Extraction of Features from Lung Image for the Detection of Covid-19

Sushmitha M*
Department of Biomedical Engineering, Saveetha Engineering College, Anna University, Chennai
sushmitham@saveetha.ac.in

Anjana S
Department of Biomedical Engineering, Saveetha Engineering College, Anna University, Chennai
honouranja@gmail.com

Kavi Bharathi R
Department of Biomedical Engineering, Saveetha Engineering College, Anna University, Chennai
kavibharathisai@gmail.com

Lavanya B
Department of Biomedical Engineering, Saveetha Engineering College, Anna University, Chennai
lavanyabaskaran0627@gmail.com

Bhavadharani V
Department of Biomedical Engineering, Saveetha Engineering College, Anna University, Chennai
bhavadharaniv@saveetha.ac.in

Abstract- The World Health Organization has identified two new human cases of Coronavirus type 2 (CoV-2)-induced severe acute respiratory syndrome and a new pandemic illness known as coronavirus disease (COVID-19). It was initially discovered in Wuhan, China, in December 2019, and afterwards spread widely. The novel COVID-19 continues to increase with confirmed cases that may result in deaths worldwide as a result of respiratory failure and alveolar damage. The common symptoms of SARS- CoV- include dyspnea, chest pain, fatigue and less common symptoms include headache, vomiting and nausea. The diagnosis of SARS-CoV2 infection was the key step to identify the infection and then prevention and quarantine are next steps to be considered as the way to stop the spreading of virus, since there is no specific vaccine, drugs or antiviral against the COVID-19 virus. The patients with COVID-19 are helped with oxygen therapy and antivirals like Lopinavir, Ritonavir, Ribavirin and Favipiravir are currently used. The essential way to control and reduce the human transmission of the virus is by implementing the precautionary steps and hygienic measures. Machine learning technologies were used by healthcare organizations and medical professionals all around the world to combat the pandemic by lowering the need for human intervention and enhancing the pandemic medication development process's screening, prediction, and forecasting capabilities. One method for finding the COVID-19 is X-ray imaging. We provide a supervised machine learning method in this study for identifying and categorizing COVID-19 infection from x-ray pictures. It is less necessary to manually label x-ray pictures when COVID-19 cases can be accurately detected and distinguished from non-COVID-19 instances. Performance metrics including Accuracy, Sensitivity, Specificity, F1-score, and Positive Predictive Value (PPV) are then used to assess the outcomes.

Keywords-COVID-19, machine learning, X-ray images, supervision

I. INTRODUCTION

The word corona was derived from the Latin word “corona” that means “crown”. They are large in size and are members of the corona viridine family. Corona viruses live in close proximity to the positive strand of RNA viruses. In the current situation the detection of coronavirus disease is the major challenge in the world; its rate of spread is very high. The number of persons with COVID-19 diagnoses has grown dramatically as the illness spreads around the globe.

The number of people infected are from 2 to 20 million with 250,000 deaths in the first four months of the pandemic. Because the symptoms of the disease are not shown immediately, it was very difficult to detect the infected patients. The bilateral distribution of ground glass opacities with or without consolidation in the posterior and peripheral lungs is the key characteristic of COVID-19.

All countries have made a serious effort to stop the COVID-19 virus from spreading quickly among people, quarantining more than 100 million individuals as a result. To diminish rapid spreading of the COVID-19 among people, all governments around the world have taken the severe effort by placing more than 100 millions of people under quarantine [21,26].

It is not possible to control the pandemic without the efforts of common people, government efforts to tackle the situation alone are not enough to control the virus. The virus needs to be identified promptly in order to let the infected person be isolated, receive treatment, find contacts, and send emergency alerts to other regular people earlier.[24].

II. LITERATURE REVIEW

Corona viruses live in close proximity to the positive strand of RNA viruses. In the current situation the detection of coronavirus disease is the major challenge in the world; its rate of spread is very high. The patients with COVID-19 are helped with oxygen therapy and antivirals like Lopinavir, Ritonavir, Ribavirin and Favipiravir are currently used. As the disease is spreading across the world the number of people diagnosed with the COVID-19 has increased in an exponential manner.

The number of people infected are from 2 to 20 million with 250,000 deaths in the first four months of the pandemic. Because the symptoms of the disease are not shown immediately, it was very difficult to detect the infected patients. The major feature of COVID 19 is a bilateral distribution of ground glass opacities in the posterior and peripheral lungs, with or without consolidation. All governments worldwide have made a significant effort to slow down the rapid spread of COVID-19 among individuals by quarantining more than 100 million people [4,9].

