# Enhancing Healthcare Outcomes with Explainable AI (XAI) for Disease Prediction: A Comprehensive Review

Salman Muneer[1], Atta Rasool[2]

[1]The Islamia University of Bahawalpur, Bahawalpur, Pakistan.
[2]The Capital University of Science & Technology, Islamabad, Pakistan

Salman Muneer: mrsalman2013@gmail.com

***Abstract***- Nowadays, diseases are caused by harmful ingredients such as origins, pathogens, microbes, microorganisms, and viruses that can enter the human body. Early prediction of diseases is crucial in saving human lives, and it remains one of the most challenging issues in clinical data analysis. The recent advancements in Artificial Intelligence (AI) have significantly improved the efficiency and accuracy of disease prediction. AI is a field of computational intelligence that encompasses various applications such as expert systems, Natural Language Processing (NLP) etc.. While some previous AI approaches are manual and prone to errors, Explainable Artificial Intelligence (XAI) forecast approaches allow for automatic identification and forecast, thereby minimizing detection faults compared to depending exclusively on human skill. This paper provides a comprehensive review of advanced approaches such as RF, VGG-16, SVM, SMOTE, ECNNs, CNN, NB, and XG Boost, all of which utilize XAI methods. Among these techniques, the Random Forest (RF) method with explainable techniques such as SMOTE and ADASYN achieved the highest accuracy of 98%. This finding highlights the potential of XAI in enhancing the accuracy and reliability of disease prediction models. By providing explanations for the predictions made by the model, XAI techniques increase transparency and trustworthiness, enabling medical professionals to better understand the reasoning behind the diagnosis.

***Keywords:*** Explainable artificial intelligence (XAI), disease forecast, comprehensive review

## 1 Introduction

In the healthcare sector, disease prediction analyses are essential. Diseases can cause distress, ailment, dysfunction, and even lead to a person's demise, significantly impacting their physical and mental well-being and altering their method of life. The clinical system involves investigating the causes of a disease, which is identified by analyzing warning signs and indications. Identification is the system of determining a illness's pathology from its signs and indications, while identification is recognizing a illness based on a human's symptoms. Physical examinations and basic health tests provide the necessary information for disease identification and forecast, which are the primary challenging and critical methods for healthcare specialists to complete. These procedures are often time-taking and tough, requiring healthcare experts to collect empirical information to reduce the uncertainty in health identification. Faults in the identification system can delay or overlook the proper treatment of patients, especially in cases of severe medical issues. Unfortunately, not all doctors are experts in every area of medication [1].

Therefore, there was a need for an automated system that could leverage both human expertise and machine precision to achieve accurate disease diagnosis at a lower cost. While categorizing disorders based on various factors can be complex for human professionals, AI has the potential to assist in identifying and managing such cases. Various AI techniques, such as deep learning and machine learning [2], are currently being utilized in the medical field to accurately diagnose and predict illnesses. However, precise disease prediction still presents challenges. To overcome these challenges, a new concept called "Explainable Artificial Intelligence (XAI)" has been introduced.

Explainable Artificial Intelligence (XAI) refers to a collection of approaches and systems that allow individuals to understand and have faith in the outcomes and productivity produced by Machine Learning (ML) techniques. The concept of XAI encompasses the description of an AI model, its expected impacts, and any possible biases it may contain. Its purpose is to define model that is assisted by AI. The foundation of XAI is based on three key steps: 1) Possibility of Interpretability, which enables understanding of a system's complete logic or a specific decision or forecast (local); 2) Pertinency, which allows for the application of a technique to any sort or class of AI method (model-agnostic) or to a unique sort or period; and 3) Duration of Interpretability, which describes when the clarification of a system occurs, either later it has been proficient and is dissimilar to its interior structure, or is inherently easy to comprehend (intrinsic) [3].

