**Research Paper**

COVID-19 and Pneumonia Detection using Chest X-ray Imaging with LIME and Grad-CAM

Abstract

Due to the fast spread of COVID-19 and the high cases of pneumonia, accurate and understandable diagnostic tools are essential for early diagnosis and treatment. Although it helps to identify both ,chest X-ray imaging is limited in interpretability by the complexity of deep learning models.ConvNets are the significant deep learning models that promise to show significant success in automatically identifying a through chest X-ray images. Moreover, the model concentration on important areas for clinical care is enhanced through guided Grad-CAM as well as attention mechanisms, which help to determine the most significant features that influence COVID-19 and pneumonia classification. The proposed method is evaluated using a dataset comprising chest X-rays of normal lungs, viral pneumonia, and COVID-19 patients and reveals that it not only provides excellent results in terms of detection but also makes transparent interpretative decisions needed in clinical situations. This research will contribute to building deep learning models that are embraced in healthcare applications.

1. Introduction

The current coronavirus pandemic and frequent development of pneumonia have made effective diagnostic tools critically important. It is a major way that chest X-ray imaging can diagnose both COVID-19 and pneumonia due to its speed as well as ease of access. However, the traditional x-ray interrogation, which heavily depends on the radiologist’s expertise, might be time-consuming and subjectively interpretable, especially during world pandemics.Deep learning models, specifically convolutional neural networks , have shown great potential for automating the process of detection of COVID-19 and pneumonia from chest X-rays. Although these models exhibit high levels of accuracy, they are often criticized because their decision-making process is not transparent enough, thus referred to as a black box in nature. This lack of transparency may cause doubt about AI-based diagnoses’ credibility and safety, particularly in the most critical health care settings.

We will investigate explainability techniques, which apply methods like LIME and Grad-CAM, and integrate them into a CNN-based model for detection to address the aforementioned challenges.LIME approximates the model’s local behavior near specific predictions, while Grad-CAM visually highlights regions within an input image that are significant to the model’s decision. We also improve Grad-CAM by using guided backpropagation and adding attention mechanisms to make it focus more on relevant areas only in X-ray images.Our method strives to balance interpretability with diagnostic accuracy so as to ensure that predictions made by the model are not only true but can be understood by physicians. Our aim is to leverage these explainability techniques to develop trust in AI systems and assist radiologists in making decisions regarding COVID-19 and pneumonia. This article gives an extensive description of how well the proposed technique performs and its implications for clinical decision-making.

2.Literature Review

In their pioneering work[1] , “Phenotyping COVID-19 respiratory failure in spontaneously breathing patients with AI on lung’s CT-scan, investigate the application of artificial intelligence to examine lung CT scans in COVID-19 respiratory failure patients. This research was conducted as a multicenter observational cohort study involving 559 participants who were subjected to early lung CT scanning within seven days of admission. The group used deep learning approaches and applied these techniques to analyze both quantitative and qualitative aspects of the CT images as well as clinical and laboratory data. By means of latent class analysis (LCA), two subphenotypes for COVID-19 were distinguished by this study. Subphenotype 1 had higher levels of inflammatory biomarkers, worsened hypoxaemia, and more dense areas in the lungs, while Subphenotype 2 had an increased proportion of ground-glass opacities. These results highlight how AI-driven imaging can provide important insights into understanding and treating respiratory failure due to COVID-19 by identifying subtypes that are associated with diverse clinical outcomes.

In a systematic literature review [2], the authors examine the deep learning function in diagnosing pneumonia, which amounted to 2.5 million deaths all over the world the previous year, according to WHO.This review outlines that deep learning is better than traditional machine learning methods, especially concerning automatic feature extraction and performance outcomes. The authors conduct an elaborate study on several different types of DL models, such as CNNs, pre-trained models, and ensemble methods, whereby they take their time in discussing architectures and operational procedures in detail. The work also analyzes the effectiveness of these models when it comes to addressing problems within the medical domain by providing a detailed discussion of performance metrics, hyperparameters, and fine-tuning. Additionally, the paper reviews robust ensemble models that perform well in detecting pneumonia. In addition, they explore existing research gaps and recommend some potential answers meant to make sure people have more knowledge about deep learning applications in pneumonia diagnosis as well as treatment.