It is not possible to control the pandemic without the efforts of common people, government efforts to tackle the situation alone are not enough to control the virus. The virus needs to be identified promptly in order to let the infected person be isolated, receive treatment, find contacts, and send emergency alerts to other regular people earlier.[7].

The majority of the current techniques for the detection of COVID-19 and diseases associated with it make use of

2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 979-8-3503-6016-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICACITE60783.2024.10617174

cutting-edge technology and a hybrid model. In order to reduce errors in classifying diseases, advanced technology like Deep Neural Networks is utilized to build neural networks and modify the network parameters.

Convolutional neural networks, on the other hand, typically struggle to categories images that have some tilt or rotation. The training process will be laborious if the CNN has several layers and the computer's GPU isn't up to standard [18]. We propose a novel Kaze characteristics-based KNN classifier-based COVID-19 identification method.

An automated method for finding and describing 2D features at several scales is called KAZE. The KNN classifies based on distance and is very helpful for massive data categorization. Categorization problems can be effectively handled with this supervised machine learning method. To save the important time of the medical specialists and take advantage of KNN's high efficacy, it is advised to classify and evaluate COVID-19, especially in the epidemic zone [19,20]

The critical and useful modalities for diagnosing the stage of COVID-19 are the radiology and imaging. The diagnosis on the basis of appearances as the disease has progressed is focused by imaging research. Identifying infected patients from healthy individuals is crucial for stopping the spread of COVID-19. Because infected patients are not properly identified, secondary COVID-19 infections are primarily caused by this. Only when the condition is identified as soon as feasible can the right therapy be started with a high likelihood of success.

The reverse transcription- polymerase chain reaction (RT-PCR) tests are a manual diagnostic tool that can be used to stop the spread of COVID-

[19] It has not a high enough sensitivity rate to prevent the pandemic in an effective manner. The CT scan or x-ray is a valuable tool which provides the complement for the RT-PCR test and it is widely used by many hospitals. In some cases, with the normal CT result the confirmed positive in RT-PCR test however, these combinations are used to identify the patient during onset using the imaging results can be taken at a lower radiation and lower price.

[20] The medical images which are used in COVID-19 diagnosis are analyzed by the expert. Because of long working hours the doctors may experience fatigue and it may result in wrong diagnosis. Many lung problems are identical to COVID-19 in CT results. During the outbreak, the existence of severely lacking experienced medical staff and finding difficulties in identifying infected patients in time. Thus there is a high demand in the automatic diagnosis system to overcome these difficulties. [1,20]

III. Methodology

Preventing epidemics depends on properly and rapidly identifying compromised patients. Artificial intelligence technology has a big impact on disease identification and diagnosis in a quicker manner by extracting data from several reports and assisting in the analysis of DNA sequences for accurate treatment methods. Many AI companies have released AI-powered diagnosis tools that can help physicians with viral diagnosis and disease outbreak monitoring. XR Health has launched an interactive virtual telehealth support group in the United States to assist persons who are isolated as a result of the illness. This will make it possible for individuals with similar ailments to get in touch with one another and help from medical specialists. For instance, study into the virus's protein structure and the development of fresh medicinal strategies can be aided by Google Deep Mind's AI algorithms.

The implementation of AI systems for diagnosis, analysis, and therapy poses logistical issues due to dataset fluctuations, fresh population, and the harmful consequences of updates to medical results algorithms. When compared to pneumonia that was not COVID-19.

SOFTWARE

SPECIFICATIONMATRIX

LABORATORY

MATLAB (matrix laboratory) is the name of the numerical computing platform and the fourth programming language. It is able to manipulate matrices, visualize functions and data, construct algorithms, develop user interfaces, and communicate with programming languages such as C, C++, Java, and Fortran, using the MathWorks application MATLAB.

An extra toolbox in MATLAB, which is primarily intended for numerical computing but also has access to symbolized computational capabilities, uses the MuPAD symbolic engine. Simulink, a separate programme, adds prototype design for embedded and dynamic systems as well as graphic multi-domain simulations.

