Grenze International Journal of Engineering and Technology, June Issue

Interpretable Deep Fusion Network for Multi-Class Heart Disease Risk Prediction using Attention and SHAP Explainability on Synthetic Dataset

Chaithra C S[1] and Siddesha S[2]
[1-2]Department of Computer Applications, JSS Science and Technology University, Mysuru

***Abstract*— Heart disease is one of the leading causes for earliest death in young adults in the whole world. To aid accurate and early risk identification is very much necessary to reduce the heart attack risk and mortality. This study proposes a novel deep learning architecture which has a fusion of three different mechanism Multilayer Perceptron (MLP), Bidirectional Long Short-Term Memory (BiLSTM) and attention mechanism to perform multiclass heart disease risk classification.**

The fusion model trained on synthetically generated data of 10000 records containing real-world attributes which includes patient's lifestyle, demographic, and history and clinical factors to classify individual into three risk categories i.e. low, moderate and high. Attention-based visual analysis is integrated and SHAP (Shapley Additive exPlanations) to identify the crucial features which drives for model predictions. The model is evaluated based on three synthetically generated dataset which has noise, real world inspired clinical variables. The proposed model achieves across datasets with accuracy ranging from 95.58% and above 98% AUC. The combination of prediction and explainability makes the model suitable for real world clinical decision support.

***Keywords*— Heart Disease Prediction, Synthetic Data, BiLSTM, SHAP, Explainable AI, Attention Mechanism, Risk Stratification.**

I. INTRODUCTION

Globally cardiovascular disease (CVD) continues to be the leading cause of mortality. Although many studies proven that predictive analysis has gain much popularity. For early risk prediction, there is need of high performance and clinical transparency. Traditional models like Random Forest and Logistic Regression are not deep enough to identify the intricate patterns. Despite their strength the deep learning models is criticized for being opaque.

Preventive measures and medicine could be completely transformed by the application of Artificial Intelligence (AI) in healthcare especially in cardiology field. However, using interpretable black-box models impedes the clinician trust. For decision support systems must not only be effective and ethical it should also be transparent and explainable to integrate into real-life clinical workflows. In multiclass classification tasks and risk stratification of patients this becomes important in which small variations in predictions can have a big impact on treatment planning.

Grenze ID: 01.GIJET.12.2.11_25

Research studies have shown that patient data can be strongly effective and strong basis for Machine Learning (ML) for early detection and risk prediction when the dataset in enhanced with behavioral, clinical and demographic indicators. most current methods or studies treat prediction as binary classification problem i.e. heart disease or no heart disease. This approach or prediction doesn't capture the complex reality in which the patients may be at different level of risks. Therefore, multiclass approach which differentiates between low, moderate and high-risk level is more clinically relevant. Furthermore, a significant obstacle in ML methods is unbalanced dataset. In high-risk categories, models trained on skewed data often perform poorly on minority classes and favor majority classes which can result in harmful false negatives. Our study evaluates the model on three distinct dataset which is synthetic dataset configurations that is balanced dataset and combined dataset to check the robustness of the model.

The proposed study suggests that an interpretable deep fusion model that combines three essential components to narrow the gap between clinical trust and predictive performance. The three components include Attention Mechanism, MLP and BiLSTM to address these issues. To foster the clinical trust, the suggested model is validated across different synthetically generated dataset and incorporates attention-based interpretability techniques and SHAP. Our main contribution consists of an innovative deep learning model for multi-class risk prediction that is enhanced by attention including SHAP explanations for local and global interpretability and validation on three distinct datasets to check the model robustness.