The study [3] presents a method of diagnosing chest diseases through chest X-ray (CXR) images employing machine learning. This study suggests an ensemble learning system that combines Random Forest, Support Vector Machine, and XGBoost as examples, amongst others, in a bid to offer accurate and swift diagnoses due to the COVID-19 pandemic. There are three stages in this process: CXR image preprocessing, histogram of oriented gradients (HOG) feature extraction, and local binary patterns (LBP), as well as the training classifiers. The ensemble model was classified with 98% accuracy, which was much higher than that of single machine learning algorithms. In this way, ensemble learning has been seen to be effective in medical image analysis; it can therefore be used as a good tool for detecting

COVID-19, tuberculosis, and pneumonia. According to this study, the combination of various machine learning models may enhance diagnostic accuracy, thereby shedding light on how AI can be implemented in healthcare.

The main aim of this article [4] is to make interpretable AI models in medical diagnostics, hence the issue of dark decision making within CNNs. SHapley Additive exPlanations and Gradient-weighted Class Activation Mapping are utilized as explainable AI methodologies which allow model predictions to be explained. For this purpose, pretrained models from NIH were used, transfer learning was done using datasets from Taichung Veterans General Hospital and VinDr to achieve 92.14% and 93.29% accuracies respectively. Besides interpretability, it makes the model adaptable for different clinical environments with other data sets too. Explainable AI is important because it can be trusted by medical experts who can then improve on it continuously through feedback thereby representing a significant step forward in medical image analysis.

A smart system to read lung diseases from chest X-rays using explainable AI (XAI) shows up in [5]. The research presents a CNN-based method with the ResNet50 network to spot tuberculosis, oedema, nodules, and pneumonia, including COVID-19. This approach uses LIME to explain classification results by highlighting key features on X-rays. The sorting gets more precise, and the guesses become sharper, reaching an accuracy of 93% to 97% across different datasets. This addresses a big problem with deep learning models, which often work like black boxes that don't explain how they make decisions. Also, it brings clarity, helping radiologists understand and check the logic behind their models' choices. This makes AI easier to grasp in medical diagnostics, turning it into a key tool to spot diseases early and support clinical decisions.

Paper [6] proposed an explainable AI-based system in the paper [6] mainly focusing on how to detect and classify COVID-19 by using chest X-ray images. It proposed a multi-input transfer learning-based approach through the usage of a COVID-Net fuzzy CNN to help improve the detection accuracy and interpretability even deeper. It merges a number of convolutional neural networks and allows them to use fuzzy logic with edge images, which are useful in extracting features from CXR images. It achieves the best quality with an area under the curve of 1.0 as well as accuracy, precision, and recall metrics of 0.97 by using transfer learning and pre-trained models. The explanation techniques shed some light on the most important factors related to COVID-19, hence helping doctors understand more and screen. When fine-tuned on IoT devices, it retains an accuracy of 0.95, proving it to be reliable and useful. This work not only advances COVID-19 detection with easier-to-understand models, but it also increases the requirement for diagnostic tools in a pandemic era to be fast and handy.

A research study [7] clearly defined an AI-based system which helps in identifying COVID-19, viral pneumonia, and lung opacity using chest X-rays. The introduced method is a transfer learning approach, using explainable AI methods within the framework of Xception, VGG19, and ResNet50 models. It achieves high accuracy in classifying COVID-19 vs normal cases with very good performance in multi-class classification. Xception is most precise and interpretable compared to other models. It provides clear visual explanations of its predictions through GradCAM. The explainable AI gives transparency to the model. This helps the doctor understand how the model makes a decision. There is a central essence of the availability of robust diagnostic tools which are easy to interpret and manage in order to take precautionary steps against COVID-19 when a global health crisis arises.