## 2 Related work

Explainable Artificial Intelligence (XAI) is a crucial area within AI that focuses on developing techniques to provide explanations for the inner workings of learning algorithms, models, and knowledge-based inference techniques. XAI has been utilized by numerous researchers to predict diseases. In this regard, some notable works include Li et al.'s use of multivariate logistic regression with a feature selection method to predict the impact of breathing crevice exhaustiveness on postoperative cardiopulmonary problems and hospital stay in main-phase non-small-cell lung tumor patients. Binary logistic regression is another effective method used by authors to analyze the predictability of essential postoperative cardiopulmonary problems in non-small cell lung tumor patients based on oxygen desaturation (EOD) and heart rate [4].

In the field of XAI, Amoroso et al. proposed a framework for breast cancer treatment that demonstrated the ability to identify important medical features of patients and the appropriate oncological treatments using a grouping and dimensionality reduction technique. Similarly, the authors [5] developed an explainable spinal posture classifier that was independent of specific pathologies. They used LIME to explain the predictions made by their machine learning model [31-36], which utilized SVM and RF algorithms. In [6], the authors built an RF system for AD detection and drive perception. They utilized SHAP to select the primary characteristics for the classifier and a fuzzy rule-based method to produce human-readable forms that make it easier for patients and doctors to understand the AI model [40-46]. SHAP also enabled the authors to provide local explanations of feature influences for individual patient identification/movement predictions.

In their study, Sarp et al. proposed a CNN-based approach to classify chronic wounds in [7]. They also employed the XAI technique LIME to interpret the CNN-enabled system. The proposed approach utilized Transfer Learning (TL) and achieved satisfactory results with respect to precision (95%), recall (94%), and F1-score (94%). The model generated a heatmap using LIME that provided visual cues to clinicians based on the input wound image. In [8], Tan et al. developed a neural network called LNN for diagnosing fenestral otosclerosis on temporal HRCT bone slices.

Chen et al. developed an XAI model for medical identification using Electronic Medical Record (EMR) data. Their suggested model incorporated entity-aware CNN systems and groups of BN, achieving a Top-3 prediction precision of greater than 88%. Specifically, the model's explainability was achieved through a Bayesian network (BN) that established connections between diseases and symptoms. To evaluate the model's interpretability, three certified surgeons examined the relationships obtained from the health information graph [37-39].

In their study, Zhang and colleagues [9] proposed a deep learning (DL) system for predicting mechanical ventilation (MV) risk in COVID-19 patients within 24 hours of hospitalization. The researchers utilized all relevant patient data, containing demographic data, lab outcomes, health data, and medical procedures, and employed the attention mechanism to weight the data. The performance of the suggested DBNet model was compared with other machine learning (ML) and DL models, and it demonstrated superior predictive accuracy with an accuracy of 80% and an F1 precision 79.8%.

## 3 Methodology

In recent times, accurate prediction of diseases has gained a lot of importance due to various factors such as ecological changes and unhealthy lifestyles. Detecting health issues at an early stage can help in preventing severe medical problems. To overcome this challenge, the emerging technology of

Explainable Artificial Intelligence (XAI) has proven to be a valuable tool in detecting various diseases as demonstrated in a recent study. XAI models have been shown to exhibit excellent performance in identifying different health conditions, as illustrated in Table 1 of the research paper. The ability of XAI to provide explanations for its predictions allows medical professionals to understand the reasoning behind the decision-making process. This feature is particularly useful when dealing with complex medical scenarios where multiple factors can affect the diagnosis.

**Table 1:** XAI-enabled approaches and their accuracy

| Sr. No. | Ref. | Approach | XAI based Algorithm | Accuracy |
|---|---|---|---|---|
| 1 | [10] | RF | SMOTE and ADASYN | 98% |
| 2 | [11] | VGG-16 | LIME | 97% |
| 3 | [12] | SVM | Virtual operative assistant | 92% |
| 4 | [13] | SMOTE | SMOTE with Oversampling | 90% |
| 5 | [14] | ECNNs | Bayesian network ensembles | 88.8% |
| 6 | [15] | CNN | Visually Interpretable Network (VINet), LRP, CAM, VBP | 82.15% |
| 7 | [16] | NB | Context-free grammar (CFG) | 81.5% |
| 8 | [17] | XGBoost | SHAP | 78.9% |

In the research paper, Table 1 presents various AI-based approaches such as [10], [11], [12], [13], [14], [15], [16], and [17], that have been employed to analyze different datasets using XAI methods. The results indicate that the RF technique in combination with XAI techniques such as SMOTE and ADASYN achieved the highest accuracy of 98%. It is noteworthy that the utilization of XAI techniques has improved the accuracy of disease prediction models, making them more reliable and transparent. The RF technique, in particular, has proven to be a successful algorithm in predicting various diseases with high accuracy rates. The incorporation of XAI techniques such as SMOTE and ADASYN has further improved the performance of the model, indicating the potential of XAI to enhance the accuracy and efficiency of disease detection systems.

