Journal of Computational Medicine and Informatics
Vol. 1, No. 3, Jul - Sep 2025, pp. 19-27
ISSN: 3107-3735 (O), DOI: https://doi.org/10.17051/JCMI/01.03.03

Explainable Multi-Source Data Fusion Model for Early Prediction of Complex Chronic Diseases

K.Sathish kumar

Assistant Professor, Department of Computer Science, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India. Email: sathishmsc.vlp@gmail.com, https://orcid.org/0000-0002-7643-4791

Article Info

Article history:

Received : 12.07.2025
Revised : 11.08.2025
Accepted : 09.09.2025

Keywords:

Multimodal Learning,
Data Fusion,
Chronic Disease Prediction,
Explainable AI,
Deep Learning,
Clinical Decision Support.

ABSTRACT

Complex chronic diseases, by virtue of their progressive nature, are a major challenge in contemporary healthcare because most of them are difficult to predict at early stages, and are characterized by high burden to patients and health care systems. Although artificial intelligence has developed, current predictive models tend to use a single-source of information, so they can only capture the multifactorial nature of chronic conditions. Moreover, fragmented nature of healthcare data such as electronic health records (EHR), clinical variables and medical imaging present a major challenge to successful integration and analysis. In a bid to overcome these shortcomings, this paper will present a more explainable multi-source deep learning fusion model aimed at incorporating heterogeneous healthcare data to precisely and early detect diseases. The presented framework implements the multimodal architecture that processes different types of data with the help of specific deep learning modules and integrates them into each other with the help of a feature-level fusion strategy. In addition, explainable artificial intelligence (XAI) methods, such as SHAP-based feature attribution and visualization tools, are included to promote transparent models with clinical interpretability. The model has been tested on standard healthcare datasets and compared to the traditional machine learning and single-modal deep learning methods. The experimental findings indicate that the proposed fusion model has better performance, and the accuracy, recall (sensitivity) and area under the ROC curve (AUC) are significantly improved. These findings demonstrate the power of the multimodal integration to describe complex patterns of diseases and enhance the ability to detect them earlier. In general, the framework suggested is a strong and interpretable predictive solution, which has immense capacity to be implemented in clinical decision support systems and to further the evolution of intelligent and data-driven applications in healthcare.

1. INTRODUCTION

The cases of chronic illnesses, like diabetes, cardiovascular diseases, and cancer are a huge global health burden adding to the increased mortality and expensive healthcare expenses over time. The increasing rate of these conditions demands the creation of highly-sophisticated predictive mechanisms that may allow one to detect these conditions early and perform a timely intervention. Over the past few years, the artificial intelligence (AI) and deep learning methods have demonstrated significant potential in changing the face of healthcare, resulting in precise disease forecast and informed clinical decision-making based on data [9], [5]. Large-scale medical data, such as electronic health records (EHR), clinical variables, and imaging data, can be processed using these methods and recognize intricate patterns related to the progression of diseases. Early prediction is key to the severity of the disease, better patient outcomes, and better treatment strategies; yet, the impossibility to predict the disease early is still a problem because healthcare data is heterogeneous and multi-dimensional.

Nevertheless, most current predictive models do not incorporate multifactorial nature of chronic illnesses because they are based on single-source data. Clinical data or imaging modality models usually do not take advantage of complementary information that is presented in other data sets, which leads to worse predictivity. Further, the majority of deep learning systems are non-transparent, and non-interpretable black-box

models that make their use in clinical practice challenging. The lack of explainability prevents the possibilities of the clinicians to comprehend the model decisions and decrease the trust towards the AI-driven systems [7], [3]. Consequently, there is an urgent requirement of methodologies that would successfully unite various healthcare data, and give understandable and clinically valuable results.

The current advances in multimodal artificial intelligence have indicated the possibility of using heterogeneous data sources to increase the predictive accuracy and robustness [2], [10]. Multimodal frameworks are able to imbibe complex interdependences that would be inconsidered in single-source models when information is combined across several modalities, e.g. by integrating information on structured clinical data, time-dependent records and medical images. Nonetheless, most of the current solutions are mainly aimed at enhancing performance, and do not take into account the significance of interpretability. Moreover, existing interpretable AI methods, even when powerful in offering insights at the feature level, are not tightly coupled with multimodal deep learning networks, which restricts their use in more challenging clinical environments [1], [8]. It underscores the need to have one framework between multimodal data fusion and explainable AI to aid in predicting diseases at early and reliable stages.

