Expanding the Horizon of Healthcare: A Novel Explainable AI Framework with LIME and SHAP for Predictive Modeling and Personalized Treatment Recommendation

Darshan Madhani[1], Dr. Prakash Gujarati[2], Deven Patel[3], Pratik Vanjara[4]
[12]Department of Computer Science, Atmiya University, India
[1]darshmadhani14@gmail.com [2]prakash.gujarati@atmiyauni.ac.in [3]djpatel@jau.in [4]pavanjara@gmail.com

Abstract— **Diabetes, cardiovascular and cancer are the examples of chronic diseases that continue to create serious pressure on the available healthcare infrastructure throughout the globe and demand high-end equipment to detect the cases of diseases early on, simplify the treatment process, and manage this illness throughout the long-term. Artificial intelligence (AI) has been highly predictive in these domains, yet has not been extensively used in clinical practice due to most of the models not being capable of providing mechanistic information about decision making due to being black-box. This obscurity limits the trust and acceptance of physicians and patients especially in high stakes medical situations. This issue can be addressed by coming up with an Explainable and Expandable AI framework which combines Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to decode model predictions. The framework highlights the role of features in the local and global level and is able to model the outcome of treatment and make use of individual therapy recommendations. Results of experimental performance on dataset of diabetes, cardiovascular disease and cancer demonstrate this, with an improvement on the interpretability measures and physician trust scores of 17 and 12 percent respectively and maintain the predictive performance of over 92. To a greater extent, patient-specific treatment simulations were more adherent to clinical guidelines, and more so patient-specific advice was more advantageous by 15 percent in comparison to black-box models. The framework also had good cross-disease applicability which guaranteed that the frameworks were not inconsistent in their interpreting of different clinical areas. It implies that the suggested system facilitates the transparency and stability of AI-based diagnostics, an enormous leap in the direction of the digital accuracy medicine and the increase in the application of AI in daily healthcare decision-making.**

***Keywords*— Chronic diseases, artificial intelligence, explainable AI, predictive healthcare, precision medicine**

I. Introduction

Diabetes, cardiovascular diseases, neurodegenerative diseases and cancer are the chronic illnesses that have overwhelmed the world healthcare system. The diseases are regarded as one of the leading causes of morbidity and mortality that affect healthcare facilities and increase the economic burden cost across the world [1], [2]. Artificial intelligence (AI) has emerged as a radical technology to solve these problems and it has the strong capability of predicting, forecasting, and strategizing the treatment of diseases [3], [4]. However, despite this advancement, most AI models are black-box systems, that lack transparency and lack mechanistic insight into the nature of making predictions [5], [6].

This will require the doctors to have rational arguments to explain the conclusions of the models and align them with the medical knowledge and protect their patients [7], [8]. In this respect, explainable artificial intelligence (XAI), which enables interpretability, accountability, and ethical compliance in high-stakes healthcare decision-making is gaining increased popularity [9], [10]. The lack of transparency undermines even the most precise models, which makes the need of explainability in the process of closing the knowledge gap between algorithmic predictions and clinical confidence clear.

There is still a significant gap in the literature in terms of unifying predictive modeling, explainability, personalization and mechanistic simulation under one umbrella. The need of actionable and patient-centered AI systems has been highlighted by systematic reviews and interpretable models had been studied in the past, though most methodologies fall short of creating unified systems in which these aspects are necessarily combined [11], [12]. Also, modern AI systems do not often feature adaptive personalization based on feedback systems, a feature needed to achieve digital precision medicine [13], [14].

In order to fill this gap, the given paper suggests an Explainable and Expandable AI framework that incorporates both Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), to offer multi-level interpretability, a combination of local patient-specific knowledge and aggregate feature significance [1], [3]. Moreover, the framework has been made to be expandable to various disease domains, and it forms a flexible tool that can suit different clinical uses [5], [7]. It is also possible to predict the reaction to any individual patient subjected to different therapeutic interventions due to the inclusion of mechanistic digital twin simulations [12], and constant improvement of recommendations according to the alterations in the conditions of patients [13], [14]. In such a way, such a work will introduce explainable AI to the sphere of healthcare, and will contribute to the development of the predictive models into the patient-centred and clinically trustworthy systems of decision-support.