## 4 Conclusion

In this research, various AI-enabled approaches, including [1], [2], [3], [4], [5], [6], [7], and [8], were employed to several datasets using XAI methods to achieve maximum accuracy in disease prediction. The results showed that these techniques achieved accuracies ranging from 78.9% to 98%. Upon comparing the results of these techniques, it was observed that the RF technique achieved the highest accuracy of 98% among all the techniques utilized. The other techniques, including [2], [3], [4], [5], [6],[7], and [8], exhibited accuracies ranging from 97% to 78.9%. The findings of this study indicate the potential of AI-based techniques combined with XAI methods in disease prediction. The high accuracy achieved by the RF technique highlights its effectiveness in predicting diseases, indicating its potential as a reliable tool for healthcare professionals. This study also emphasizes the importance of utilizing XAI methods in disease prediction models to enhance their accuracy and transparency. The results of this study provide valuable insights for researchers and medical professionals working towards improving disease prediction models.

## 5 References

[1] Usuelli M. The Lombardy region of Italy launches the first investigative COVID-19 commission. The Lancet. 2020 Nov 14;396(10262):e86-7.

[2] Abualigah L, Zitar RA, Almotairi KH, Hussein AM, Abd Elaziz M, Nikoo MR, Gandomi AH. Wind, solar, and photovoltaic renewable energy systems with and without energy storage optimization: a survey of advanced machine learning and deep learning techniques. Energies. 2022 Jan 14;15(2):578.

[3] Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Information Fusion. 2022 Jan 1;77:29-52.

[4] Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, Darwish O. An explainable machine learning framework for lung cancer hospital length of stay prediction. Scientific reports. 2022 Jan 12;12(1):1-0.

[5] Dindorf C, Konradi J, Wolf C, Taetz B, Bleser G, Huthwelker J, Werthmann F, Bartaguiz E, Kniepert J, Drees P, Betz U. Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (Xai). Sensors. 2021 Sep 21;21(18):6323.

[6] Zhang Y, Weng Y, Lund J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. Diagnostics. 2022 Jan 19;12(2):237.

[7] D. Zhang, L. Zou, X. Zhou, F. He, Integrating Feature Selection and Feature Extraction Methods With Deep Learning to Predict Clinical Outcome of Breast Cancer, IEEE Access. 6 (2018) 28936–28944. https://doi.org/10.1109/ACCESS.2018.2837654.

[8] Tan W, Guan P, Wu L, Chen H, Li J, Ling Y, Fan T, Wang Y, Li J, Yan B. The use of explainable artificial intelligence to explore types of fenestral otosclerosis misdiagnosed when using temporal bone high-resolution computed tomography. Annals of Translational Medicine. 2021 Jun;9(12).

[9] Aslam N. Explainable Artificial Intelligence Approach for the Early Prediction of Ventilator Support and Mortality in COVID-19 Patients. Computation. 2022 Feb 28;10(3):36.

[10] Roscher R, Bohn B, Duarte MF, Garcke J. Explainable machine learning for scientific insights and discoveries. Ieee Access. 2020 Feb 24;8:42200-16.

[11] Rucco, M.; Viticchi, G.; Falsetti, L. Towards personalized diagnosis of glioblastoma in fluid-attenuated inversion recovery (FLAIR) by topological interpretable machine learning. Mathematics 2020, 8, 770.

[12] Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. PloS one. 2020 Feb 27;15(2):e0229596.

[13] Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine learning-based model to predict the disease severity and outcome in COVID-19 patients. Scientific programming. 2021 Apr 20;2021.