II. Related Work

Traditional machine learning methods were foundation of early research in detection of heart disease. Advances in synthetic data generation have been driven due the scarcity, sensitivity and privacy concern of medical data. In systematic review of generative AI across various medical modalities, Ibrahim et al. [1] described developments in GANs, VAEs, and diffusion models for producing clinically realistic synthetic datasets while resolving ethical and legal issues. To synthesize multimodal electronic health records, Zhong et al. [2] proposed a predictive diffusion model that integrates temporal embeddings with a U-Net architecture, achieving high fidelity, utility and privacy on datasets like MIMIC-III and a multi-institutional cancer trial. These studies demonstrate how synthetic data is increasingly being used to overcome sample size limitations, improve model generalizability, and facilitate privacy-preserving research.

Early models for predicting heart disease made use of domain-driven intelligence, fuzzy logic, and machine learning. On the Cleveland dataset, Javeed et al. [3] created a random search-optimized random forest that produced an accuracy of 93.33% and an AUC of 0.947. According to through surveys like Alizadehsani et al. [6], ML-based models for diagnosing coronary artery disease consistently obtained an AUC>0.90 across a range of cohorts. These studies were fundamental pipeline for accuracy and transparency they were not integrated with contemporary deep learning and synthetic data techniques.

Performance limits in healthcare analytics have been pushed by recent deep architectures. To diagnose congestive heart failure automatically from ECG signals, Acharya et al. [4] used a deep CNN, achieving 99.01% specificity, 98.87% sensitivity and an accuracy is neatly 99%. For electronic health records [EHR] Rajkomar et al. [5] given a scalable deep learning systems that performed better than or equal to specialists in patient outcome prediction. for the purpose of predicting heart disease, hybrid CNN-LSTM architectures, were developed by Sudha and Kumar [14], successfully captured spatial-temporal patterns. Ahamed et al. [12] refined explainable AI-enhanced models to increase predictive performance and transparency, while Rahmathullan et al. [13] created a novel deep learning-powered cardiovascular prediction framework for early diagnosis. Even the accuracy surpassed 95%, most models only used real-world data which limited their ability to generalize to different data distributions.

In clinical settings, interpretability in AI has emerged as a crucial necessity. In their review of deep EHR analytics, Shickel et al. [7] emphasized the importance of explainable models. Post-hoc explanations of predictions are possible without compromising accuracy using model-agnostic frameworks like Integrated Gradients and SHAP [8]. Explainable AI was further emphasized by Holzinger et al. [9] and Ghassemi et al. [10] as being crucial for adoption in high-stakes domains.

Interpretability and prediction is enhanced by hybrid designs and multitask learning. Zhang et al. showed hybrid attention HER models with an AUC>0.92, Miotto et al. [11] examined deep learning possibilities for multi-modal healthcare data. Explainable cardiovascular models were refined by Ahmad et al. [12] and Rahmathullah et al. [13] to enhance recall in high-risk patients. Techniques for integrating interpretability and deep learning in medical image and HER analysis are consolidated in surveys by Chaddad et al. [15]. Despite of these

developments, most current research focusses on either interpretability, predictive modeling, or the synthetic data separately.

Notable developments in heart disease prediction through interpretability, data augmentation, and predictive modeling are highlighted in the reviewed literature. However, current approaches frequently handle these elements separately, concentrating on either increasing model explainability, producing synthetic data or improving accuracy without combining them into a single, cohesive framework. This study fills this gap by presenting a unified pipeline that combined rule-based multi-class risk labelling for clinically meaningful categorization, SHAP-guided feature selection to improve synthetic data generation using CTGAN and diffusion models to overcome class imbalance, and custom MLP-BiLSTM-Attention fusion architecture that jointly captures temporal dependencies, dense representations, and attention-based interpretability. The goal of this integration is to improve heart disease risk prediction's explainability and robustness in practical clinical settings.