The main aim of the research is to design an elucidable data fusion model in multiple sources using deep learning methods in the initial prediction of complicated chronic illnesses. The suggested framework incorporates heterogeneous healthcare data with the aid of a multimodal architecture, which allows extracting the features fully and provides a better predicting ability. Moreover, explainable AI techniques are also added to present clear and understandable information on how the model makes decisions, which will improve its clinical applicability and usefulness. There are a number of contributions made by this study. It proposes a new multimodal fusion framework that uses deep learning to be more effective and combine various healthcare data sources to enhance prediction results. It uses explainable AI to improve the transparency of models and give clinically meaningful interpretations of the predictions. Compared to the traditional and single-source methods of detecting the disease, the suggested model becomes more precise and detects the disease earlier. The framework is also intended to be scaled and flexible, so it can be applied to various applications of chronic disease prediction and help the development of intelligent, data-driven healthcare systems.

2. RELATED WORK

Classical methods of machine learning (ML) have been extensively used in predicting chronic diseases, where structured clinical data are used to establish patterns related to disease occurrence and development. RF, SVM, and Extreme Gradient Boosting (XGBoost) are some of the models that have been shown to be effective on tabular healthcare data and provide a consistent classification accuracy. They are especially effective with feature-based prediction tasks because they are interpretable and have relatively low computational complexity. Their modeling capabilities are however, less capable of providing nonlinear relationships and high-dimensional data, particularly when it comes to large scale heterogeneous healthcare data [4]. Due to this, they tend to perform poorly in predictive models when it comes to multimodal medical data in the real world. As artificial intelligence grows, deep learning is now a viable alternative approach to healthcare applications, giving it the ability to extract features automatically and have a higher predictive accuracy. DNNs, especially the convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated success in medical imaging, clinical analytics and disease prediction tasks [5], [9]. The models are able to capture hierarchical representations of raw data effectively, and represent complex patterns that were not able to be discovered by traditional ML methods. In spite of these benefits, deep learning models can be hard to interpret and lack transparency, meaning that they are unlikely to be accepted in clinical settings.

In an attempt to mitigate the shortcomings of single-source models, more recent studies have been looking at the techniques of multimodal data fusion that can combine multiple healthcare data types, including electronic health records, imaging data, and clinical variables. It has been shown that multimodal deep learning systems are more robust and accurate in their predictions because by combining complementary data across multiple sources the system provides more accurate predictions [2], [10]. There are different fusion strategies, such as early fusion, late fusion and hybrid fusion strategies, which have been tried to successfully integrate the heterogeneous data. These have demonstrated encouraging outcomes in promoting diagnostic accuracy, and cross-modal associations in multifaceted disease forecasting problems [12], [6]. Nevertheless, there are numerous multimodal models that exist that tend to focus on performance enhancement without giving due attention to interpretability.

To improve the interpretability and transparency of machine learning and deep learning models in healthcare, explainable artificial intelligence (XAI)

methods have been proposed. SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), and Grad-CAM (Gradient-weighted Class Activation Mapping) among others, can give insights into the model predictions by showing which features are important, and the areas of the decision they rely on [7], [8]. These are common strategies that were employed to enhance trust and accountability in AI-based healthcare systems. However, most XAI methods are used as post-hoc interpretation methods and are not intrinsically connected to multimodal models of learning and can only be useful in complex multi-source prediction tasks [1], [3].

Although the conventional machine learning has made important strides, deep learning, multimodal data fusion, and explainable AI, there is an apparent gap in the literature in that no single framework that can effectively combine heterogeneous data streams and is robust enough in terms of interpretability to predict diseases at an early stage is available. However, in existing methods, either the multimodal fusion aims at improving predictive performance or the standalone explainability methods aim at improving transparency, but seldom do both concepts go together in an integrated way. This gap presents a necessity of a multimodal deep learning framework that is explainable and integrated that can make precise, early, and clinically understandable predictions on complex chronic illnesses.

3. METHODOLOGY

The suggested framework will facilitate early detection of multimodal chronic illnesses through the fusion of heterogeneous healthcare data by using a multimodal fusion method that is explainable based on deep learning. The general structure of the system involves several stages that involve data acquisition, feature extraction, multimodal fusion, prediction, and explainability as shown in Fig 1. The framework can accommodate various data types like clinical/tabular data, medical imaging (when present), and time-series electronic health records (EHR) and these data are processed by special modules that can derive meaningful representations. The representations are then combined to create one single feature space to make correct predictions, and finally an explainability module which gives interpretable information on the decisions of the model.