[14] Chen, J.; Dai, X.; Yuan, Q.; Lu, C.; Huang, H. Towards Interpretable Clinical Diagnosis with Bayesian Network Ensembles Stacked on Entity-Aware CNNs. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 3143–3153.

[15] Gu, D.; Li, Y.; Jiang, F.; Wen, Z.; Liu, S.; Shi, W.; Lu, G.; Zhou, C. VINet: A Visually Interpretable Image Diagnosis Network. IEEE Trans. Multimed. 2020, 22, 1720–1729.

[16] Kroll, J.P.; Eickhoff, S.B.; Hoffstaedter, F.; Patil, K.R. Evolving complex yet interpretable representations: Application to Alzheimer's diagnosis and prognosis. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020.

[17] Yoo, T.K.; Ryu, I.H.; Choi, H.; Kim, J.K.; Lee, I.S.; Kim, J.S.; Lee, G.; Rim, T.H. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. Transl. Vis. Sci. Technol. 2020, 9, 1–14

[18] AsadUllah, M., Khan, M. A., Abbas, S., Athar, A., Raza, S. S., & Ahmad, G. (2018). Blind channel and data estimation using fuzzy logic-empowered opposite learning-based mutant particle swarm optimization. Computational intelligence and neuroscience, 2018.

[19] Khan, F., Khan, M. A., Abbas, S., Athar, A., Siddiqui, S. Y., Khan, A. H., ... & Hussain, M. (2020). Cloud-based breast cancer prediction empowered with soft computing approaches. Journal of healthcare engineering, 2020.

[20] Rehman, A., Athar, A., Khan, M. A., Abbas, S., Fatima, A., & Saeed, A. (2020). Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine. Journal of Ambient Intelligence and Smart Environments, 12(2), 125-138.

[21] Khan, M. A., Abbas, S., Rehman, A., Saeed, Y., Zeb, A., Uddin, M. I., ... & Ali, A. (2020). A machine learning approach for blockchain-based smart home networks security. IEEE Network, 35(3), 223-229.

[22] Khan, M. A., Abbas, S., Atta, A., Ditta, A., Alquhayz, H., Khan, M. F., & Naqvi, R. A. "Intelligent cloud based heart disease prediction system empowered with supervised machine learning," Computers, Materials & Continua, vol. 65, no.1, pp. 139–151, 2020.

[23] Khan, M. A., Umair, M., Saleem, M. A., Ali, M. N., & Abbas, S. (2019). CDE using improved opposite based swarm optimization for MIMO systems. Journal of Intelligent & Fuzzy Systems, 37(1), 687-692.

[24] Saleem, M., Khan, M. A., Abbas, S., Asif, M., Hassan, M., & Malik, J. A. (2019, July). Intelligent FSO link for communication in natural disasters empowered with fuzzy inference system. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.

[25] Ata, A., Khan, M. A., Abbas, S., Khan, M. S., & Ahmad, G. (2021). Adaptive IoT empowered smart road traffic congestion control system using supervised machine learning algorithm. The Computer Journal, 64(11), 1672-1679.

[26] Siddiqui, S. Y., Athar, A., Khan, M. A., Abbas, S., Saeed, Y., Khan, M. F., & Hussain, M. (2020). Modelling, simulation and optimization of diagnosis cardiovascular disease using computational intelligence approaches. Journal of Medical Imaging and Health Informatics, 10(5), 1005-1022.

[27] Fatima, A., Adnan Khan, M., Abbas, S., Waqas, M., Anum, L., & Asif, M. (2019). Evaluation of planet factors of smart city through multi-layer fuzzy logic (MFL). The ISC International Journal of Information Security, 11(3), 51-58.

[28] Hussain, S., Abbas, S., Sohail, T., Adnan Khan, M., & Athar, A. (2019). Estimating virtual trust of cognitive agents using multi layered socio-fuzzy inference system. Journal of Intelligent & Fuzzy Systems, 37(2), 2769-2784.

[29] Asif, M., Khan, M. A., Abbas, S., & Saleem, M. (2019, January). Analysis of space & time complexity with PSO based synchronous MC-CDMA system. In 2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET) (pp. 1-5). IEEE.