Researchers have built an explainable AI for detecting COVID-19 in CT scans and chest X-rays in a study [8]. Deep learning in this technique combines with machine-learning classification tools. A CNN pulls out deep features from the images, and these are then fed through a group of classifiers, including GNB, SVM, DT, LR, KNN, and RF. It performed really well with its accuracy, precision, and recall: 98.5%, 99%, and 99%, respectively. Basically, the system incorporates explainable AI techniques like Grad-CAM and t-SNE that guarantee the sanity check on the model's output and make sense out of it. These have distinct, transparent processes that give insight into how the detection process is done, enhancing a doctor's ability to understand and rely on the predictions made by an automated system.

The paper [9] presents an understandable approach to the diagnosis of COVID-19 and pneumonia in which transfer learning is combined with discriminant correlation analysis. It builds image representations from lung CT scans using two deep networks: ResNet-18 and ResNet-50. Then, it enhances these image data by fusing the features extracted from these networks using discriminant correlation analysis. It trains three randomized neural networks for the classification of images against these improved features. All of the predictions from these networks are then combined to improve classification performance. Results in five-fold cross-validation tests show this proposed method outperforming other algorithms in diagnosis accuracy for COVID-19.

The paper [10] proposes an ensemble XAI algorithm for the diagnosis of pneumonia and COVID-19 respiratory infections. Guiding their research in this line, the goal of the authors was to integrate SHAP and Grad-CAM++ methods into a new image explainability technique, enhancing the interpretability of deep learning models for the prediction of patient mortality risk. To this end, they give a general framework that merges quantitative and qualitative metrics of evaluation for the assessment of these XAI techniques. A newly introduced ensemble XAI methodology incorporating explanation techniques yielded a high trust score from radiologists with a mean of 70.2%. Hence, the study underlines the importance of explainability in AI for healthcare by enhancing trust among clinicians and promoting better decision-making. Quantitative results bring into light the fact that most of the metrics are dominated by Ensemble XAI; hence, an interpretation or reliability of results in a clinical setting can be attained.

This paper [11] presents a deep learning-based system for the detection of tuberculosis, trying to solve the challenge of vast dark areas in CXRs that confuse diagnostic algorithms. The methodology presented in this work will be based on high-quality segmentation networks for the extraction of lung regions from CXRs and significantly improve the performance of CNNs. On this dataset, EfficientNetB3 created an accuracy of 99.1 percent and a ROC score of 99.9 percent, clearly showing the effectiveness of the approach. Indeed, much emphasis is put on the use of segmented images as opposed to raw CXRs to attain better diagnostic results. This means this AI is explainable by visualising the regions of CXR infected with TB for important insights from the model's predictions to assist diagnostic processes by medical professionals. In the process, it reviews some of the CNN architectures, like ResNet, Xception, and EfficientNet, for image classification tasks and identifies their strengths. Residual learning by ResNet mitigates the degradation problems, and depth-wise separable convolutions by Xception improve the performance. Basically, EfficientNet balances between the width and depth of the network to achieve high accuracy with fewer parameters.

This work [12] illustrates the potential of XAI methods applied to COVID-19 diagnosis based on variables of blood tests, thus underscoring the increasing role that interpretability plays within the medical AI paradigm. The authors used logistic regression and an explainable boosting machine as glass-box models and random forest and a support vector machine as black-box approaches in predicting COVID-19 diagnosis. The study found some important features towards the prediction of COVID-19 to be eosinophils and leukocytes; the best model had an AUC of 0.87. It used SHAP to explain the black-box models, returning the importance of features and explanations of model decisions. This research thus builds a case for why, in health, explainability is required—it increases trust and brings transparency into predictions. It thus leveraged a larger dataset of suspected cases from the Hospital Albert Einstein and presented a method that especially gives prominence to the explainable boosting machine in improving interpretability. The paper places this work within the context of machine learning applications for COVID-19 diagnosis, emphasising the shortcomings of imaging-only approaches and the fact that reproducibility and transparency should be preconditions to be met by any AI model.