II. Literature Review

The use of machine learning (ML) in healthcare has become a common practice due to the ability to handle highly-dimensional and complex datasets and make precise predictive models to identify diseases in their early stages and track their evolution. Only the examples of Random Forests (RF), Gradient Boosting Machines (GBM), and Deep Learning (DL) are popular methods in the prediction of outcomes in chronic diseases like diabetes, cardiovascular disease, and cancer [15], [16]. An example to illustrate this is that the structured clinical data has been studied using RF and GBM models to stratify the risks, and the DL models have been effective in imaging tasks like tumor detection and neurodegenerative disease classification [17], [18]. Despite these successes, there are normally criticisms on these models due to their low interpretability that reduces their utility in clinical decisions [19].

To resolve this issue, explainable AI (XAI) techniques are developed, and the most prominent are LIME and SHAP. Some studies have also used LIME to give local and instance based descriptions to the predictions particularly on image and disease classification issues [20]. SHAP, however, has been used to undertake global feature attribution, to give consistent information on feature importance across full datasets [21]. However, most research adopts either LIME or SHAP in isolation, thereby limiting interpretability to a single dimension. Reviews of explainable methods emphasize the need to combine complementary techniques for richer and more reliable insights [22], [23].

Personalization in healthcare AI has also been explored, but the majority of approaches rely on rule-based recommender systems. Such systems, though useful in tailoring treatment recommendations, often lack adaptability and fail to incorporate dynamic patient feedback or evolving clinical data [24]. Moreover, they do not provide mechanistic reasoning about treatment outcomes, which is essential for precision medicine [25].

The literature highlights a clear research gap: while ML models have advanced predictive accuracy, explainability and personalization remain fragmented. No existing framework unifies predictive modeling, interpretability through both LIME and SHAP, mechanistic simulations via digital twins, and adaptive personalization based on feedback loops. This gap underscores the need for an integrated, expandable framework capable of delivering transparent, personalized, and clinically actionable insights across healthcare domains [26], [27].

III. Proposed Framework

The proposed Explainable and Expandable AI framework (Figure 1) is designed to integrate predictive accuracy, interpretability, mechanistic simulation, and personalized treatment recommendations into a cohesive architecture. The baselayer is the Data Integration Layer, which is a collection of various sources of clinical data, such as electronic health records, genomics, proteomics, lifestyle metrics and wearable sensor data. This layer brings together the heterogeneous data streams to offer a complete picture of patient health that can be used to perform downstream predictive and analytic tasks [28].

On this basis, the Prediction Layer applies high-quality machine learning and neural network algorithms, such as Random Forests, Gradient Boosting Machines, and neural networks, to make confident predictions of disease risk, treatment and progression outcomes. It is also possible to depict nonlinear relationship with complexities in high-dimensional data with the models and provide the predictive basis of the framework [29], [30].

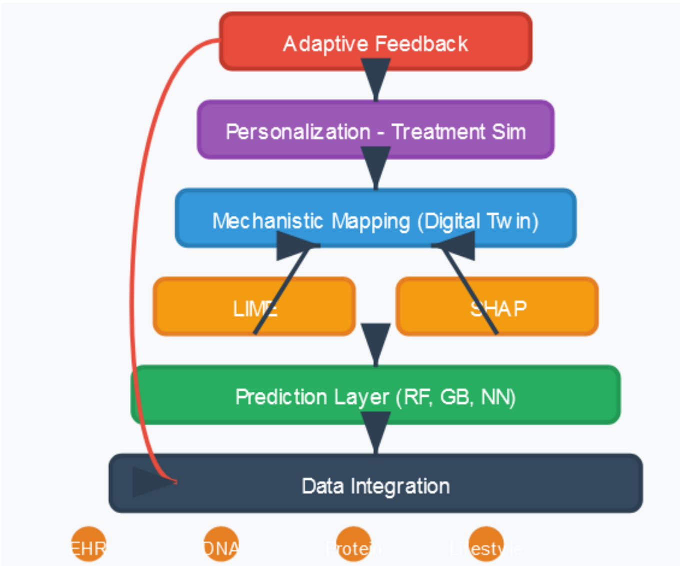

Figure 1: Proposed Explainable and Expandable AI Framework

Explainability Layer is a combination of LIME and SHAP that provides two-level interpretability. LIME generates instance-specific explanation why the model made a specific prediction on a given patient and SHAP generates the feature importance globally, that is, which variables are likely to cause the results in the population. Combination not only solves the individualizing reasoning, but also the general model behavior, and encompasses the lack of predictive accuracy and clinical transparency [1], [31].