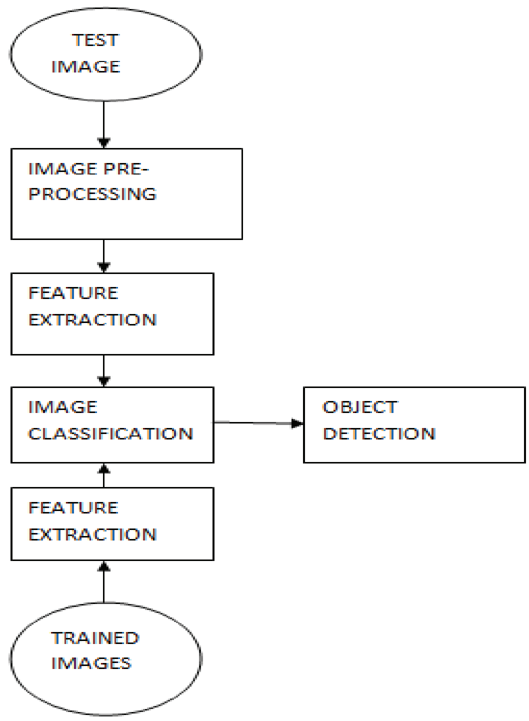


Fig. 1. Algorithm used in MATLAB

MODULE 1:

INPUT IMAGES

Here, we'll give the system instructions to accept the collected dataset as input and train the computer to tell COVID-19 images apart from non-COVID-19 images.

MODULE 2:

IMAGE PRE PROCESSING

Data preparation is the process of transforming raw data into something that can be used by a machine learning technique. It is the initial and most significant step in the process of creating a machine learning model. When working on a machine learning project, although it's not constantly possible that we are presented with the clean and prepared data. Additionally, you must format and clean your data every time you work with it. In order to do this, we need a data pre-treatment activity. To make the data clean and ready for a machine-learning model, pre-processing is required. This step also increases the precision and efficacy of the model.

Identifying incomplete data, encoding categorizing data, importing libraries, obtaining the dataset, etc. generating test and training sets from the dataset

MODULE 3:

FEATURE EXTRACTION

With the help of feature extraction, the amount of the equivalent information in the gathering of information is reduced. Finally, the data reduction quickens. while also allowing the model to be generated with less machine effort during the machine-learning process' learning and generalization phases.

When an investigation with a large number of variables uses a lot of memory and processing power, a classification algorithm may overfit to training instances and perform badly on new data. Feature extraction is a technique for developing approximations that get around these problems while still accurately describing the data. Many machine learning experts believe that properly tuned feature extraction is the key to effective model construction.

MODULE 4:

CLASSIFICATION

Performance measure for the developed model

In order to assess the model's performance on X-Ray images,

1. Accuracy = true predictions/total number of cases,
2. True positive or positive sensitivity (recall),
3. Particularity = real negative/negative,
4. The Positive Prediction Value (PPV) (Precision) is defined as True Positive/(True Positive + False Positive,
5. F1-score generates a balanced average result by combining precision and recall.

MODULE 5:

OBJECT DETECTION

Counting objects in a scene, identifying and tracking their exact locations, and accurately labelling them can all be done using detection and recognition.

IV. Results And Discussion

The feature extraction code trains the image as covid and non-covid and then it compare with the input image and gives the results.

A. INPUT IMAGE

An efficient way for finding novel COVID-19 on chest x-ray images that come from procedures termed radiography is automated detection.

