Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Multi Disease Diagnostic Analysis for Chest X-Ray Images with Explainable AI

Priyadharsini R[1], Beulah A[1], Prithika Priyadharshini H P[2], Rudrashree SB[2]

[1]Department of Artificial Intelligence and Data Science, Rajalakshmi Engineering College
[2]Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering, Chennai

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:
Integrating Explainable Artificial Intelligence (XAI) in the Multi Disease Dia- gnostic Model for Chest X-ray Images enables clearer, more interpretable diagnoses for stakeholders. By providing insights into model reasoning, XAI enhances trust and accountability, improving diagnostic transparency and accuracy, which is critical for healthcare providers and patients alike. In this project, various XAI tools such as LIME, GRAD-CAM were applied to improve reliability and interpretability in multi-disease diagnosis using chest X-ray images. An open-source dataset comprising chest X-ray images and associated metadata was used, and preprocessing techniques such as resizing, normalization, and data augmentation were applied to ensure data quality and relevance. Deep learning models were implemented, leveraging each model's strengths to improve diagnostic accuracy. The XAI methods LIME and GRAD-CAM were combined to improve model transparency and offer comprehensible findings about how decisions are made of the model. The interpretability aids various stakeholders, such as healthcare professionals, patients and data analysts, in understanding the reasoning behind model predictions, fostering trust and enabling more precise, data driven decisions in multi-disease diagnostic scenarios.

Keywords: Chest X-Ray Images, Interpretable Model, Deep Learning Model, Explainable AI.

1. Introduction

Artificial intelligence has become a disruptive force in healthcare, particularly in medical imaging. Chest X-rays are crucial for diagnosing various pulmonary and cardiovascular disorders, but interpreting these images requires significant expertise. This study aims to develop a robust Explainable AI model [2,3] that enhances interpretability, making AI-generated diagnoses more understandable and reliable for patients, physicians, and data analysts. By integrating advanced deep learning with explainability tools, the model supports informed, accurate medical decisions and fosters trust in AI-driven healthcare solutions [5,6].

Healthcare requires not only precise AI predictions but also clear explanations behind those predictions. Explainable AI addresses the "black box" issue associated with com- plex AI models, enhancing transparency and trust. Understanding AI decisions is crucial for accountability, high care standards, and minimizing risks, encouraging clinicians to integrate AI into their workflows and empowering patients to understand diagnoses and treatment recommendations [30]. The primary objective of this project is to integrate XAI into a Multi-Disease Diagnostic Model [25] for Chest X-

ray images, improving the clarity, interpretability, [8] and trustworthiness of AI-generated diagnoses [10]. Using XAI tools like LIME, GRAD-CAM, and SHAP [19] the project aims to validate explanations with interpretable models, assessing XAI techniques' effectiveness and accuracy.

This project employs XAI outputs to guide attention-weighted segment selection, ensuring the model focuses on clinically significant areas [29]. This comprehensive approach ensures AI-driven diagnostics are reliable and transparent, empowering healthcare professionals with data-driven insights and fostering trust in medical decision-making.

2. Interpretable Model

Interpretable models like Decision Trees and Random Forests are crucial for providing transparency and clarity in machine learning, especially in healthcare.[14][15] These models help demystify complex decision-making processes, making AI results more understandable for medical professionals and patients. Decision Trees offer clear visualizations of decision paths, allowing clinicians to see which factors lead to diagnoses or predictions [22]. This transparency builds trust and helps in verifying AI decisions. Random Forests enhance this further by using an ensemble of Decision Trees, improving prediction accuracy and reducing overfitting [20]. They provide a ranked list of feature importance, helping clinicians understand significant factors across scenarios [18].

However, analyses have shown that while these interpretable models offer better transparency and facilitate understanding, they often have slightly lower predictive accuracy compared to more complex models. Deep learning models, leveraging advanced neural network architectures, have shown superior performance in predictive tasks. To address the predictive accuracy gap while maintaining transparency, we integrated deep learning models with Explainable AI (XAI) techniques such as LIME (Local Interpretable Model agnostic Explanations), GRAD-CAM (Gradient-weighted Class Activation Mapping).

Combining these XAI techniques with deep learning models allows us to maintain high predictive performance while providing transparent insights. LIME helps explain individual predictions by approximating the model behavior with an interpretable local surrogate model around each prediction. GRAD-CAM provides visual explanations, high- lighting the regions in the input image that are most influential in making a prediction [21]. SHAP values explain the contribution of each feature to the final prediction, offering a comprehensive view of feature importance [19].