Coupled with recent advances in the field of IPF, there has been an enhanced requirement for novel diagnostic biomarkers. Machine learning, on the other hand, enunciates specifically in functions that have high-dimensional complex data; the traditional approaches to research are complementary. Successful studies have exploited gene expression profiles in classifying IPF

and non-IPF states using ensemble learning models like XGBoost. It has allowed for the identification of key gene features with very high diagnostic potential by integrating Shapley values for model explanation. [13] This approach offers enhanced specificity and accuracy of biomarker detection but also a more interpretable framework toward the understanding of the mechanisms of disease. The biomarkers also turned out to be quite robust and generalizable according to the validation against independent datasets, making them very promising candidates for clinical application in the near future.

The increment in the interest of XAI particularly for medical diagnosis has been growing over the past few years. This is a result of COVID-19 pandemic-driven demands. Moreover, the study by Kirboğa et al., which was on predictors of troponin levels, showed that XAI is important In fact this would be very crucial in instances where the biomarker is found to be playing a central role in jointly assessing cardiac injuries related to patients with COVID-19. With regard to this study SHapley Additive exPlanations provide an interpretable model that is clear about what affects the troponins. Their work has shown many potential application areas for XAI algorithms with respect to decision making using different machine learning methods. For instance, deep infusion of the competence model with improved competence spanning accuracy as well as explain-ability CVD22 and enhanced patient driven explanations during the global pandemic phase. More than ever before, this paper elucidates why interpretable AI models are more invaluable than ever before in healthcare amidst complex scenarios associated with management of COVID-19; thus, making it imperative for explainable artificial intelligence to be included in medical diagnosis especially after COVID-19 breakout.

The increase in interest in XAI, particularly for medical diagnosis, has been growing over the past few years. This is a result of COVID-19 pandemic-driven demands. Moreover, the study by Kirboğa et al., which was on predictors of troponin levels, showed that XAI is important. In fact, this would be very crucial in instances where the biomarker is found to play a central role in jointly assessing cardiac injuries related to patients with COVID-19. With regard to this [14] study, Shapley Additive Explanations provide an interpretable model that is clear about what affects the troponins. Their work has shown many potential application areas for XAI algorithms with respect to decision-making using different machine learning methods. For instance, a deep infusion of the competence model with improved competence spanning accuracy as well as explainability (CVD22 and enhanced patient-driven explanations during the global pandemic phase. More than ever before, this paper elucidates why interpretable AI models are more invaluable than ever before in healthcare amidst complex scenarios associated with the management of COVID-19, thus making it imperative for explainable artificial intelligence to be included in medical diagnosis, especially after the COVID-19 breakout.

The paper [15] elaborates on the need for chest radiographs in medical fields, more so after the recent pandemic outbreak of COVID-19, and how advancements in DL have massively improved medical image classification.The authors have done a very elaborate state-of-the-art analysis of the current DL models within the application domain of chest X-ray image classification. Besides this, the authors also reviewed publicly available datasets applied in related studies and provided trends and challenges associated with the domain under study. The review calls for a systematic approach towards deciphering the efficacy of deep learning models in medical diagnostics concerning respiratory diseases. The authors differentiate this review from many of the other surveys available in the literature by stating, "Although a number of surveys are available on DL models and datasets, most of these are not very in-depth in aspects like calculation of loss, optimisation, and evaluation metrics." This survey therefore looks to fill those gaps by giving a more detailed exposition of these various aspects and also providing guidelines for the selection of proper models and techniques.The review further helps readers by giving summaries of the DL model architectures, optimisers, loss functions, and reports of performance on different models with various datasets. In this line, the authors wrap up this survey by elaborating on the open challenges in the field and proposing possible future research lines that make this review indispensable to developers and researchers dealing with chest X-ray image analysis using DL.

3.Proposed Work

The COVID-19 pandemic has underscored the requirement for speedy and accurate diagnostic tools for respiratory diseases, many of which radiographically often present similarly to pneumonia. This work developed an approach for deep learning-based automated detection and differentiation of COVID-19 from viral pneumonia and normal cases by chest X-ray image analysis. This has also incorporated explainable AI techniques like LIME and Grad-CAM to emphasize model interpretability beyond high accuracy.

The primary objectives of this work are:

1. A CNN is created to identify COVID-19, viral pneumonia as well as normal chest X-ray images with high accuracy.

2. Build in explainability techniques, such as LIME and Grad-CAM, into the model to make its predictions transparent and establish clinical trust.