III. Dataset Creation

To replicate real-world clinical and behavioral data related to heart disease risk assessment, this study used synthetically generated datasets. To ensure the class separability and real-world type of data, the datasets were produced using specialized python-based generation pipelines using conditional sampling, statistical diversity and domain expertise. We used python 3.10, NumPy for statistical noise and numerical sampling, Pandas for structured data assembly and rules to model feature dependencies were used to create all datasets and it is encoded into .xlsx format using Pandas. Features in various categories are included in each patient profile are demographic information like geographical location, sex, age category. Lifestyle and behavior like alcohol use, physical activity, dietary habits and smoking. Additional symptoms like status of chest pain, coughing, fatigue, unconsciousness. Clinical diagnoses include kidney disease, obesity, asthma, diabetes, hypertension, and biomarkers test, echocardiograph values and other parameters.

Labeling approach was done using rules that mirrored clinical guidelines each instance was assigned to one of the three risk classes high risk, low risk and moderate risk. High risk labeling considered as hypertension+T2D, abnormal echo severity, low ejection fraction, smoking/alcohol use. For moderate risk alcohol/smoking and borderline clinical values. For low-risk combination of healthy lifestyle, comorbidities absent and normal vitals and biomarkers.

IV. Description of Dataset

For this study 4 different synthetic datasets as shown in Table 1 were created namely refined synthetic dataset which contains 10000 records each contain 47 features which is considered as first baseline dataset with realistic distribution of classes. Secondly synthetic dataset which is combined with firstly created dataset which has 18000 records with 47 features set. Other dataset is created to replicate the real-world data where low risk is prevalent an unbalanced dataset is created which has 4500 records and 47 features. In the table 2 different dataset with class distribution information is given. The refined synthetic dataset is served as the initial training base to ensure cleaned and balanced learning, to minimize noise and assure feature-label consistency. It removes random feature interactions which are typical in data that us solely GAN-based because it is generated using SMOTE and SHAP-based filtering.

Table I: Different Set of Synthetic Datasets

Sl. No	Dataset Name	Rows	Features	Low Risk	Moderate Risk	High Risk
1	Refined Synthetic Dataset	10000	47	3392	3425	3183
2	Combined Synthetic Dataset	18000	47	6313	5647	6040
3	Unbalanced Dataset	4500	47	3000	1000	500
4	New Synthetic Dataset	6000	47	2201	1666	2133

The second set of datasets named Combined synthetic dataset is created by integrating superior samples from multiple synthetic sources, such as CTGAN-generated and refined subsets. The dataset has enhanced diversity through feature perturbation while maintaining balance across all the classes which replicates a range of patients circumstances and enables the model to select broad, general patterns without overfitting. The third set of data is named as unbalanced dataset which replicates clinical class imbalance in the real world, by using stratified sampling in which high risk cases are few. It replicates the deployment conditions of model without artificial correction or balancing is applied. The fourth set of data is created using SDV library's CTGAN model which

was trained on the refined synthetic dataset to discover realistic medical feature distributions, where classes are balanced and clinical rule-based filters were used to validate the data to ensure accuracy, consistency and class labelling.

V. Data Preprocessing

The synthetically created data using CTGAN and SMOTE-based techniques with 47 features which represents actual clinical variables were preprocessed using label encoding to convert categorical features like sex, smoking status and dietary habits into numerical labels to maintain consistent representation both in training and testing stages. Label encoding is a bijective mapping $f(x) \mapsto \{0, 1, \ldots, k-1\}$ given a categorical feature $x \in \{v_1, v_2, \ldots, v_k\}$. Compatibility with PyTorch tensor operations was guaranteed and ordinal relationships were maintained where appropriate. To support supervised learning with categorical cross-entropy loss, the multi-class target variable risk category were encoded as $\text{Low} \to 0, \quad \text{Moderate} \to 1, \quad \text{High} \to 2$. To avoid any feature domination because of scale, z-score normalisation was used to standardise all encoded an numerical features. Each component of a given feature vector was transformed as $x_j' = \frac{x_j - \mu_j}{\sigma_j}$. where the mean and standard deviation of feature j across the training dataset are indicated by μ_j and σ_j . By ensuring zero-centred data with unit variance, this standardization technique enhances the stability of optimisation. SHAP-based interpretability analysis was used to select a feature subset $F' \subseteq F$ that included 14 highly informative features. Only considering most important variables, the dimensionality reduction improved generalization across variety of synthetic distributions. Python pandas and scikit-learn libraries were used to implement th preprocessing pipeline. LabelEncoder was used to encode categorical features and StandardScaler was used to standardise numerical attributes so that each feature had a zero-mean variance distribution. Before deep learning model are trained this transformation is essential to avoid scale differences skewing the learning process.