3. Deep Learning Model

Deep learning models [1,4] are powerful tools that leverage artificial neural networks with multiple layers to process and analyze data effectively. These models have revolutionized various fields, including healthcare, by improving the accuracy and efficiency of tasks such as medical image analysis. In particular, deep learning architectures like ResNet50 and DenseNet121 [16,17] are widely used for analyzing chest X-rays to diagnose multiple diseases. ResNet50 addresses this problem by incorporating residual connections, which allow gradients to flow more easily through the network [26]. These connections help maintain gradient strength, facilitating more efficient learning and enabling the network to train deeper layers effectively. As a result, ResNet50 achieves high performance in image classification tasks by learning complex features from large datasets. In DenseNet121, each layer receives inputs from all previous layers, allowing the network to leverage

previously learned features continuously. This dense connectivity reduces redundancy and promotes the reuse of features, making the network more computationally efficient. By minimizing the need for new feature extraction at each layer [12], DenseNet121 effectively reduces the number of parameters and computational cost while maintaining high accuracy in classification tasks. These optimizers are particularly effective in fine-tuning deep learning models, leading to quicker and more stable convergence.

Loss functions such as categorical cross-entropy and binary cross-entropy are essential for guiding model training. These functions measure the difference between the predicted outputs and the actual labels, providing a quantitative assessment of prediction accuracy [23]. Categorical cross-entropy is used for multi-class classification problems, where each sample belongs to one of several categories. Binary cross-entropy is used for binary classification problems, where each sample belongs to one of two categories. By minimizing these loss functions during training, the model learns to make more accurate predictions, ultimately improving classification performance.

These advanced architectures and techniques significantly enhance the effectiveness of deep learning models in medical image analysis. By leveraging ResNet50 and DenseNet121 [7], along with adaptive optimizers and appropriate loss functions, deep learning models [9] can accurately classify chest X-rays and diagnose multiple diseases. This capability is vital in healthcare, where precise and efficient diagnostics can improve patient outcomes and streamline clinical workflows [13]. Overall, the integration of these elements makes deep learning models highly effective tools for multi-disease diagnostics using chest X-rays.

4. Explainable AI

Explainable Artificial Intelligence [XAI] enhances the transparency and interpretability of AI models [27], particularly in healthcare, where trust and reliability are essential. By providing insights into AI-driven diagnostics, XAI enables clinicians to validate predictions, ensuring their accuracy and fostering integration into clinical workflows to improve patient outcomes [11]. Its significance is amplified in high-stakes environments, addressing ethical concerns, regulatory compliance, and continuous model refinement. Various XAI techniques, such as LIME, SHAP, and Grad-CAM, offer different perspectives on model interpretability [28].

LIME [Local Interpretable Model-agnostic Explanations] approximates complex models locally by perturbing input data and analyzing output variations. When applied to chest X-rays, it highlights critical regions influencing the diagnosis, aiding physicians in understanding AI decisions. LIME's flexibility and model-agnostic nature allow for granular insights into individual predictions, helping detect biases and errors while guiding future improvements in ensemble models. Its consistency ensures fair representation of feature contributions, aiding clinicians in making informed, evidence-based decisions.

Grad-CAM [Gradient-weighted Class Activation Mapping] visually explains AI decisions by generating heatmaps that highlight image regions most influential in predictions. Applied to chest X-rays, it helps clinicians verify diagnoses by showing which areas contribute most to the model's output. Grad-CAM also detects model biases, ensuring the AI focuses on medically relevant regions. By integrating LIME, SHAP, and Grad-CAM, this project enhances AI interpretability in chest X-ray

diagnostics, providing complementary insights that enable medical professionals to validate and trust AI-driven decisions, ultimately improving healthcare outcomes [24].

5. Experiments and Discussions

The experiments aimed to evaluate various machine learning models in the context of medical imaging, particularly chest X-ray interpretation. Models tested included traditional techniques like logistic regression, random forest, and decision trees, alongside a complex architecture like Multi-Layer Perceptron. Additionally, deep learning models such as ResNet50 and DenseNet121 were incorporated for their advanced feature extraction capabilities. Interpretability was assessed using XAI techniques: LIME was applied to traditional and simpler neural network models, while GRAD-CAM was used with deep learning models. These techniques made model predictions more transparent and understandable, crucial for medical applications where trust and accuracy are paramount.