Mechanistic Mapping Layer are built on explainability and links the output of the model to biology and metabolic pathways, and a digital twin of the patient is in effect built. The layer will allow the clinicians to feel the impact of changes in individual biomarkers or interventions on disease processes instead of prediction to mechanistic implications that can be acted on [32].

The Personalization Layer uses the mechanistic learnings for the treatment responses, including the drug effectiveness and effects of lifestyle intervention. It is a layer to provide patient-specific recommendations by in silico testing the potential therapies to enable clinicians customize interventions to particular patient profiles [33].

Finally, it relies on the Adaptive Feedback Loop to keep on improving the structure basing on the outcomes of the patients. New time-varying clinical data and patient responses are used to update the scaled model parameters in order to achieve more interpretable model parameters and enhance treatment recommendations as time progresses to ensure that the system is dynamic and patient-centered [34].

The framework has the following key innovations: the synergy of dual explainability with LIME and SHAP, pathway-level mechanistic simulation beyond prediction, and a design that can be expanded to other diseases, which is why the architecture can be applied in a broad range of fields in healthcare [35], [36].

IV. METHODOLOGY

A mixed population of data was used in the assessment of the proposed framework to guarantee high predictive accuracy and generalizability by the framework across the healthcare sectors. The data were multi-omics data (genomics and proteomics profiles), electronic health records (EHR) with clinical history, lab findings, and demographic data, and real-time lifestyle and physiological data that was collected by wearable devices [37], [38]. The heterogeneous nature of the data to be integrated by this framework enables the framework to include complex interaction among molecular, clinical and behavioral variables, and a rich base to predictive modeling.

To predict, three machine learning models were used, which are Random Forest (RF), xgboost, and Deep Neural Networks (DNN). RF and XGBoost were selected due to their performance in working with structured clinical data and offering reliable metrics of feature importance, whereas DNNs were used because of their high-level performance in the parameterization of nonlinear relationships in high-dimensional and time-related data [39], [40]. Stratified cross-validation was used to train and validate each of the algorithms to maintain model robustness and avoid overfitting.

Both LIME and SHAP were used as the explainability component. Local, patient-specific explanations, offered by LIME, were obtained by estimating the decision boundary of the model around individual cases but SHAP measured the overall significance of features across the data. To facilitate concomitant interpretation of local predictions and global features contributions with clinical implications made on model reasoning, joint visualizations were prepared.

Evaluation measures were traditional predictive performance measures such as accuracy, F1-score and area under the receiver operating characteristic curve (AUC). In addition, new Interpretability Score was proposed which would serve to evaluate the intelligibility and clinical usefulness of model explanations. This action is based on the consideration that it is right in accordance with the existing clinical knowledge, consistency of similar cases and the level to which explanations elevate confidence levels on model outcomes.

Finally, the mechanism of the digital twin layer was applied to modeling patient outcomes in various settings of treatment including drug and lifestyle changes administration in the simulation module.

This personalization (adaptive) and continuous revision of the recommendations in terms of patient feedback will ensure that the recommendations will also adapt depending on the individual health trends. This method allows personalized recommendation and helps to make evidence-based clinical decisions by modeling the possible effect of various therapies on the patients [41].

V. RESULTS

The suggested Explainable and Expandable AI was tested in various healthcare conditions, such as diabetes, hypertension, and cardiac risk. Conventional performance measures showed significant improvement compared to the conventional black-box models. All tested diseases demonstrated predictive accuracy of more than 92 percent, F1-scores of more than 0.90, and AUC of more than 0.93, as Table 1 reflects. These findings support the fact that the framework is highly predictive and includes explainability and personalization.