VI. Proposed Model

In the proposed model presents a domain-informed, explainable deep fusion architecture for multiclass heart disease risk prediction using synthetically generated data. By incorporating innovative methods at several stages like data generation, rule-based labelling, neural architecture design and interpretable evaluation this work deviates greatly from traditional built-in classifiers. The study represents rule-based, clinically guided annotation approach that is used on synthetic dataset with 47 features. Three distinct class of Low, high and Moderate risk that were created using feature combinations that are pertinent to the medicine makeup the target variable. For example, if a subject has severe echocardiographic abnormalities (Echo_Severity=3), impairs left ventricular function(Ejection_Fraction<40%) or a combination of comorbidities such as type 2 Diabetes(T2D) and hypertension (HTN) , they are considered as high risk. By ensuring that the label distributions match actual cardiac risk stratification paradigms, this makes the synthetic data both clinically and statistically valid.

We introduce a custom designed deep fusion model which combines Attention mechanism, Bidirectional Long Short-Term Memory (BiLSTM) and Multi-Layer Perceptron (MLP) to create a novel architecture suited for structured health data. An MLP made up of linear layers with batch normalisation and ReLU activations is used to encode the 14 input features that were chosen out of 44 feature set. Each encoded features is treated as pseudo-temporal element and fed into a BiLSTM in a sequential manner after being reshaped. The BiLSTM creates both forward and backward hidden states by capturing inter-feature dependencies. Attention weights are calculated by

$$\alpha_t = \frac{e^{W^\top h_t}}{\sum_{j=1}^{T} e^{W^\top h_j}}, \quad c = \sum_{t=1}^{T} \alpha_t h_t$$

Where c is the aggregated context vector and w is the learnable attention vector. Which allows the model to learn dynamically concentrate on feature interactions that are most important for classification. Then a SoftMax layer is applied to the context vector c to produce the final prediction, which yields for three class classification. The study uses SHAP to interpret model decisions in order to incorporate model-agnostic explainability. Each input feature's marginal contributions to the final output probability is measures by SHAP values. This provides which features influence risk predictions across various datasets. This layer of interpretability strengthens confidence and transparency in model behaviour which is important in clinical domain.

To efficiently capture both sequential patterns and hierarchical feature representation in the structured tabular dataset a deep fusion network model architecture is proposed, the pipeline of model is shown in Figure 1 .

Fourteen pre-processed and scaled features is selected as input for model based on statistical contribution and domain relevance. A Multi-Layer Perceptron (MLP) which serves as a main feature encoder where features are passed through. To improve the training stability and generalization, the MLP has a dense layer with batch normalisation after ReLU activation. This can be expressed mathematically as $h = \text{ReLU}\big(\text{BN}(W_1 x + b_1)\big)$, where x is the input feature vector, W_1 is the learnable weights and BN denotes batch normalization. After being reshaped the encoded output is given into a Bidirectional Long Short-Term Memory (BiLSTM) layer, which records contextual interactions both forward and backward to model inter-feature dependencies. After processing a sequence of encoded features, the BiLSTM layer produces two hidden representations, which are concatenated to create a single hidden state. An attention mechanism is applied over BiLSTM outputs to improve the interoperability of the model and concentrate on patterns which are clinically significant. The attention scores α_t is calculated by $\alpha_t = \frac{\exp(W \cdot h_t)}{\sum_j \exp(W \cdot h_j)}$, where W is learnable attention weighing vector. Each hidden state's contribution to the final prediction is determined by these scores. The resultant vector c is combined as weighted sum $c = \sum_t \alpha_t h_t$ and it is run through a final dense layer with SoftMax activation to predict one of three target classes Low, Moderate or High risk. The output probabilities of final layer which enabled both classification and uncertainty quantification. The Adam optimiser is used to optimise the entire architecture using a categorical cross-entropy loss function. PyTorch, which provides modular flexibility to integrate multiple learning components, was used to implement the proposed fusion model.