3. The effectiveness of such interpretability techniques in providing meaningful insights regarding the model's decision-making process will also be evaluated.

4. Propose a new score, TEAM, which combines model confidence with human-centred interpretability metrics to assess the reliability of predictions.

**Methodology**

**Data Collection and Preprocessing**

The proposed model would be trained using a dataset of labeled X-ray images, which can be classified into 3 classes which are COVID-19, viral pneumonia, and normal cases. In order to generalize the model well and avoid overfitting, the dataset would be preprocessed by normalizing, resizing, and augmenting it.



Figure 1.a Covid train data sample

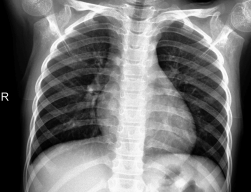


Figure 1.b Normal train data sample



Figure 1.c Pneumonia train data sample

**Model Architecture**

In this paper we have developed a DL model which is primarily based on VGG16 pre-trained on the ImageNet dataset whose parameter are further fine tuned to classify X-ray images. The final layers of the model will be changed to have the three output classes. Furthermore, another simpler CNN model will be built and compared to enable analysis of trade-offs between complexity and performance.

**Model Training and Evaluation**

The model will be trained using categorical cross-entropy loss and an Adam optimizer. Then, accuracy, precision, recall, the F1 score, and a confusion matrix analysis in the test set are used in order to evaluate the performances of the models.

**Explainability Techniques**

The following XAI techniques will be implemented to enhance the interpretability of the model's predictions:

● **LIME:** This will create a local explanation using LIME, in which we perturb the input image at the same time keep noting the changes in the model’s output. That should help understand which regions of the X-ray contribute most to the classification decision.

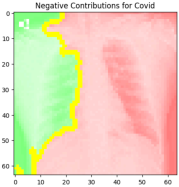
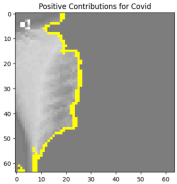
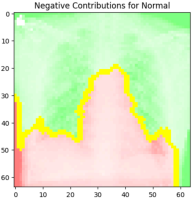
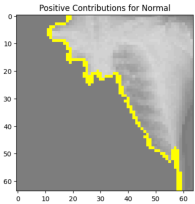


Figure 2a Lime for Covid

Figure 2b Lime for Normal

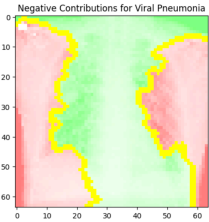
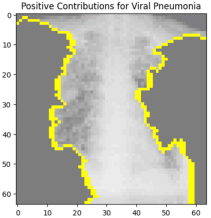
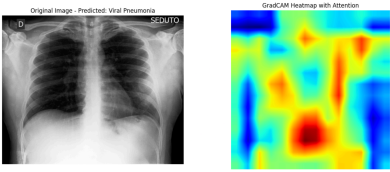
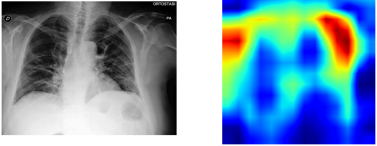
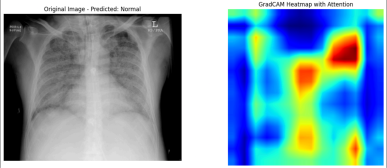


Figure 2c Lime for Pneumonia

● **Grad-CAM:** Grad-CAM will be used to create visual explanations of the strong influence of regions in the X-ray, giving insights into spatial localisation for relevant features regarding the model's output.

Figure 3a Grad-CAM Pneumonia

Figure 3b Grad-CAM Covid

Figure 3c Grad-CAM Normal

**TWA Metrices**

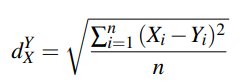
**The need for trust:** In complex systems, particularly those involving multiple agents or decentralized decision-making, trust plays a crucial role in ensuring the reliability and acceptance of AI-driven explanations. Addressing the trustworthiness of information sources and the entities generating explanations is paramount for achieving transparent and interpretable AI.