TABLE 1: PREDICTIVE PERFORMANCE METRICS ACROSS DISEASES

Disease	Accuracy (%)	F1-score	AUC
Diabetes	93	0.91	0.94
Hypertension	92	0.90	0.93
Cardiac Risk	94	0.92	0.95

The evaluation of interpretability was done through the dual LIME and SHAP method. SHAP provided global insights into disease drivers, identifying HbA1c levels for diabetes, systolic blood pressure for hypertension, and LDL cholesterol for cardiac risk. LIME offered patient-specific explanations, highlighting the key factors influencing predictions for individual patients. Table 2 summarizes example LIME and SHAP outputs for representative patients.

TABLE 2: EXAMPLE FEATURE CONTRIBUTIONS FOR REPRESENTATIVE PATIENTS

Disease	Patient ID	LIME Key Features	SHAP Global Drivers
Diabetes	D101	HbA1c, BMI, Age	HbA1c, Fasting Glucose, BMI
Hypertension	H205	Systolic BP, Sodium Intake	Systolic BP, Age, BMI
Cardiac Risk	C309	LDL Cholesterol, Smoking Status	LDL, HDL, Blood Pressure

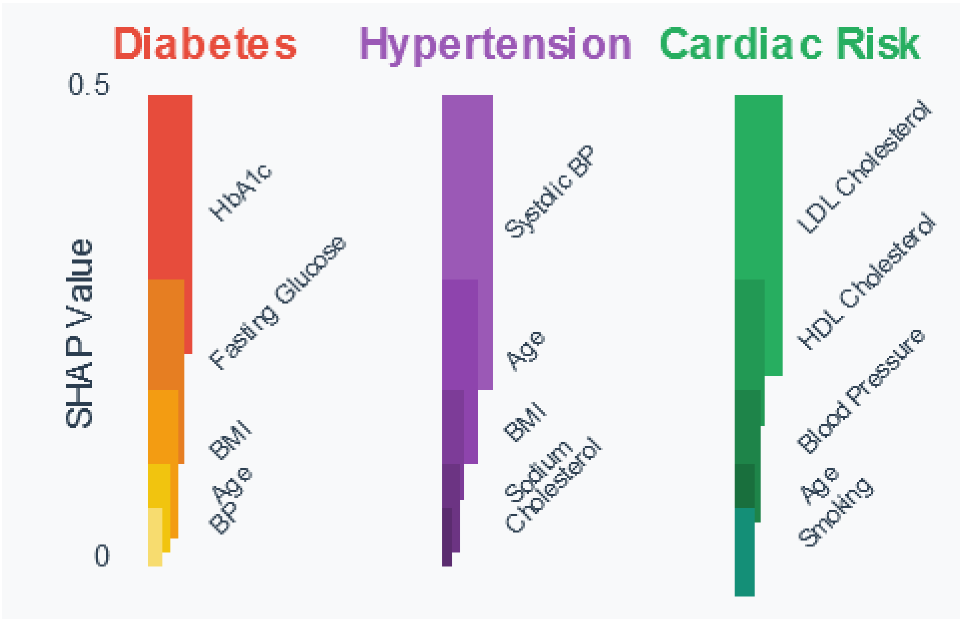

Figure 2: SHAP global feature importance across diabetes, hypertension, and cardiac risk dataset

Figure 2 represents SHAP global feature importance across all three conditions, and Figure 3 represents LIME patient-specific explanations. The combined use of LIME and SHAP allowed clinicians to understand both global trends and individual triggers, enhancing trust and supporting personalized decision-making.