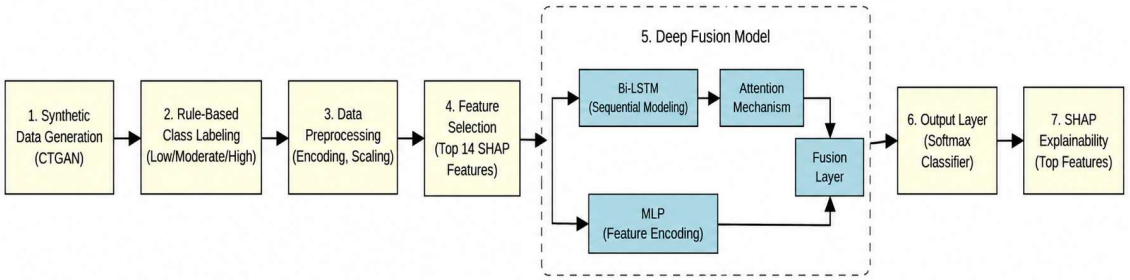

Figure 1: Fusion Model Architecture

A. Training process

Hence it is a multiclassification the categorical cross-entropy loss function is used to train the proposed fusion model. $Let y_i \in \{0,1,2\}$ represent the actual label of i[th] sample and let $\hat{y}i = \left[\hat{y}i^{(0)}, \hat{y}i^{(1)}, \hat{y}i^{(2)}\right]$ represents the predicted probability vector for each the three classes. For a single sample the cross-entropy loss is determined by:

$$\mathcal{L}(y_i, \hat{y}i) = -\sum c = 0^2 \mathbb{1}[y_i = c] \log\left(\hat{y}i^{(c)}\right)$$

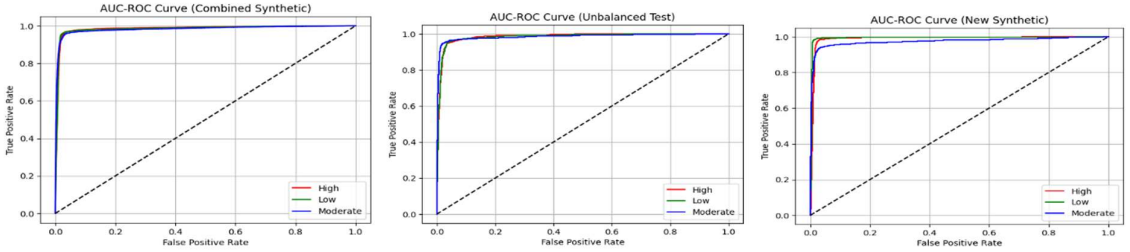

Figure 2: a) ROC-AUC Curves on Combine Synthetic data, b) Figure : ROC-AUC Curves on Unbalanced Synthetic data c) Figure : ROC-AUC Curves on New Synthetic data

To avoid local minima and speed up convergence, the model is optimized using the Adam optimizer in conjunction with learning rate scheduler. Dropout layers and early stopping based on validation loss were added to avoid overfitting. Ten thousand samples with a balanced class distribution from the Refined Synthetic Dataset are used to train the model. Statistical corelation analysis and feature relevance, guided by SHAP scores from

preliminary experiments, were used to select only 14 features out of 47 total. To stabilize gradients during optimisation, each input vector $x_i \in R^{14}$ is standardised to zero mean and unit variance.
Figure 2 a), 2 b), 2 c) shows the ROC performance for the High, Low and Moderate risk classes across three datasets. The proposed model's discriminative ability is outstanding and is demonstrated by all curves' AUC values which are close to 1.0 consistent predictive performance across dataset compositions is demonstrated by the overlap of class-wise curves.