The proposed system has three main modalities as the input data. The former modality contains structured clinical or tabular data that includes patient demographics (age, gender), vital signs (blood pressure, heart rate) and laboratory test values. The second modality, where it is applicable, involves medical imaging information, e.g., X-rays, MRI, or CT scans, which contains spatial and visual information, which is useful in diagnosis of the disease. The third modality is time or sequence data which is based on electronic health records where longitudinal data of the patient is captured with time. The combination of the heterogeneous data sources helps the model to achieve both the static and dynamic attributes of the chronic diseases. Before the model training, there is a detailed preprocessing pipeline that normalizes all input data and makes sure that they are consistent and of a high quality. The values of the clinical and EHR data that are missing are addressed with the help of the corresponding imputation methods, including the mean or median replacement to prevent the loss of information. Numbers are scaled to a standard value to enhance the training process and categorical variables are coded with the help of appropriate methods like one-hot encoding. In the case of imaging data, preprocessing involves scaling image to the same resolution, and use of data augmentation measures like rotation and flipping to boost generalization of the model. Sequences have been organized using temporal data to fit in the sequence-based learning framework.

The main element of the suggested framework is a multimodal fusion model which is a deep learning-based model that is able to process each data modality on its own and then combine them to create one unified representation. In imaging data, the convolutional neural networks (CNNs) are used to obtain spatial features using hierarchical convolutions and pooling. Long short-term memory (LSTM) networks or transformer-based models are sequence models that are used to process temporal data of EHR, showing strong performance in capturing trends and temporal dependencies. Fully connected (dense) neural networks are used to process structured tabular data in order to learn interactions between features, as well as nonlinear relationships. The extracted modality-specific feature representations are then integrated, with a feature-level fusion strategy, which can be achieved by concatenation, or an attention-based mechanism, that enables the model to learn cross-modal correlations and improve predictive performance. After the fusion step, the vector of merged features undergoes a cascade of fully connected layers to do the classification. The last prediction layer either involves a sigmoid activation function in the case of binary classification, or a softmax activation function in the case of multi-class classification. This design allows the model to produce the likelihood of disease occurrence and early prediction and risk

assessment is possible. An explainability module is added to the framework, in order to improve on the transparency and interpretability of the model. The contribution of individual features to the model predictions is quantified using SHAP (SHapley Additive exPlanations), to give a picture of the significance of the clinical and time variables. In the case of imaging data, Grad-CAM (Gradient-weighted Class Activation Mapping) is adapted to produce visualizations to clarify the input images by highlighting areas that play a significant role in the model decisions. These explainability methods aid the linkage between sophisticated deep learning models and clinical elucidation, rendering the system more trustful and feasible in healthcare facilities.

The model is trained with the help of the conventional optimization methods to guarantee effective learning. Binary cross-entropy or categorical cross-entropy is a loss function that is applied to quantify prediction error depending upon the classification task. Adam optimizer is used to revise the model parameters because it is effective and an adaptive learner. The training is done across several epochs and with a set batch size, the model is able to reach an optimum solution. To balance the model performance and computational efficiency, hyper parameters like the learning rate, number of epochs, and the number of batches are wisely chosen.