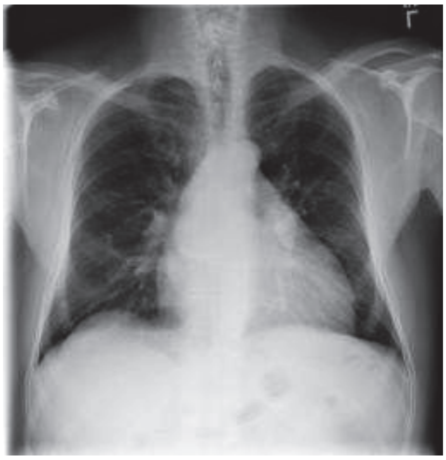


Fig. 2. Input image

B. RESIZED IMAGE

The input image will be given, and the image will be scaled for processing

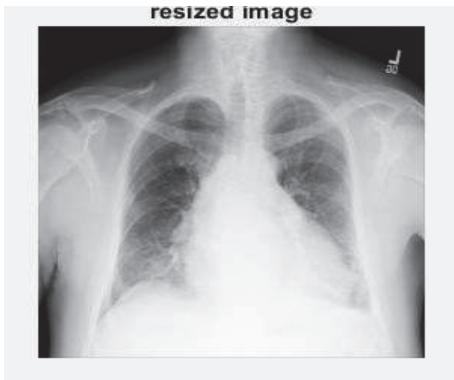


Fig. 3. Resized image

C. DOUBLED IMAGE

The only aspect of images that is still present is the essential grayscale. Whatever colour information is present, the result is always a pure black image. Red, green, and blue (RGB) three-pixel groups, commonly referred to as channels in digital imaging, are the building blocks of a digital image. However, we double the image for a float number as the output rather than using grey scaling.

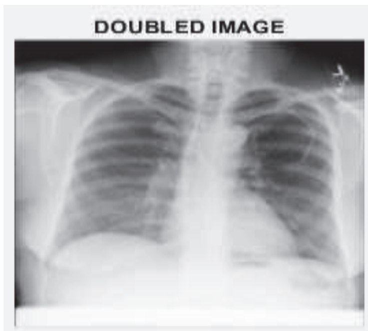


Fig. 4. Doubled image

D. KAZE FEATURE

Multi scale image processing is a crucial technique for computer vision applications. By screening the original image with the proper function as the time or scale rises, you can construct an image's scale space. KAZE is a nonlinear scale space-based technique for recognizing and describing two-dimensional features.

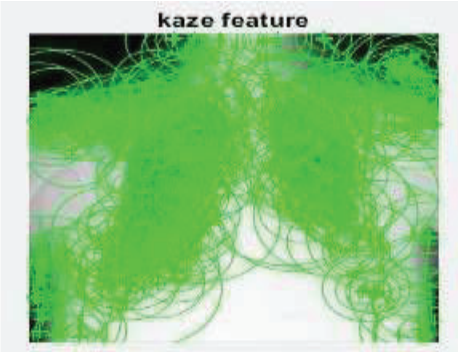


Fig. 5. KAZE feature image 1

output.5.1 KAZE feature image

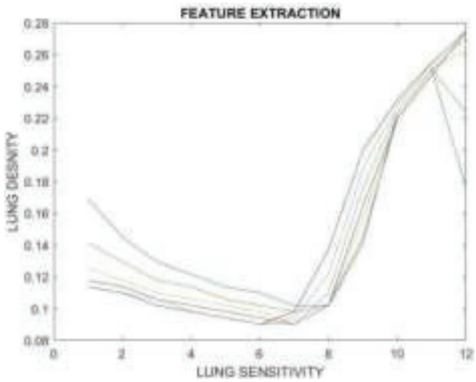


Fig. 6. Feature extraction graph

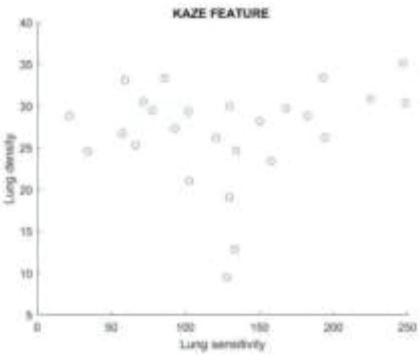


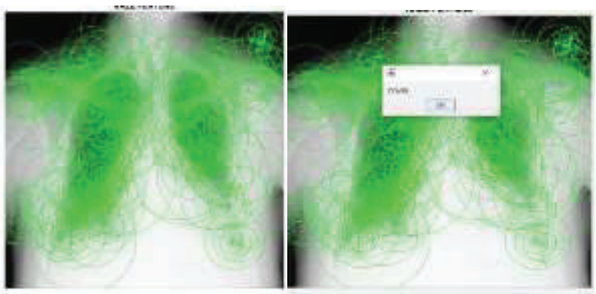
Fig. 7. KAZE feature graph

CLASSIFIED IMAGE

Fig. 8. Output

(a) (b)

(c)



(d) (e)

Fig. 9. (a)-input image ,(b)-resized image, (c)-gray converted image,(d)-kaze image, (e) classified image

(a)