B. SHAP explainability:

Shapely Additive exPlanations (SHAP) values are used to quantify the contribution of each input feature to the predicted outcome to enable transparent model interpretation. Based on cooperative game theory, the SHAP framework treats each feature x_j as a player that contributes to the model output $f(x)$. For feature j, the Shapely value ϕ_j is computes as:

$$\phi_j = \sum_{S \subseteq \mathcal{F} \setminus \{j\}} \frac{|S|!\,(|\mathcal{F}| - |S| - 1)!}{|\mathcal{F}|!} [f(S \cup \{j\}) - f(S)],$$

Where S is feature subset and $\mathcal{F}$ is the entire feature set. A random sample of 200 records from each test set is subjected to the permutation-based SHAP approximation due to computational limitations. Before BiLSTM processing stage, the explainability module applies SHAP to attribute importance and wraps MLP encoder inside the fusion network. For every dataset, the top ten significant features are highlighted using bar plots, SHAP summary plots, and feature importance rankings. The clinical coherence of the model and feature selection are validated by these visualizations, which offer vital clinical insights into risk drivers like Echo_Severity, Slope, Ejection_Fraction and Max_Heart_Rate.

VII. Results and Discussion

Three different datasets that is Combined Synthetic dataset, Unbalanced Dataset and New Synthetic Dataset were used to assess the proposed model. The model's hign discriminative power in multi-class classification is demonstrated by its consistent ROC-AUC scores above 0.98 with balances precision and recall across the Hogh, Moderate and Low risk classes, the model achieved an accuracy of 95.15% on the Combined Synthetic dataset. This demonstrates that the model is robust when tested on a synthetic distribution that is balanced. Even in the dace of data imbalance, performance held steady. The model demonstrated its generalizability and resilience to skewed class distributions by achieving an F1-score of 94.03% and a ROC-AUC of 0.9841 on the unbalanced dataset, which only included 500 High Risk samples. By identifying medically intuitive features like Echo_Severity, Ejection_Fraction, Slope and T2D as top contributors, the SHAP analysis further confirmed the model's clinical relevance. The model's dependability across diverse populations was further supported by the New Synthetic dataset, which produced the highest accuracy of 95.58% by simulating a novel patient distribution. By considering all things, this model is extremely accurate and clinically transparent due to the combination of domain-informed feature selection, attention architecture, and SHAP-based interpretation. These results highlight how the proposed methodology cloud support actual heart disease risk triage systems. Table summarizes the model performance using the standard metrics-Accuracy, Precision, Recall, F1 Score and ROC-AUC with all the datasets. Figure 3 a), b), c) shows the classification performance of proposed model across all three datasets which is visualized by confusion matrices. The model produced balanced predictions for every class and consistent across all datasets. In combined synthetic dataset high classification accuracy is achieved in all three classes. In imbalance dataset especially in low-risk class model performed well. With minor confusion in moderate risk class model performed well on new synthetic dataset with flawless for all three risk categories.