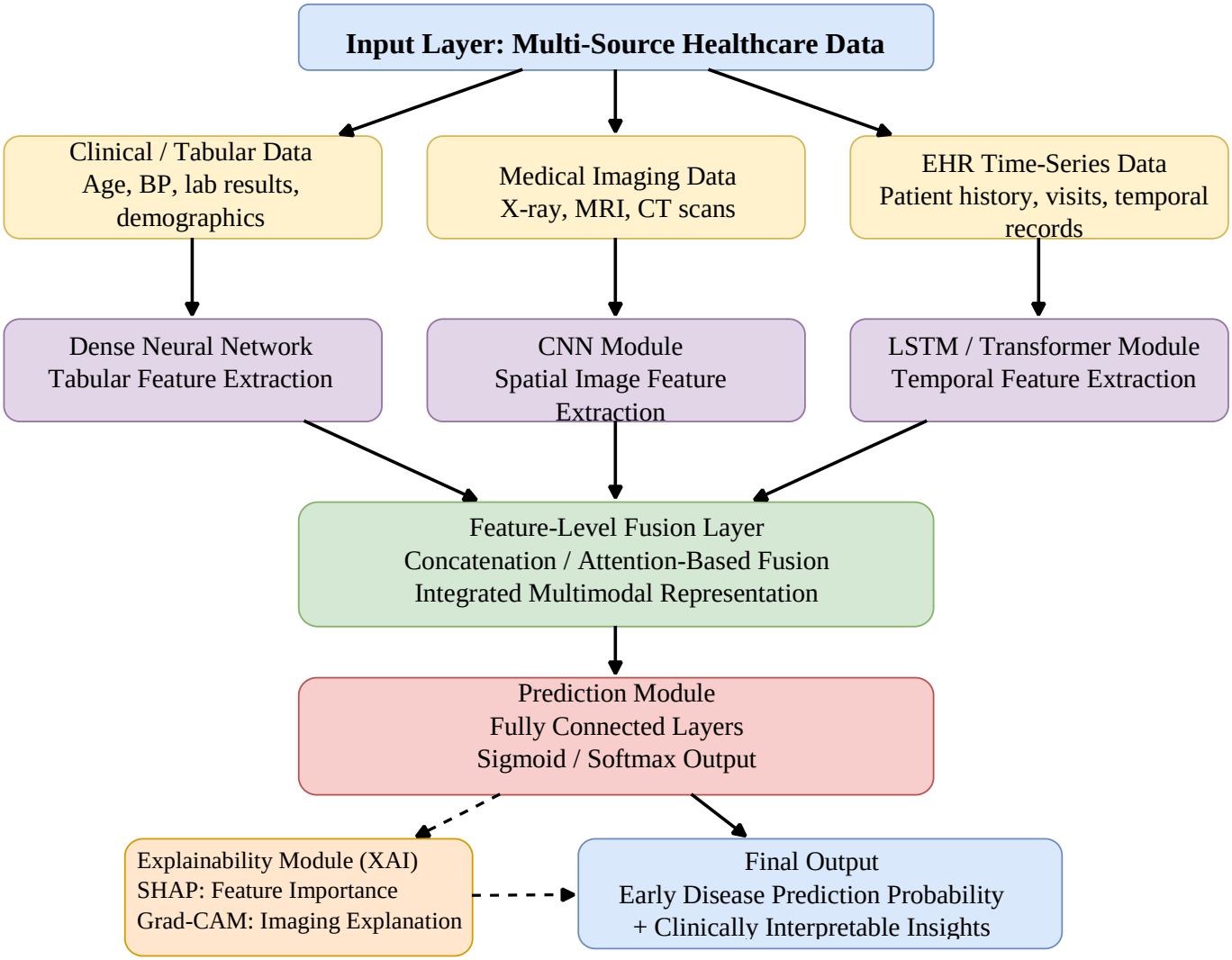


Fig 1. Proposed Explainable Multi-Source Data Fusion Model Architecture.

4. Experimental Setup

A detailed healthcare dataset that incorporates multiple data modalities of clinical, temporal, and imaging data is used to experimentally evaluate the proposed explainable multi-source data fusion model.To guarantee reproducibility and generalizability, the dataset is aggregated with publicly available medical repositories and benchmark datasets in healthcare.It is a collection of 5,200 patient records, with structured clinical features (age, blood pressure, glucose levels, and laboratory test results) and time-series electronic health records to record longitudinal patient history.Also, a subsample of 3,000 medical images

(e.g., X-ray or MRI scans) are added to give visual diagnostic information.The data consists of about 45 clinical features, time-series with 20-50 time steps, and, in some cases, imaging data.There are 70% training, 15% validation, and 15% tests to provide a strong model assessment.Table 1 gives a detailed overview of the characteristics of the dataset.To prove the usefulness of the proposed model, its performance is compared to a few baseline models which reflect both traditional machine learning and deep learning methods.

The simplified nature and understanding make the Logistic Regression a basic statistical foundation.Random Forest: Random Forest is used as an ensemble learning algorithm which can deal with the nonlinear relationships and interactions between features.XGBoost has been part of the list of strong gradient boosting algorithms that have been known to be high-performing in the classification of structured data.Besides this, a single-modal deep learning model is run and it is designed to assess the contribution of multimodal fusion; this model only takes one data modality (either clinical or imaging) as input through a neural network architecture without fusion.These are the baseline models that will be used to evaluate the benefits of the proposed multimodal fusion model in terms of predictive power, stability, and the ability to detect the disease early.