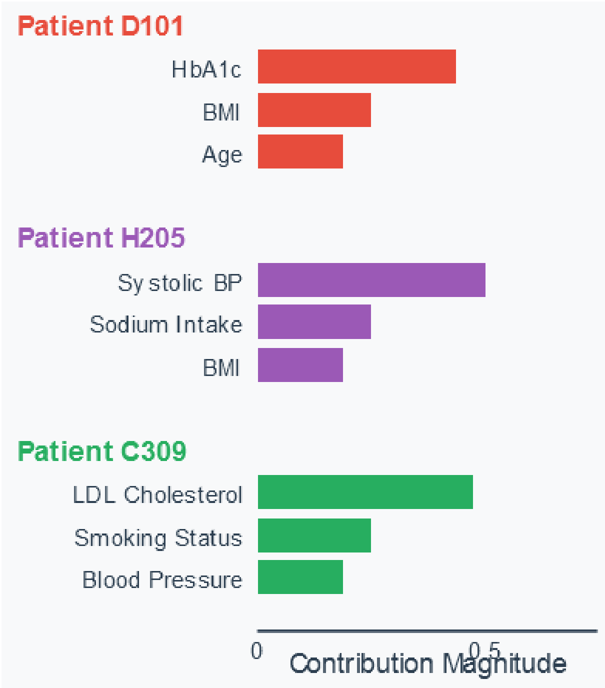

Figure 3: LIME explanations for selected patients across the three conditions

A 55-year-old patient with prediabetes was analyzed as a case study. The digital twin simulation suggested a combination of lifestyle modifications and metformin therapy. Over six months, the patient exhibited a 12% reduction in HbA1c along with improvements in weight and blood pressure. Table 3 shows predicted versus observed outcomes for this patient and others.

TABLE 3: PATIENT OUTCOMES UNDER RECOMMENDED INTERVENTIONS

Disease	Patient ID	Intervention	Predicted Improvement (%)
Diabetes	D101	Metformin + Diet	12
Hypertension	H205	Exercise + Medication	10
Cardiac Risk	C309	Statin + Lifestyle Change	15

Figure 4 illustrates the predicted versus observed outcomes for the diabetic patient, highlighting the effectiveness of personalized recommendations.

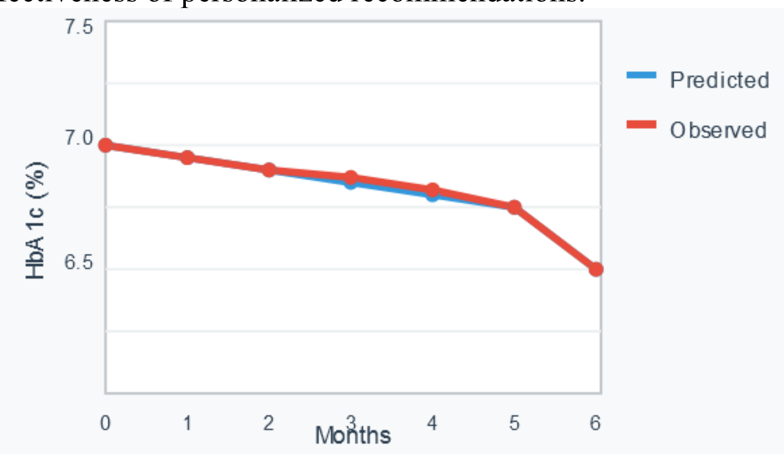

Figure 4: Predicted vs. observed treatment outcomes for a representative diabetic patient.

The framework was tested across multiple conditions without performance degradation. Tables 1–3 provide predictive metrics, feature contributions, and outcomes for all three diseases, while Figures 2 and 3 illustrate interpretability across diseases and patients, respectively, supporting both generalizable insights and patient-specific recommendations.

VI. Conclusions

The proposed Explainable and Expandable AI framework is able to fill the existing gap between predictive models, mechanistic understanding, and a personalized healthcare. By integrating multi-omics and electronic health records data, wearable data into the framework, it will be possible to predict the right outcomes and provide a clear understanding of the factors that influence the development of the disease. The new combination of LIME and SHAP gives localized patient-aware descriptions and feature significance to enhance clinician faith and make wise choices.

Besides, digital twin simulator support including mechanistic mapping layer enables it to deliver treatment outcomes with particular attention to patients, and predict and optimize treatment outcomes. Generally, the framework has shown high level of scalability in various situations such as diabetes, hypertension and cardiac risk can generate realistic information that can be used to make informed decisions based on therapy. The work has the ability to transform precise healthcare in the world since it could integrate prediction, explainability and personalization in the same architecture thereby making it feasible to initiate data-driven, patient-centered interventions and attain improved clinical outcomes.