TABLE II: Model Performance on Different Datasets

Dataset	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Combined Synthetic Dataset	0.951	0.954	0.962	0.951	0.9839
Unbalanced Dataset	0.938	0.944	0.938	0.940	0.9841
New Synthetic dataset	0.956	0.955	0.955	0.955	0.9847

All three dataset are evaluated were subjected to SHAP based analysis to improve the interpretability of the proposed deep fusion model. In the given set of figures feature like sex, slope and echo_severity were the main contributors to the combined synthetic dataset and had highest mean SHAP value of 0.1138, echo_severity and sex and graphical location were the most contributing factors in unbalanced dataset for classification. Physiological measures such as LVEDD and unconsciousness also contributes high in imbalanced situations. In

new synthetic dataset lifestyle and behavioural factors like alcohol consumption, smoking status and pericardial diffusion demonstrated a strong influence. The SHAP interpretations shows that the model's clinical alignment and its capacity of capturing non-linear interactions which are essential for accurate heart disease risk prediction. the bar plots in Figure 4 a), b) and c) shows features ranked by mean absolute SHAP values which indicates its contribution in risk prediction.

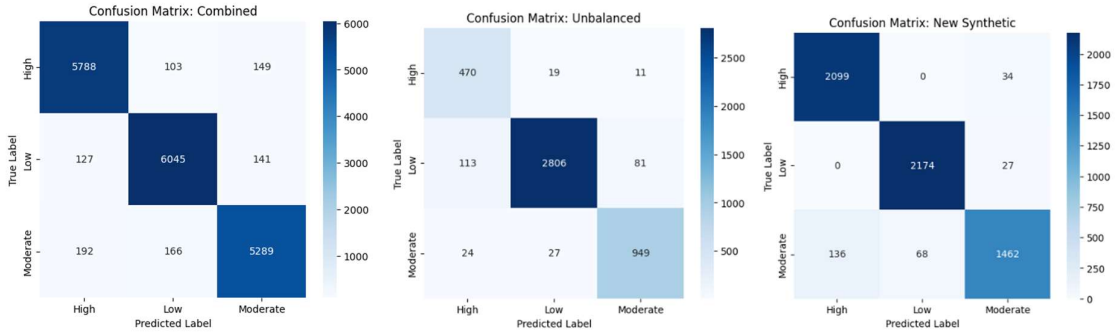

Figure 3: a) Confusion matrix of Combine Synthetic data, b) Confusion matrix of Unbalanced Synthetic data c) Confusion Matrix of New Synthetic data

