017).

[9] Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).

[10] Asnaoui, K. E., Chawki, Y. & Idri, A. Automated methods for detection and classification pneumonia based on X-ray images using deep learning. arXiv preprint arXiv:2003.14363 (2020).

[11] Yao, Z. *et al.* A machine learning-based pulmonary venous obstruction prediction model using clinical data and CT image. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 609–617 (2021).

[12] Bhandary, A. *et al.* Deep-learning framework to detect lung abnormality—A study with chest X-ray and lung CT scan images. *Pattern Recognit. Lett.* **129**, 271–278 (2020).

[13] Kuchana, M. *et al.* Ai aiding in diagnosing, tracking recovery of COVID-19 using deep learning on chest CT scans. *Multimed. Tools Appl.* **80**, 9161–9175 (2021).

[14] Alshazly, H., Linse, C., Abdalla, M., Barth, E. & Martinetz, T. Covid-nets. Deep CNN architectures for detecting COVID-19 using chest CT scans. *medRxiv* (2021).

[15] Joaquin, A. Using deep learning to detect pneumonia caused by ncov-19 from X-ray images. https://towardsdatascience.com/using-deep-learning-to-detect-ncov-19-from-x-ray-images-1a89701d1acd (2020).

[16] Wang, D., Mo, J., Zhou, G., Xu, L. & Liu, Y. An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images. *PLoS One* **15**, e0242535 (2020).

[17] D. Keidar, D. Yaron, E. Goldstein et al., COVID-19 classification of X-ray images using deep neural networks. Eur. Radiol. **31**, 9654–9663 (2021). https://doi.org/10.1007/s00330-021-08050-1Shelke, M. Inamdar, V. Shah et al., Chest X-ray classification using deep learning for automated COVID-19 screening. SN Comput. Sci. **2**, 300 (2021). https://doi.org/10.1007/s42979-021-00695-5

[18] S. Sanket, M. Vergin Raja Sarobin, L. Jani Anbarasi et al., Detection of novel coronavirus from chest X-rays using deep convolutional neural networks. Multimed. Tools Appl. (2021). https://doi.org/10.1007/s11042-021-11257-5

[19] K. Hammoudi, H. Benhabiles, M. Melkemi et al., Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. J. Med. Syst. **45**, 75 (2021). https://doi.org/10.1007/s10916-021-01745-4Abbas, M.M. Abdelsamea, M.M. Gaber, Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl. Intell. **51**, 854–864 (2021). https://doi.org/10.1007/s10489-020-01829-7

[20] R. Rekha, Comparative analysis of COVID-19 X-ray images classification using convolutional neural network, transfer learning, and machine learning classifiers using deep features. Pattern Recognit. Image Anal. **31**, 313–322 (2021). https://doi.org/10.1134/S1054661821020140

[21] Senthil, P., M. Suganya, Ishwar Baidari, and S. P. Sajjan. "Enhancement Sushisen algorithms in images analysis technologies to increase computerized tomography images." International Journal of Information Technology 14, no. 1 (2022): 375-387.

[22] Senthil, P. "ENHANCED BIG DATA CLASSIFICATION SUSHISEN ALGORITHMS TECHNIQUES IN HADOOP CLUSTER (META)." Journal of Computer-JoC, Available Online at: www. journal. computer 1, no. 1 (2016): 14-20.

[23] Senthil, P., M. Stanly, and S. S. Inakshi. "Improve Multidimensional 5G OFDM Based MIMO Sushisen Algorithms Merge Multi-Cell Transmission." International Journal of Recent Engineering Science 7, no. 2 (2020): 17-21.

[24] Senthil, P. "Image Mining Brain Tumor Detection using Tad Plane Volume Rendering from MRI (IBITA)." Journal of computer science 1, no. 1 (2016): 1-13.

[25] Senthil, P., and M. Suganya. "Exchanged Nonlinear Third Order Differential Equation Ordinary Differential Equation." Journal for Research| Volume 4, no. 05 (2018).