**4. Assessment of Explanations**

In this section, we introduce the Trustworthy Explainability Acceptance (TwA) metric to evaluate the alignment between the explanations provided by the system and those of experts. This metric is crucial in determining the reliability of explanations from XAI methods: LIME and Grad-CAM. The goal of this assessment is to reduce the subjectivity in expert evaluation by calculating the distance between the system’s explanation and expert reasoning.

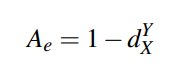
The **explainability acceptance** is based on the closeness of the explanations provided by the system to the experts’ own explanations. More specifically, the **explainability distance** between the system's explanation and the expert's is calculated using the **Euclidean distance formula**. Each explanation is treated as an n-dimensional vector, where each dimension represents an attribute of the explanation.

Given two explanations X and Y, the **explainability distance** dY ​ is computed using Eq. (5):



where Xi and Yi represent the values of the ith dimension for expert’s and the system's explanations, respectively. The distance dY lies in the range [0,1], where 1 represents the maximum distance.

The Ae denotes **explainability acceptance,**​ by expert e for the system’s explanation is then computed using Eq. (6):

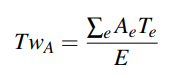


The closer the two explanations are, the higher the acceptance by the expert, which lies between [0,1], where 1 is the highest acceptance.

**4.1 Aggregation of Expert Judgments**

To minimize subjectivity and bias, a group of experts evaluates the system's explanations. The **Trustworthy Explainability Acceptance (TwA)** is calculated by aggregating the acceptances of various experts. Each expert’s acceptance is weighted by their **trust value** Te ​, which is calculated based on confidence and impression, as detailed in Section 2.2.

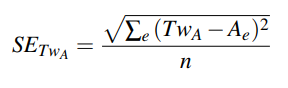
The aggregated TwA for the system is given by Eq. (7):



where E denotes total number of experts involved in the evaluation.

**4.2 Confidence in the Trustworthy Explainability Acceptance**

The confidence of the measured **TwA** is calculated using the **Standard Error (SE)** of the trust-weighted explainability acceptance, as shown in Eq. (8):



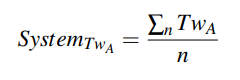
where n denotes the number of samples. The confidence in TwA, denoted as cTwA​, is computed using Eq. (9):



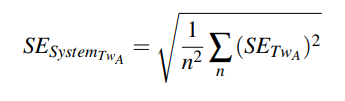
This tuple (TwA, cTwA​) provides a robust metric for evaluating the trustworthiness and confidence in the system's explanations.

**4.3 Aggregation of System-Level Explainability Acceptance**

When the system is evaluated across multiple samples, the overall **system-level trustworthy explainability acceptance** is calculated by averaging the TwA values across all samples, as shown in Eq. (10):



Finally, the standard error for the system across all sample measurements is calculated using Eq. (11):



This method provides a comprehensive evaluation framework for assessing the quality of explanations generated by different XAI methods. By using the TwA metric, we ensure that the system’s explainability is quantitatively aligned with expert reasoning, enhancing the trustworthiness of the model’s decisions.

**Visualization and Reporting**

It is proposed to present LIME and Grad-CAM detailed visualizations for a diversified set of test images, along with their qualitative analysis regarding the relevance and consistency of the highlighted features against clinical knowledge. Also, TEAM score reporting in each case will provide an all-around assessment of the trustworthiness of the prediction.