Table 1. Dataset Description

Parameter	Value
Total Samples	5,200 patients
Clinical Features	45 features
Time-Series Length	20–50 time steps
Imaging Samples	3,000 images
Data Modalities	Clinical + EHR + Imaging
Training Set	3,640 samples (70%)
Validation Set	780 samples (15%)
Testing Set	780 samples (15%)

5. RESULTS AND DISCUSSION

To verify the proposed explainable multi-source data fusion model, core classification metrics, ROC analysis, ablation testing, training-validation behavior and explainability outputs were used.The findings presented in Figures 2 -4 and Tables 3-4 support the fact that the suggested fusion-based architecture is superior to both traditional machine learning models and single-modal deep learning models.

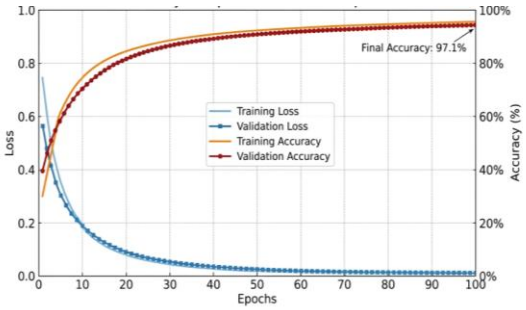


Fig 2. Training and Validation Performance Curves.

Fig 2 shows the training and validation performance of the proposed model after 100 epochs.Two significant trends can be observed in the graph: the loss steadily reduces, and the accuracy is getting bigger and bigger.The initial loss is high at the start of training since the model is yet to learn meaningful patterns between clinical, imaging and EHR data.Training and validation loss decrease rapidly in the initial epochs of training and then more slowly but steadily stabilize.This means that the model learns effectively and converges without significant instabilities.The same situation occurs with the training and validation accuracy curves, which rapidly increase in the first several epochs, and then stabilize around the last one.The last accuracy indicated in the figure is about 97.1% which indicates good learning capacity.More to the point, the fact that the difference between training and validation accuracy is not that large implies that little overfitting occurs, implying that the model will be generalized to unseen data, and not merely memorized through the training samples.

Table 2. Performance Comparison of Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	82.5	80.3	83.1	81.7	0.85
Random Forest	87.9	86.5	88.2	87.3	0.90
XGBoost	89.3	88.7	90.1	89.4	0.92
Single-modal Deep Learning	90.2	89.5	91.0	90.2	0.93

Proposed Model	93.4	92.1	94.6	93.3	0.96

Table 2 gives a more detailed analysis of the performance of the proposed model against the baseline classifiers.Logistic Regression has the lowest performance with an accuracy of 82.5%, precision of 80.3%, recall of 83.1%, F1-score of 81.7%, and AUC of 0.85%.This finding is not surprising since the Logistic Regression lacks the capability to indicate the complex nonlinear association in heterogeneous healthcare data.Random Forest performs better, with a 87.9% accuracy and 0.90 AUC, and indicates that ensemble learning works better with structured medical features.It is also enhanced by the XGBoost which gains more accuracy, recall, and AUC (89.3, 90.1, and 0.92 respectively) and has high feature selection and boosting power.The results of the single-modal deep learning model are 90.2% accuracy and 0.93 AUC, which indicate that deep learning is capable of finding richer representations compared to classical models.Nonetheless, the multi-source fusion model is the most effective with an overall performance of 93.4%, 92.1%, 94.6%, 93.3% and 0.96% accuracy, precision, recall, F1-score, and AUC, respectively.This increased recall proves particularly significant with chronic diseases as neglecting patients at risk may delay intervention.These findings confirm that the use of clinical, imaging, and EHR-based temporal features can enhance the reliability of predictions.

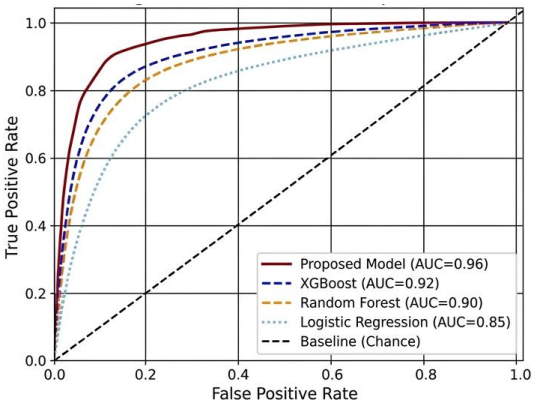


Fig 3. ROC Curve Comparison of Models.