[30] Ihnaini, B., Khan, M. A., Khan, T. A., Abbas, S., Daoud, M. S., Ahmad, M., & Khan, M. A. (2021). A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning. Computational Intelligence and Neuroscience, 2021.

[31] G. Ahmad, S. Alanazi, M. Alruwaili, F. Ahmad, M. A. Khan et al., "Intelligent ammunition detection and classification system using convolutional neural network," Computers, Materials & Continua, vol. 67, no.2, pp. 2585–2600, 2021.

[32] Hanif, M., Naqvi, R. A., Abbas, S., Khan, M. A., & Iqbal, N. (2020). A novel and efficient 3D multiple images encryption scheme based on chaotic systems and swapping operations. IEEE Access, 8, 123536-123555.

[33] Thakker, D.; Mishra, B.K.; Abdullatif, A.; Mazumdar, S.; Simpson, S. Explainable Artificial Intelligence for Developing Smart Cities Solutions. Smart Cities 2020, 3, 1353–1382.

[34] Asif, M., Abbas, S., Khan, M.A., Fatima, A., Khan, M.A. and Lee, S.W., 2021. MapReduce based intelligent model for intrusion detection using machine learning technique. Journal of King Saud University-Computer and

Information Sciences.

[35] F. Alhaidari, S. H. Almotiri, M. A. Ghamdi, M. A. Khan, A. Rehman et al., "Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach," Computers, Materials & Continua, vol. 67, no.1, pp. 1269–1285, 2021.

[36] Naz, N. S., Khan, M. A., Abbas, S., Ather, A., & Saqib, S. (2020). Intelligent routing between capsules empowered with deep extreme machine learning technique. SN Applied Sciences, 2(1), 1-10.

[37] A. H. Khan, M. A. Khan, S. Abbas, S. Y. Siddiqui, M. A. Saeed et al., "Simulation, modeling, and optimization of intelligent kidney disease predication empowered with computational intelligence approaches," Computers, Materials & Continua, vol. 67, no.2, pp. 1399–1412, 2021.

[38] Komninos, N.; Bratsas, C.; Kakderi, C.; Tsarchopoulos, P. Smart City Ontologies: Improving the effectiveness of smart city applications. J. Smart Cities 2016, 1, 31–46.

[39] Rizvi, S. S. R., Sagheer, A., Adnan, K., & Muhammad, A. (2019). Optical character recognition system for Nastalique Urdu-like script languages using supervised learning. International Journal of Pattern Recognition and Artificial Intelligence, 33(10), 1953004.

[40] Hussain, S., Naqvi, R. A., Abbas, S., Khan, M. A., Sohail, T., & Hussain, D. (2021). Trait based trustworthiness assessment in human-agent collaboration using multi-layer fuzzy inference approach. IEEE Access, 9, 73561-73574.

[41] Q. Khan, S. Abbas, M. A. Khan, A. Fatima, S. Alanazi et al., "Modelling intelligent driving behaviour using machine learning," Computers, Materials & Continua, vol. 68, no.3, pp. 3061–3077, 2021.

[42] N. Tabassum, A. Ditta, T. Alyas, S. Abbas, H. Alquhayz et al., "Prediction of cloud ranking in a hyperconverged cloud ecosystem using machine learning," Computers, Materials & Continua, vol. 67, no.3, pp. 3129–3141, 2021.

[43] Ghazal, T.M., Abbas, S., Ahmad, M. and Aftab, S., 2022, February. An IoMT based Ensemble Classification Framework to Predict Treatment Response in Hepatitis C Patients. In 2022 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-4). IEEE.

[44] Abbas, S., Fatima, A., Asif, M. and Saleem, M., Energy Optimization in Smarts Homes by using Fuzzy Inference System.

[45] Khan, T.A., Khan, M.S., Abbas, S., Janjua, J.I., Muhammad, S.S. and Asif, M., 2021, April. Topology-Aware Load Balancing in Datacenter Networks. In 2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob) (pp. 220-225). IEEE.

[46] Alyas, T.A.T., 2018. Data Breaches Security Issues for Cloud Based Internet of Things. International Journal for Electronic Crime Investigation, 2(1), pp.7-7.