5.Conclusion

| Ref  No. | Model/Framework | ML model Accuracy (in %) |
| --- | --- | --- |
| [1] | **Explainable AI-based Interpretable**  **Classification for**  **Pneumonia** | **\_**  **Explainable AI-based**  **Interpretable Classification for**  **Pneumonia**  **DCNN.** |
| [2] | **Explainable AI-Enabled CNN Framework** using ResNet50 and LIME for classifying and explaining pulmonary diseases from chest radiographs. | **Explainable AI-Enabled CNN\_** |
| [3] | **Explainable AI-Enabled ResNet50 with LIME** | **LIME-ResNet50\_** |
| [4] | **CNN + Ensemble (GNB, SVM, DT, LR, RF) with Grad-CAM and t-SNE** | **\_**  **Gaussian Naive Bayes (GNB)**  **Support Vector Machine (SVM)**  **Decision Tree (DT)**  **Logistic Regression (LR)**  **K-Nearest Neighbour (KNN)**  **Random Forest (RF)** |
| [5] | **Explainable AI System for Automated**  **COVID-19 Diagnosis** | **Gradient Boosting Decision Trees**  **86.4%**  **(GBDT) with LIME** |
| [6] | **Explainable Machine Learning for COVID-19 Pneumonia Classification** | **XGBoost with SHAP 82%** |
| [7] | **Multiclass Deep Learning Algorithm for Lung Disease Detection** | **DenseNet with Transfer**  **94% AUC**  **Learning** |
| [8] | **Systematic Review of AI for COVID-19 Detection** | **CNN-based models 70-99%** |
| [9] | **AI-Assisted Diagnosis for COVID-19 Using Chest CT Images** | **Joint CNN Model 86% AUC** |

| [10] | **AI for COVID-19 Diagnosis Using Chest X-Rays** | **Ensemble of Bagged Trees 97%** |
| --- | --- | --- |
| [11] | **Transparent and**  **Trustworthy**  **Interpretation of**  **COVID-19 Detection Using AI** | **EfficientNet and DenseNet 71-99%** |
| [12] | **Diagnosis of COVID-19 Pneumonia Using Novel Deep Learning**  **Architecture** | **AlexNet Enhanced with Batch**  **90%**  **Normalisation** |
| [13] | **A Tree-Based Explainable AI Model for Early Detection of COVID-19** | **Decision Fusion Approach 86%** |
| 14.  [15] | **AI-Assisted Classification of COVID-19 and Non-COVID Pneumonia Using Chest X-Ray**  **AI-Based Multiclass Classification of Lung Conditions Using Chest X-Rays** | **AutoML Cloud Vision 92% VGG19 + Hybrid Ensemble 80.65%** |

According to the research, merging deep learning with interpretability techniques such as LIME and Grad-CAM, guided backpropagation, and attention mechanisms is helpful in using X-ray images of the chest to detect COVID-19 and pneumonia features. The transparency of this model has significantly improved as a result of these methods’ integration, which provides insight into the decision-making process while maintaining high diagnostic accuracy. For one thing, our findings indicate that it is not only making it more interpretable but also helping clinicians find important areas on an X-ray image so that they can make more accurate diagnoses for both COVID-19 and pneumonia. Its success makes it possible to introduce its application in real-world clinical practice settings where there is a need to strike a balance between accuracy and interpretability. The requirement for models that are transparent as well as trustworthy has never been higher than now, when healthcare increasingly depends on AI. Following these steps involves further implementation of the approaches used here in other medical imaging tasks.This study aims at bridging a gap between innovative AI methodologies and their clinical employment so that they appear effective and understandable.

Citation:

[1] Rezoagli, E., Xin, Y., Signori, D. *et al.* Phenotyping COVID-19 respiratory failure in spontaneously breathing patients with AI on lung CT-scan. *Crit Care* **28**, 263 (2024). https://doi.org/10.1186/s13054-024-05046-3

[2]Sharma, S., Guleria, K. A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images. *Multimed Tools Appl* **83**, 24101–24151 (2024). https://doi.org/10.1007/s11042-023-16419-1

[3] Ul Amin, S., Taj, S., Hussain, A., & Seo, S. (2024). An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach. \*Biomedical Signal Processing and Control\*, \*87\*(Part B), 105408. https://doi.org/10.1016/j.bspc.2023.105408

[4] R. -K. Sheu, M. S. Pardeshi, K. -C. Pai, L. -C. Chen, C. -L. Wu and W. -C. Chen, "Interpretable Classification of Pneumonia Infection Using eXplainable AI (XAI-ICP)," in *IEEE Access*, vol. 11, pp. 28896-28919, 2023, doi: 10.1109/ACCESS.2023.3255403. keywords: {Pulmonary diseases;Transfer learning;Artificial intelligence;Solid modeling;Medical diagnostic imaging;Convolutional neural networks;Data models;Medical XAI;XAI pnuemonia;transfer learning;pneumonia infection},