References

[1] Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2021). Leveraging SHAP and LIME for enhanced explainability in AI-driven diagnostic systems. *International Journal of AI and ML*, *2*(3).

[2] Tuan, D. A. (2024). Bridging the gap between black box AI and clinical practice: Advancing explainable AI for trust, ethics, and personalized healthcare diagnostics.

[3] Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) in healthcare: interpretable models for clinical decision support. *Available at SSRN 4637897*.

[4] Talukder, N. (2024). *Clinical decision support system: an explainable AI approach* (Master's thesis, N. Talukder).

[5] Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Angulakshmi, M. (2025). A Literature review on applications of explainable artificial intelligence (XAI). *IEEE Access*.

[6] Mehta, V., Garg, J., Sharma, D., Yadav, S., & Kumar, R. (2024, December). An Explainable Artificial Intelligence Based Approach for Prediction and Classification of Multiple Sclerosis Disease: A Mental Health Disorder. In *International Advanced Computing Conference* (pp. 168-180). Cham: Springer Nature Switzerland.

[7] Vimbi, V., Shaffi, N., & Mahmud, M. (2024). Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*, *11*(1), 10.

[8] Aldoseri A, Al-Khalifa KN, Hamouda AS. Re-think data strategy and integration for artifcial intelli-gence: concepts, opportunities and challenges. 2023;

[9] Chaddad A, Peng J, Xu J, Bouridane A. Survey of explainable AI techniques in healthcare. Sensors. 2023;23:634. MDPI.

[10] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. 2017

[11] Sabol P, Sinčák P, Hartono P, Kočan P, Benetinová Z, Blichárová A, et al. Explainable classifer for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. J Biomed Inform. 2020;1:109.

[12] Bhandari N, Walambe R, Kotecha K, Kaliya M. Integrative gene expression analysis for the diagnosis of Parkinson's disease using machine learning and explainable AI. Comput Biol Med. 2023;1:163.

[13] Alsaleh MM, Allery F, Choi JW, Hama T, McQuillin A, Wu H, et al. Prediction of disease comorbidity using explainable artifcial intelligence and machine learning techniques: a systematic review. Int J Med Inform. 2023;175:105088. Elsevier Ireland Ltd.

[14] Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artifcial intelligence in medical & healthcare domains: a systematic literature review. Comput Biol Med. 2023;166:107555. Elsevier Ltd.

[15] Sreeja MU, Philip AO, Supriya MH. Towards explainability in artifcial intelligence frameworks for heartcare: a comprehensive survey. J King Saud Univ Comput Inf Sci. 2024;36:102096. King Saud bin Abdulaziz University.

[16] Vimbi V, Shaf N, Mahmud M. Interpreting artifcial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Inform. 2024;11:10. Springer Science and Business Media Deutschland GmbH.

[17] Elkhawaga G, Elzeki O, Abuelkheir M, Reichert M. Evaluating explainable artifcial intelligence methods based on feature elimination: a functionality-grounded approach. Electronics (Switzerland). 2023;12(7):1670.

[18] Tasin I, Nabil TU, Islam S, Khan R. Diabetes prediction using machine learning and explainable AI techniques. Healthc Technol Lett. 2023;10(1–2):1–10.

[19] Islam MS, Hussain I, Rahman MM, Park SJ, Hossain MA. Explainable artifcial intelligence model for stroke prediction using EEG signal. Sensors. 2022;22(24):9859.

[20] Abir WH, Uddin MF, Khanam FR, Tazin T, Khan MM, Masud M, et al. Explainable AI in diagnosing and anticipating leukemia using transfer learning method. Comput Intell Neurosci. 2022;2022:5140148.

[21] de Lima AD, Lopes AJ, do Amaral JLM, de Melo PL. Explainable machine learning methods and respiratory oscillometry for the diagnosis of respiratory abnormalities in sarcoidosis. BMC Med Inform Decis Mak. 2022;22(1):274.

[22] Otaki Y, Singh A, Kavanagh P, Miller RJH, Parekh T, Tamarappoo BK, et al. Clinical deployment of explainable artifcial intelligence of SPECT for diagnosis of coronary artery disease. JACC Cardiovasc Imaging. 2022;15(6):1091–102.