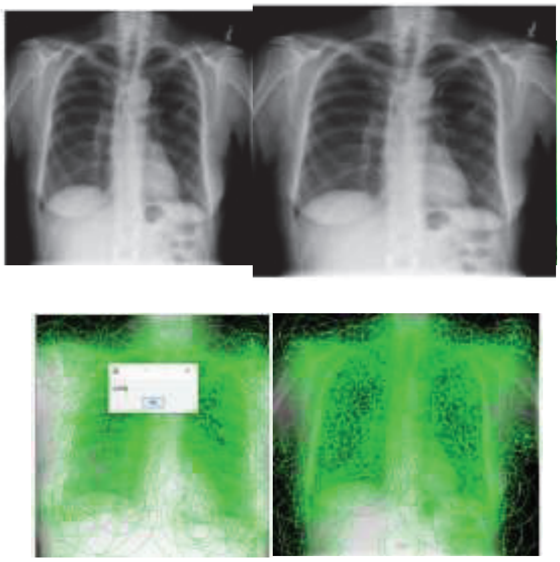


Fig. 10. (a)-input image, (b)-resized Image, (c)-gray converted image (d)-kaze image (e)classified image.

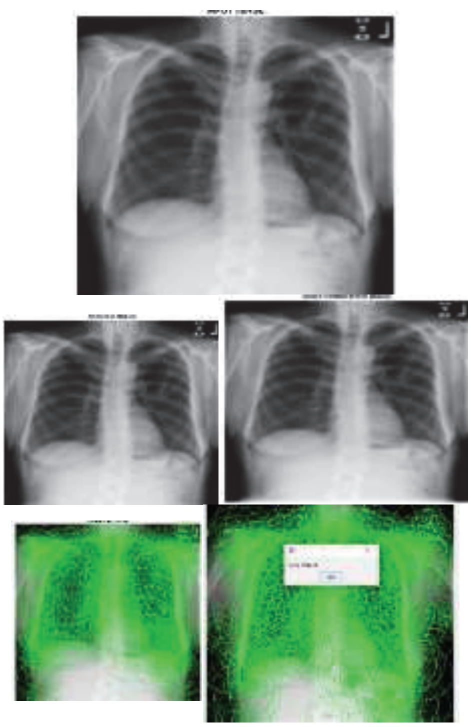


Fig. 11. (a)-input image ,(b)-resized image, (c)- gray converted image,(d)-kaze image,

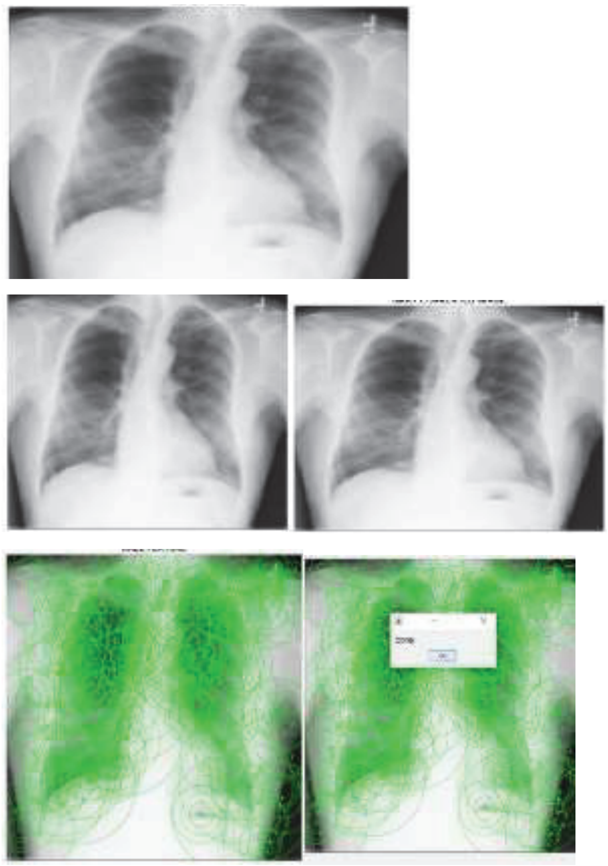


Fig. 12. (a)-input image ,(b)-resized image, (c)-gray converted image,(d)-kaze image,