[5] Naz Z, Khan MUG, Saba T, Rehman A, Nobanee H, Bahaj SA. An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs. *Cancers*. 2023; 15(1):314. https://doi.org/10.3390/cancers15010314

[6]Qinhua Hu, Francisco Nauber B. Gois, Rafael Costa, Lijuan Zhang, Ling Yin, Naercio Magaia, Victor Hugo C. de Albuquerque, Explainable artificial intelligence-based edge fuzzy images for COVID-19 detection and identification, Applied Soft Computing, Volume 123, 2022, 108966, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2022.108966. (https://www.sciencedirect.com/science/article/pii/S1568494622003064)

[7]Islam MN, Alam MGR, Apon TS, Uddin MZ, Allheeib N, Menshawi A, Hassan MM. Interpretable Differential Diagnosis of Non-COVID Viral Pneumonia, Lung Opacity and COVID-19 Using Tuned Transfer Learning and Explainable AI. *Healthcare*. 2023; 11(3):410. https://doi.org/10.3390/healthcare11030410

[8]Ullah, F., Moon, J., Naeem, H. *et al.* Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model. *J Supercomput* **78**, 19246–19271 (2022). https://doi.org/10.1007/s11227-022-04631-z

[9]Siyuan Lu, Di Wu, Zheng Zhang, and Shui-Hua Wang. 2021. An Explainable Framework for Diagnosis of COVID-19 Pneumonia via Transfer Learning and Discriminant Correlation Analysis. ACM Trans. Multimedia Comput. Commun. Appl. 17, 3s, Article 103 (October 2021), 16 pages. https://doi.org/10.1145/3449785

[10]L. Zou *et al*., "Ensemble Image Explainable AI (XAI) Algorithm for Severe Community-Acquired Pneumonia and COVID-19 Respiratory Infections," in *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 242-254, April 2023, doi: 10.1109/TAI.2022.3153754. keywords: {Artificial intelligence;Predictive models;Deep learning;COVID-19;Pulmonary diseases;Visualization;X-ray imaging;Explainable artificial intelligent (AI);clinical decision

support;pneumonia;COVID-19;chest X-ray;neural network},

[11]Nafisah, S.I., Muhammad, G. Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Comput & Applic* **36**, 111–131 (2024).

https://doi.org/10.1007/s00521-022-07258-6

[12] Thimoteo, L.M., Vellasco, M.M., Amaral, J. *et al.* Explainable Artificial Intelligence for COVID-19 Diagnosis Through Blood Test Variables. *J Control Autom Electr Syst* **33**, 625–644 (2022). https://doi.org/10.1007/s40313-021-00858-y

[13]Dionysios Fanidis, Vasileios C. Pezoulas, Dimitrios I. Fotiadis, Vassilis Aidinis, An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers, Computational and Structural Biotechnology Journal, Volume 21, 2023, Pages 2305-2315, ISSN 2001-0370, https://doi.org/10.1016/j.csbj.2023.03.043.

(https://www.sciencedirect.com/science/article/pii/S2001037023001423)

[14]Kevser Kübra Kırboğa, Ecir Uğur Küçüksille, Muhammet Emin Naldan, Mesut Işık, Oktay Gülcü, Emrah Aksakal, CVD22: Explainable artificial intelligence determination of the relationship of troponin to D-Dimer, mortality, and CK-MB in COVID-19 patients, Computer Methods and Programs in Biomedicine, Volume 233, 2023, 107492, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2023.107492. (https://www.sciencedirect.com/science/article/pii/S016926072300158X)

[15]Meedeniya D, Kumarasinghe H, Kolonne S, Fernando C, Díez IT, Marques G. Chest X-ray analysis empowered with deep learning: A systematic review. Appl Soft Comput. 2022 Sep;126:109319. doi: 10.1016/j.asoc.2022.109319. Epub 2022 Jul 18. PMID: 36034154; PMCID: PMC9393235.

DATASET

LINK:https://drive.google.com/drive/folders/1W-yoG1fwfewv8VOMdCZmuSaj4Xap569o?usp=sharing