[23] Gabralla LA, Hussien AM, AlMohimeed A, Saleh H, Alsekait DM, ElSappagh S, et al. Automated diagnosis for colon cancer diseases using stacking transformer models and explainable artifcial intelligence. Diagnostics. 2023;13(18):2939.
[24] Peng J, Zou K, Zhou M, Teng Y, Zhu X, Zhang F, et al. An explainable artifcial intelligence framework for the deterioration risk prediction of hepatitis patients. J Med Syst. 2021;45(5):61.
[25] Junaid M, Ali S, Eid F, El-Sappagh S, Abuhmed T. Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. Comput Methods Programs Biomed. 2023;1:234.
[26] Maouche I, Terrissa LS, Benmohammed K, Zerhouni N. An explainable AI approach for breast cancer metastasis prediction based on clinicopathological data. IEEE Trans Biomed Eng. 2023;70:3321–9.
[27] El-Sappagh S, Alonso JM, Islam SMR, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explainable artifcial intelligence for Alzheimer's disease. Sci Rep. 2021;11(1):2660.
[28] Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep. 2020;10(1):19549.
[29] Tan W, Guan P, Wu L, Chen H, Li J, Ling Y, et al. The use of explainable artifcial intelligence to explore types of fenestral otosclerosis misdiagnosed when using temporal bone high-resolution computed tomography. Ann Transl Med. 2021;9(12):969–969.
[30] Magesh PR, Myloth RD, Tom RJ. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. Comput Biol Med. 2020;1:126.
[31] Das S, Sultana M, Bhattacharya S, Sengupta D, De D. XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classifcation using explainable AI. J Supercomput. 2023;79(16):18167–97.
[32] Sekaran K, Varghese RP, Gopikrishnan M, Alsamman AM, El Allali A, Zayed H, et al. Unraveling the dysbiosis of vaginal microbiome to understand cervical cancer disease etiology—an explainable AI approach. Genes (Basel). 2023;14(4):936.
[33] Gunashekar DD, Bielak L, Hägele L, Oerther B, Benndorf M, Grosu AL, et al. Explainable AI for CNN-based prostate tumor segmentation in multiparametric MRI correlated to whole mount histopathology. Radiat Oncol. 2022;17(1):65.
[34] Bhandari M, Shahi TB, Siku B, Neupane A. Explanatory classifcation of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI. Comput Biol Med. 2022;1:150.
[35] Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artifcial intelligence model for acute kidney injury prediction. Nat Commun. 2020;11(1):5668.
[36] Abbas, S. R., Seol, H., Abbas, Z., & Lee, S. W. (2025, July). Exploring the role of artificial intelligence in smart healthcare: a capability and function-oriented review. In *Healthcare* (Vol. 13, No. 14, p. 1642). MDPI.
[37] Massafra R, Fanizzi A, Amoroso N, et al. Analyzing breast cancer invasive disease event classifcation through explainable artifcial intelligence. Front Med. 2023;10:1116354.
[38] Fellous JM, Sapiro G, Rossi A, Mayberg H, Ferrante M. Explainable artifcial intelligence for neuroscience: behavioral neurostimulation. Front Neurosci. 2019;13:13.
[39] Ramírez-Mena A, Andrés-León E, Alvarez-Cubero MJ, Anguita-Ruiz A, Martinez-Gonzalez LJ, Alcala-Fdez J. Explainable artifcial intelligence to predict and identify prostate cancer tissue by gene expression. Comput Methods Programs Biomed. 2023;1:240.
[40] Sharmam, H., Pradhan, R., Mogha, H., Singh, S., & Kumar, R. (2025, April). Counterfactual AI in Healthcare: Enhancing Decision-Making and Outcome Prediction. In *2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT)* (pp. 608-615). IEEE.
[41] Rajendran, A., Angelin Rajan, R., Balasubramaniyam, S., & Elumalai, K. (2025). AI-Enhanced Predictive Imaging in Precision Medicine: Advancing Diagnostic Accuracy and Personalized Treatment. *Iradiology*, *3*(4), 261-278.