LIME was used to provide explanations for models like Decision Trees and Random Forests. LIME achieves this by perturbing different image sections and observing changes in model predictions, thereby identifying regions influential to the outcome. For Decision Trees, visualized decision paths allowed clinicians to understand the factors leading to predictions. LIME visualizations for Decision Trees outline key areas that affect the model's decision, highlighting important regions for classification. GRAD-CAM heatmaps and overlays further illustrate where the model focused its attention.

Random Forests, on the other hand, enhance interpretability by aggregating results from multiple trees, thus improving prediction accuracy and reducing overfitting. LIME visualizations for Random Forests identify specific pixel ranges significant for the model's decision. GRAD-CAM visualizations for Random Forest highlight influential regions in the X-ray images, representing areas where pixel intensities significantly contributed to the model's classifications. These visualizations help to understand how the random forest model identifies diagnostic features. Class imbalance in the NIH Chest X-ray dataset was addressed to improve model performance. Ensuring equal representation for all classes reduced bias towards the dominant class and enhanced the model's ability to accurately detect examples from minority classes. This adjustment led to more balanced and precise predictions, which is crucial for detecting rare diseases in medical imaging.

For deep learning models, LIME was employed to highlight regions supporting specific diagnoses. LIME visualizations identify areas positively supporting diagnoses of conditions like Atelectasis and Emphysema, assigning weights to regions to signify their importance in the model's decision process. These visual explanations show the features consistent with the conditions, making the model's decisions clearer. GRAD-CAM overlays highlight specific areas of the chest X-rays that the model focused on during diagnosis, enhancing the interpretability and reliability of the model's predictions. By integrating these interpretable models and deep learning techniques, the study ensured reliable AI-driven insights without the "black box" effect. This approach fostered trust and accountability, enabling healthcare professionals to make better informed decisions and improving patient outcomes.

6. Conclusion

Data augmentation significantly improved model performance by increasing the size and diversity of training data, enhancing generalization and robustness. This was particularly beneficial for

interpretable models, which achieved lower accuracy. Despite their simplicity, these models, combined with XAI techniques like LIME and Grad-CAM, provided valuable insights into decision-making, aiding in visualization and transparency. Class imbalance, a common issue leading to biased predictions, was effectively mitigated using oversampling, undersampling, and cost-sensitive learning. These techniques ensured balanced attention to all classes, resulting in more accurate and reliable predictions. Deep learning models further boosted accuracy higher and beyond by capturing com plex patterns in the data. XAI methods, including LIME and Grad-CAM, were also applied to these models, enhancing interpretability. The use of more expressive LIME explanations provided clearer insights into model predictions, making even complex models more transparent. Overall, integrating data augmentation, class balancing strategies, and explainable AI techniques contributed to improved model performance, interpretability, and accountability in decision making.

References

[1] Ahmed R., Imran A.S. Knee Osteoarthritis Analysis Using Deep Learning and XAI on X-Rays. J. Med. Imaging Health Inform. 2023;13[4]:512–524.
[2] Akter S.B. Stroke Probability Prediction from Medical Survey Data: AI-Driven Analysis with Insightful Feature Importance using Explainable AI [XAI]. IEEE Access. 2023;11:12345–12356.
[3] Agughasi V.I. xAI: An Explainable AI Model for the Diagnosis of COPD from CXR Images. Expert Syst. Appl. 2023;198:113821.
[4] van der Velden H.M., Kuijf H.J., Gilhuijs K.G.A., Viergever M.A. Explainable artificial intelligence [XAI] in deep learning-based medical image analysis. Med. Image Anal. 2022;79:102470.
[5] Chaddad A., Peng J., Xu J., Bouridane A. Survey of Explainable AI Techniques in Healthcare. IEEE Trans. Neural Netw. Learn. Syst. 2023;34[9]:4025–4040.
[6] Dave D., Naik H., Singhal S., Patel P. Explainable AI Meets Healthcare: A Study on Heart Disease Dataset. Procedia Comput. Sci. 2021;184:730–739.
[7] Güngör S., Kaya M. Automatic Detection of Covid-19 from Colorized CT Images using Deep Learning. Comput. Biol. Med. 2021;137:104778.
[8] Naeem H., Bin-Salem A.A. A CNN-LSTM network with multi-level feature extraction-based approach for automated detection of coronavirus from CT scan and X-ray images. Appl. Soft Comput. 2021;113:107918.
[9] Jaiswal A., Singh M., Sachdeva N. Empirical Analysis of Heart Disease Prediction
[10] Using Deep Learning. J. Big Data. 2023;10[1]:1–22.
[11] Karim A., Islam S.M.S. Fast and Efficient Lung Abnormality Identification With Explainable AI: A Comprehensive Framework for Chest CT Scan and X-Ray Images. IEEE J. Biomed. Health Inform. 2024;28[3]:1011–1023.
[12] Lanfer E., Sylvester S., Aschenbruck N., Atzmueller M. Leveraging Explainable AI Methods Towards Identifying Classification Issues on IDS Datasets. J. Inf. Secur. Appl. 2023;68:103065.
[13] Pathak M., et al. A Robust EfficientNet Architecture for Brain Tumor Classification and Identification Using MRI Image. 11th Int. Conf. Intell. Syst. Embed. Des. [ISED]. 2023;1–5.
[14] Maxwell A., Li R., Yang B., et al. Deep learning architectures for multi-label classification of intelligent health risk prediction. BMC Bioinformatics. 2022;18[Suppl. 14]:523.