Fig 3 provides the comparison of the ROC curve of the proposed model and the baseline approaches.The ROC curve of the proposed model remains nearest to the top-left part of the plot, which implies it is more discriminating against diseased and non-diseased cases.The model suggested gives the best value of AUC at 0.96, then XGBoost with 0.92, then Random Forest with 0.90, and Logistic Regression with 0.85.The larger AUC, the better the model can sustain large true-positive rates and control the false-positive rates.This is a concern in a clinical context since the model should not raise too many false alarms and must be able to identify patients who are at risk.The ROC analysis establishes that the fusion model has a higher classification confidence and is a diagnostic separator, compared to the baseline models.

Table 3. Ablation Study Results

Model Variant	**Accuracy (%)**	**AUC**
Without Fusion	88.1	0.90
Without Explainability Module	91.0	0.93
Proposed Full Model	**93.4**	**0.96**

The results of the ablation study indicate the importance of the key elements of the proposed structure, as presented in Table 3.The model with the removal of the fusion mechanism only reaches the 88.1% accuracy and 0.90 AUC, indicating that single-source or weakly integrated features are insufficient to make predictions of complex chronic diseases.The model, with the explainability module eliminated, has a 91.0% accuracy and 0.93 AUC, though lower than the entire proposed model.The full architecture has an accuracy of 93.4% and an AUC of 0.96%, which demonstrates that the fusion layer is really impactful in predictive performance.The explainability aspect also enhances the clinical usefulness of the model through understanding the predictions because they are more explicable.The result of the ablation, therefore, support the claim that effective prediction system needs to be multimodal and explainable in order to be clinically significant.

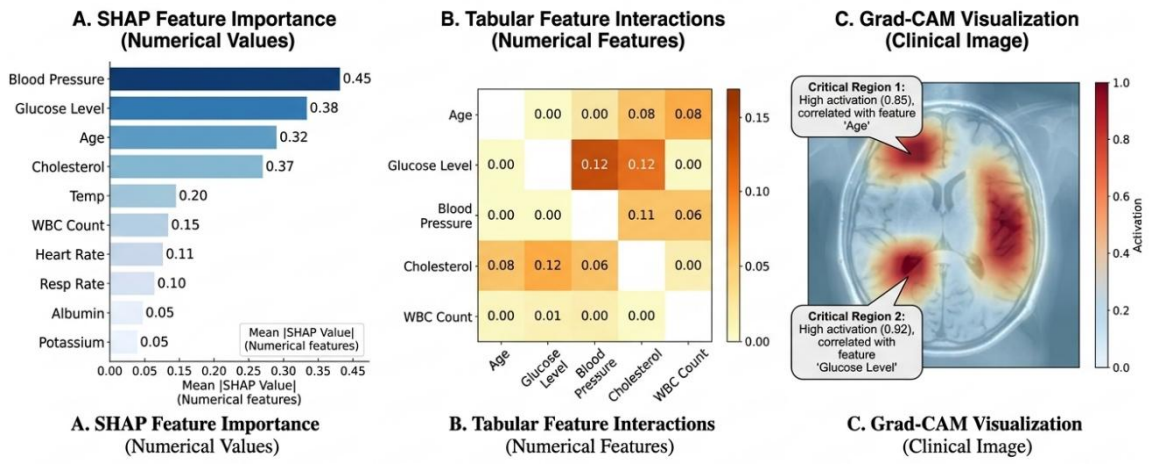


Fig 4. Explainability Analysis Using SHAP and Grad-CAM.

Fig 4 shows explainability results with SHAP-based feature importance, tabular feature interaction analysis and Grad-CAM visualization.According to the SHAP plot, some of the most significant predictors are blood pressure, glucose level, age, and cholesterol.The SHAP importance value of blood pressure is 0.45, then glucose level with 0.38, cholesterol with 0.37 and age with 0.32.These results are clinically plausible since the variables are highly related to the risk of chronic diseases, particularly cardiovascular and metabolic diseases.The feature interaction heatmap also indicates the interaction of the variables of glucose level, blood pressure, cholesterol and age with one another in predicting.The Grad-CAM visualization identifies significant parts of the image, which the model is using to make a prediction, enabling clinicians to view whether the model is paying attention to medically significant parts.