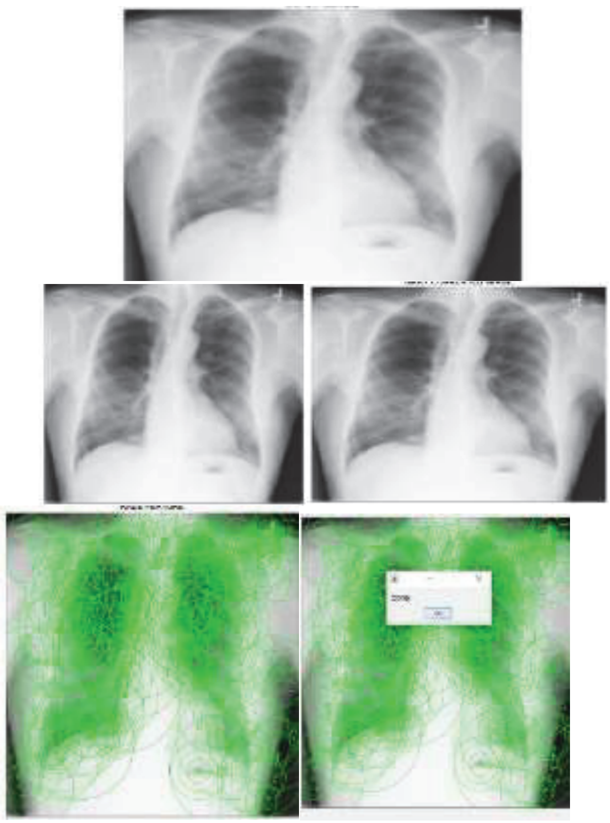


Fig. 13. (a)-input image ,(b)-resized image, (c)-gray converted image,(d)-kaze image,

OUTPUT:

The feature extraction code trains the image as covid and non- covid and then it compares with the input image and gives the results

V. CONCLUSION

The most commonly used procedures for COVID-19 diagnosis, notably RT-PCR and CT, have limitations and downsides such as lengthy processing periods and unacceptably high mistake rates. Because of the lack of information from the COVID-19 cases, most current works in the literature based on deep learning share these disadvantages. Although deep learning- based recognition approaches have performed at a cutting-edge level in computer vision, their performance declines quickly owing to data scarcity, which is the reality in this case. This research intends to address these constraints by presenting a robust and highly accurate COVID- 19 recognition method based on X-ray pictures The proposed approach makes use of machine learning algorithms like the K-nearest neighbour classifier. The KNN classifier generates superior results when combined with the kaze feature, achieving accuracy of 85.714%, sensitivity of 100%, specificity of 80%, and g mean of 87%. The COVID-19 will be distinguished from other respiratory disorders in the future by using a greater scale of data and qualities that can be explained. Additionally, we plan to use a variety of Deep Neural Network models to analyse the data and perform a comparison analysis on the ability to identify COVID-19 and related pneumonia infections. Furthermore, they create the most operationally effective system in terms of performance and memory.

REFERENCES

[1] [Ouassou H, Kharchoufa L, Bouhrim M, Daoudi NE, Imtara H, Bencheikh N, ELbouzidi A, Bnouham M. The Pathogenesis of Coronavirus Disease 2019 (COVID-19): Evaluation and Prevention. J Immunol Res. 2020 Jul 10;2020:1357983. doi: 10.1155/2020/1357983. PMID: 32671115; PMCID: PMC7352127.

[2] Lipták P, Banovcin P, Rosoľanka R, Prokopič M, Kocan I, Žiačiková I, Uhrik P, Grendar M, Hyrdel R. A machine learning approach for identification of gastrointestinal predictors for the risk of COVID-19 related hospitalization. PeerJ. 2022 Mar 21;10:e13124. doi: 10.7717/peerj.13124. PMID: 35341062; PMCID: PMC8944335

[3] Rahman MM, Khatun F, Uzzaman A, Sami SI, Bhuiyan MA, Kiong TS. A Comprehensive Study of Artificial Intelligence and Machine Learning Approaches in Confronting the Coronavirus (COVID-19) Pandemic. Int J Health Serv. 2021 Oct;51(4):446-461. doi: 10.1177/00207314211017469. Epub 2021 May 17. PMID: 33999732.

[4] Rahman, M.M., Khatun, F., Uzzaman, A., Sami, S.I., Bhuiyan, M.A.A. and Kiong, T.S., 2021. “A comprehensive study of artificial intelligence and machine learning approaches in confronting the coronavirus (COVID-19) pandemic”.International Journal of Health Services, 51(4), pp.446-461

[5] Alamrouni, A., Aslanova, F., Mati, S., Maccido, H.S., Jibril, A., Usman, A.G. and Abba, S.I., 2022. “Multi- Regional Modeling of Cumulative COVID-19 Cases Integrated with Environmental Forest Knowledge Estimation: A Deep Learning Ensemble Approach”. International Journal of Environmental Research and Public Health, 19(2), p.738.