[15] Onari M.A., et al. Comparing Interpretable AI Approaches for the Clinical Environment: An Application to COVID-19. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. [CIBCB]. 2022;1–8.
[16] Rajjliwal N.S., Chetty G. Cardiovascular Disease Detection Based on Inter- pretable and Explainable AI. IEEE Asia-Pacific Conf. Comput. Sci. Data Eng. [CSDE]. 2022;1–7.
[17] Pant T.R. Disease Classification of Chest X-Ray using CNN. Int. J. Comput. Appl. 2021;175[14]:23–28.
[18] Puente F., et al. Predicting COVID-19 Cases using Deep LSTM and CNN Models. IEEE Colombian Conf. Appl. Comput. Intell. [ColCACI]. 2023;1–6.
[19] Romalt A.A., Kumar R.M.S. Prediction of Cardiovascular Disease by Deep Learning and Machine Learning - A Combined Data Science Approach. Procedia Comput. Sci. 2022;200:1123–1130.
[20] Brdnik S., Šumak B. Current Trends, Challenges and Techniques in XAI Field: A Tertiary Study of XAI Research. 47th MIPRO ICT Electron. Conv. [MIPRO]. 2024;2032–2038.
[21] Cheng S., et al. The application of interpretable machine learning model based on comparative learning and NARMAX in epidemic research. IEEE Int. Conf. Artif. Intell. Comput. Appl. [ICAICA]. 2022;1071–1076.
[22] Abeyagunasekera S.H.P., et al. LISA: Enhance the explainability of medical images unifying current XAI techniques. IEEE 7th Int. Conf. Converg. Technol. [I2CT]. 2022;1–9.
[23] Talapaneni S., et al. Enhancing Heart Disease Prediction and Analysis: An Efficient Voting Ensemble model. Int. Conf. Commun. Comput. Sci. Eng. [IC3SE]. 2024;156–160.
[24] Saednia K., Jalalifar A., Ebrahimi S. Deep Neural Network for Annotating Abnormalities in Chest X-ray Images. IEEE Access. 2022;10:87599–87608.
[25] Saraswat D., et al. Explainable AI for Healthcare 5.0: Opportunities and Challenges. IEEE Access. 2022;10:68456–68474.
[26] Tsoumakas G., Vlahavas I. A Review of Multi-Label Classification Methods. 2022;12[4]:51–64.
[27] Tyagi S. Detecting Diabetic Retinopathy using ResNet50 and Explainable AI. J. Med. Syst. 2023;47[2]:1–10.
[28] Wu X., Bell P., Rajan A. Can We Trust Explainable AI Methods on ASR? An Evaluation on Phoneme Recognition. IEEE/ACM Trans. Audio Speech Lang. Process. 2024;32:1–13.
[29] Yang C.C. Explainable Artificial Intelligence for Predictive Modeling in Health- care. J. Healthc. Inform. Res. 2022;6:228–239.
[30] Coppola F., Faggioni L., Regge D., Giovagnoni A., Golfieri R., Bibbolino C., Miele V., Neri E., Grassi R. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. Radiol. Med. 2021;126[1]:63-71. doi: 10.1007/s11547-020-01205-y.
[31] Tang A., Tam R., Cadrin-Chênevert A., Guest W., Chong J., Barfett J., Chepelev L., Cairns R., Mitchell J.R., Cicero M.D., Poudrette M.G., Jaremko J.L., Reinhold C., Gallix B., Gray B., Geis R. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can. Assoc. Radiol. J. 2018;69[2]:120-135. doi: 10.1016/j.carj.2018.02.002.