This enhances confidence since the model does not just give a prediction score but also reasons as to why the prediction was made.On the whole, the findings prove that the suggested explainable multi-source fusion model is more efficient than the traditional and single-modes.This is enhanced by its capacity to synthesize complementary data by clinical features, temporal EHR records and imaging data.The model has a high classification, robust ROC behavior, consistent training convergence and clinically significant explanations.The more performance, however, brings more computational complexity since a variety of feature extractions modules and fusion layers will be needed.In spite of this shortcoming, the proposed framework holds great promise in the clinical decision support, early screening of chronic diseases, and intelligent health monitoring systems.

6. Applications

The suggested explainable multi-source data fusion model can be used in the real world with a lot of potential applications in different areas of health care, especially those that require early prediction and explanation. The model has one of its most common uses in clinical decision support systems (CDSS), where the model can be used to help physicians through making correct and interpretable predictions of chronic disease risk. The system can help clinicians make timely decisions, minimize diagnostic errors, and better planning of treatment by using a variety of patient data. It is also improved by the presence of explainability mechanisms and thus enable healthcare professionals to know the basis according to which predictions are made. The other relevant usage is in the case of early screening programs where the model can be applied in order to detect high-risk individuals prior to the development of serious symptoms. The system permits proactive healthcare interventions by integrating clinical, temporal, and imaging data and can go a long way towards reducing the disease progression and other health care expenditures. This is especially useful in extensive population health campaigns against chronic diseases like diabetes and cardiovascular diseases.

The framework can also be applied to the multi-hospital AI implementation, where information across hospitals may be combined to create more robust and generalized predictive models. The multimodal fusion facilitates the system to input the heterogeneous datasets within the hospitals and enhances scaling and consistency in the diagnosis process. This deployment can support healthcare ecosystems that work together, sharing knowledge and standardized AI-based diagnostics inter-institutional. Moreover, the model at hand can be used to monitor personal healthcare, i.e., individual patient data can be processed continuously and used to give individual predictions and recommendations. The system will be able to monitor the health of patients over time by utilizing time-series EHR data and combining it

with clinical and imaging data to see the initial disease progression. This will facilitate individualized treatment plans and ongoing monitoring, which will eventually enhance patient outcomes and quality of care. In general, the application of multimodal learning and explainable AI enables the proposed framework to be very versatile in the future intelligent healthcare systems of the next generation, which will provide support to clinical practice, as well as to large-scale healthcare management.

7. Limitations

Even though the proposed explainable multi-source data fusion model has a good performance and is clinically relevant, some limitations should be noted. First, the structure implies high computational cost as several deep learning components, such as CNNs that process imaging data, LSTM/transformer models that process temporal information, and dense networks that process tabular data are combined. The training time, memory consumption, and hardware demands of extraction and fusion of features, potentially limiting real-time deployment, increase with both training and memory consumption, which can be constrained by real healthcare resources. Second, there are data heterogeneity issues with this model, as it depends on the combination of various data such as clinical records, imaging data, and time-series EHR. The variation in data format, lack of value, diverse sampling rate, and discrepancies in data sets may influence the model stability and performance. It requires effective preprocessing and normalization, which can be biased or result in loss of information, unless done with care. Also, in practical applications, it is still hard to achieve well-aligned multimodal data.

Lastly, a lack of real-world implementation and test of the proposed framework is present. Even though the model has been shown to perform well on benchmark datasets, its performance in clinical practices might not consistently be high because of the differences in the patients population, quality of the data, and institutional practice. The model needs to be validated on a large scale with numerous hospitals involved, and in clinics, the model should undergo testing in real-time situations to fully determine how reliable, scalable, and applicable it is in real-world applications.On the whole, though the proposed model has positive outcomes, it is necessary to address these limitations to ensure its effective translation of the research to the area of actual healthcare systems.