[6] Bai, H.X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J.W., Tran, T.M.L., Pan, I., Shi, L.B.,Wang, D.C., Mei,J. and Jiang, X.L., 2020.“Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT”. Radiology, 296(2), pp.E46-E54.

[7] Alamrouni, A.; Aslanova, F.; Mati, S.; Maccido, H.S.; Jibril, A.A.; Usman, A.G.; Abba, S.I. Multi-Regional Modeling of Cumulative COVID-19 Cases Integrated with Environmental Forest Knowledge Estimation: A Deep Learning Ensemble Approach. Int. J. Environ. Res. Public Health 2022, 19,738.https://doi.org/10.3390/ijerph19020738

[8] Kishore Kanna, R., Yamuna Devi, K., Gomalavalli, R., & Ambikapathy, A. (2024). Smart Detection and Removal of Artifacts in Cognitive Signals Using Biomedical Signal Intelligence Applications. In V. Dutt, A. Kumar, S. Ahuja, A. Baliyan, & N. Vyas (Eds.), Quantum Innovations at the Nexus of Biomedical Intelligence (pp. 223-244). IGI Global. https://doi.org/10.4018/979-8-3693-1479-1.ch013.

[9] An automated and fast system to identify COVID‐19 from X‐ray radiograph of the chest using image processing and machine Chien, H.J., Chuang, C.C., Chen,C.Y. and Klette, R., 2016, November.

[10] SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. In 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE. Bai, H.X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J.W., Tran, T.M.L., Pan, I., Shi, L.B., Wang, D.C., Mei, J. and Jiang, X.L., 2020.

[11] Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. Radiology, 296(2), pp.E46-E54. Perumal, R. and Kaladevi,A.C., 2021.

[12] An Effective Diagnostic System for Identifying Covid-19 and Pneumonia Diseases Using Machine Learning Algorithms. Annals of the Romanian Society for Cell Biology, pp.9078- 9086. Shaban, W.M., Rabie, A.H., Saleh, A.I. and Abo-Elsoud, M.A., 2020.

[13] A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowledge-Based Systems, 205, p.106270. Arslan, H. and Arslan, H., 2021.

[14] [A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowledge-Based Systems, 205, p.106270. Arslan, H. and Arslan, H., 2021.Ramya Perumal1 A. C. Kaladevi2. “An Effective Diagnostic System forIdentifyiing Covid-19 and Pneumonia Diseases Using Machine Learning Algorithms”. 8th math 2001.

[15] Alamrouni, A., Aslanova, F., Mati, S., Maccido, H.S., Jibril, A., Usman, A.G. and Abba, S.I., 2022. Multi-Regional Modeling of Cumulative

[16] Artificial Intelligence Analysis of Deep Learning Techniques for COVID-19 Detection Kawal Arora, Dr Ankur Singh Bist, Saksham Chaurasia and Roshan Prakash Signy Advanced Technologies, INDIA Khan, M.A., 2021.

[17] An automated and fast system to identify COVID19 from X‐ray radiograph of the chest using image processing and machine Chien, H.J., Chuang, C.C., Chen,C.Y. and Klette, R., 2016, November.

[18] SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. In 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE. Bai, H.X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J.W., Tran, T.M.L., Pan, I., Shi, L.B., Wang, D.C., Mei, J. and Jiang, X.L., 2020.

[19] Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. Radiology, 296(2), pp.E46-E54. Perumal, R. and Kaladevi,A.C., 2021.

[20] An Effective Diagnostic System for Identifying Covid-19 and Pneumonia Diseases Using Machine Learning Algorithms. Annals of the Romanian Society for Cell Biology, pp.9078- 9086. Shaban, W.M., Rabie, A.H., Saleh, A.I. and Abo-Elsoud, M.A., 2020.

[21] A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowledge-Based Systems, 205, p.106270. Arslan, H. and Arslan, H., 2021.

[22] Kanna, R. K., Pradhan, N. R., Panigrahi, B. S., Basa, S. S., & Mohanty, S. (2024). Smart Assist System Module for Paralysed Patient Using IoT Application. EAI Endorsed Transactions on Internet of Things, 10.