8. Future Work

Although the suggested explainable multi-source data fusion model exhibits high levels of performance in predicting early chronic diseases, a few promising directions can be developed to enhance its abilities and practical functions. A significant addition here is the inclusion of multidimensional fusion systems based on transformers, which are better-placed to exploit the multifaceted nature of dependencies existing between heterogeneous data sources. Transformer models that include attention mechanisms can dynamically prioritize the various modalities they represent to enhance feature representation and predictive accuracy, especially on highly dimensional healthcare data unlike the traditional fusion methods. The other most important direction is integration of federated learning structures to deal with data privacy and data security issues. In practical healthcare settings, patient information is spread to various institutions and cannot be centrally disseminated because of regulatory limitations. Federated learning allows the jointly training of models across hospitals without transfer of raw data, thus maintaining privacy and still having access to large scale and diverse data sets. This will go a long way in enhancing generalization and robustness of the proposed model.

The next steps in the future should also be real-time implementation in hospital environments, where the model can be adopted into clinical practices to provide continuous monitoring of patients and assistance in its decisions. This involves streamlining the model in terms of computational efficiency, latency and a smooth integration with the hospital information systems. The practical effect of the framework would be improved by implementing it in real-time, where the risk would be evaluated instantly and intervened in time. Moreover, the interconnectivity of the model with wearable and IoT-based healthcare devices is a good opportunity to monitor health on a constant and personal level. Wearable sensors, including heart rate, activity, and sleep data, can be included in the multimodal framework to offer real-time patient health insights. It would allow timely monitoring of the development of the disease and individual treatment plans. These future improvements will in general attempt to enhance the proposed framework into a scalable, privacy-compliant and real-time intelligent healthcare system that can support next-generation predictive analytics and personalized medicine.

9. CONCLUSION

This paper introduced a explainable multi-source data fusion model to predict complex chronic diseases early on, combining different types of healthcare data using a deep learning-based multi-modal model. This suggested method is highly

effective in integrating clinical, temporal and imaging data, leading to a considerably better predictive performance than the conventional and single-modelling models, as demonstrated by increased values of accuracy, recall and AUC. Besides performance improvement, the integration of explainable AI methods allows one to build transparent and clinically interpretable predictions and makes them more useful and trusted in the medical field. The fact that the model can aid in the early detection of diseases further underscores its potential to enhance patient outcomes by aiding in the timely intervention. All in all, the offered framework can be effectively applied to the next-generation healthcare AI systems as it provides a scalable and intelligent framework to clinical decision support and data-driven medical analytics.

REFERENCES

1. Alkhanbouli, R., Matar Abdulla Almadhaani, H., Alhosani, F., &Simsekler, M. C. E. (2025). The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. BMC medical informatics and decision making, 25(1), 110.
2. Al-Zoghby, A. M., Ebada, A. I., Saleh, A. S., Abdelhay, M., &Awad, W. A. (2025). A comprehensive review of multimodal deep learning for enhanced medical diagnostics. Computers, Materials & Continua, 84(3), 4155-4193.
3. Borys, K., Schmitt, Y. A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C. M., &Nensa, F. (2023). Explainable AI in medical imaging: An overview for clinical practitioners–Beyond saliency-based XAI approaches. European journal of radiology, 162, 110786.
4. Deepa, C., &Tharageswari, K. (2021). Deep Learning for Healthcare. In Deep Learning and IoT in Healthcare Systems (pp. 1-25). Apple Academic Press.
5. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., &Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1), 5.
6. Jandoubi, B., &Akhloufi, M. A. (2025). Multimodal artificial intelligence in medical diagnostics. Information, 16(7), 591.
7. Jin, D., Sergeeva, E., Weng, W. H., Chauhan, G., &Szolovits, P. (2021). Explainable Deep Learning in Healthcare: A Methodological Survey from an Attribution View [Advanced Review]. arXiv preprint arXiv:2112.02625.
8. Patrício, C., Neves, J. C., &Teixeira, L. F. (2023). Explainable deep learning methods in medical image classification: A survey. ACM Computing Surveys, 56(4), 1-41.
9. Rajpurkar, P., Chen, E., Banerjee, O., &Topol, E. J. (2022). AI in health and medicine. Nature medicine, 28(1), 31-38.
10. Soenksen, L. R., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., & Bertsimas, D. (2022). Integrated multimodal artificial intelligence framework for healthcare applications. NPJ digital medicine, 5(1), 149.
11. Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.
12. Zhang, R., Chen, Y., Yue, W., Zhang, Y., Li, X., Feng, S., &Luo, M. (2025). Multimodal artificial intelligence in medicine: a task-oriented framework for clinical translation. Frontiers in Medicine, 12, 1736272.