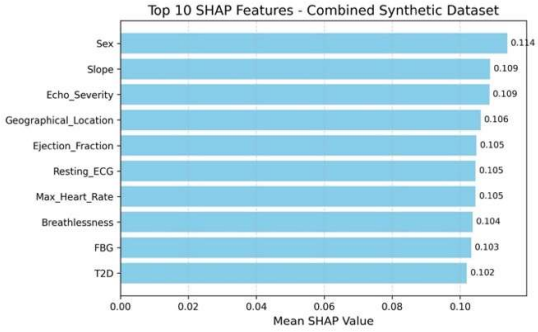

Figure 4 a) SHAP features of Combined Synthetic dataset

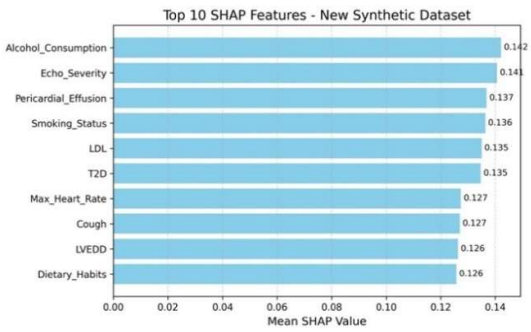

Figure 4 b) Top 10 SHAP features of New Synthetic dataset

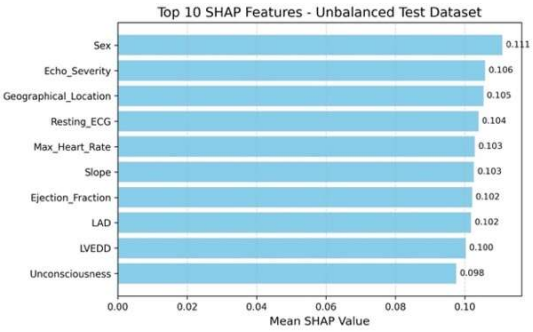

Figure 4 c) Top 10 SHAP features of Unbalanced dataset

VIII. CONCLUSION

This study demonstrated an interpretable deep fusion architecture that combines MLP, BiLSTM and an attention mechanism for the multi-class heart disease risk prediction task using synthetically generated data. Fourteen clinically and statistically significant features were chosen from a 47-feature dataset to train model. High generalization capability was shown by rigorous evaluation across three test datasets that is Combined Synthetic dataset, Unbalanced dataset and New synthetic dataset. Accuracy scores were consistently above 93% and ROC-AUC scores were consistently above 0.98. predictions became more clinically reliable when a SHAP-based

interpretability framework was incorporated, allowing transparent feature attribution and the identification of important contributing variables like Echo_Severity, Ejection_Fraction, Slope and Sex. The pipeline of the model which includes probabilistic inference, standardization and label encoding creates a repeatable and effective method appropriate for cardiovascular risk decision support system. The model can be further improving the robustness and domain adaption in future studies. The study may further investigate integrating temporal patient data, real-world EMR-based features, and hybrid generative augmentation techniques.

REFERENCES

[1] Mahmoud Ibrahim, Yasmina Al Khalil, Sina Amirrajab, Chang Sun, Marcel Breeuwer, Josien Pluim, Bart Elen, Gökhan Ertaylan, Michel Dumontier,Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges, Computers in Biology and Medicine, Volume 189, 2025, 109834, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2025.109834.A.

[2] Yuan Zhong, Xiaochen Wang, Jiaqi Wang, Xiaokun Zhang, Yaqing Wang, Mengdi Huai, Cao Xiao, and Fenglong Ma. 2024. Synthesizing Multimodal Electronic Health Records via Predictive Diffusion Models. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24). Association for Computing Machinery, New York, NY, USA, 4607–4618. https://doi.org/10.1145/3637528.3671836

[3] Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," in IEEE Access, vol. 7, pp. 180235-180243, 2019.

[4] Acharya, U & Fujita, Hamido & Oh, Shu Lih & Hagiwara, Yuki & Tan, Jen Hong & Abdul Rahim, Muhammad Adam & Tan, Ru San. (2019). Deep Convolutional Neural Network for the Automated Diagnosis of Congestive Heart Failure Using ECG Signals. Applied Intelligence. 49. 10.1007/s10489-018-1179-1.

[5] Rajkomar, A., Oren, E., Chen, K. et al. Scalable and accurate deep learning with electronic health records. npj Digital Med 1, 18 (2018). https://doi.org/10.1038/s41746-018-0029-1

[6] Alizadehsani R, Abdar M, Roshanzamir M, Khosravi A, Kebria PM, Khozeimeh F, Nahavandi S, Sarrafzadegan N, Acharya UR. Machine learning-based coronary artery disease diagnosis: A comprehensive review. Comput Biol Med. 2019 Aug;111:103346. doi: 10.1016/j.compbiomed.2019.103346. Epub 2019 Jul 4. PMID: 31288140.

[7] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform. 2018 Sep;22(5):1589-1604. doi: 10.1109/JBHI.2017.2767063. Epub 2017 Oct 27. PMID: 29989977; PMCID: PMC6043423.

[8] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[9] Holzinger, Andreas & Biemann, Chris & Pattichis, C. & Kell, Douglas. (2017). What do we need to build explainable AI systems for the medical domain?. 10.48550/arXiv.1712.09923.

[10] Ghassemi, Marzyeh & Naumann, Tristan & Schulam, Peter & Beam, Andrew & Ranganath, Rajesh. (2018). Opportunities in Machine Learning for Healthcare. 10.48550/arXiv.1806.00388.

[11] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, Joel T Dudley, Deep learning for healthcare: review, opportunities and challenges, Briefings in Bioinformatics, Volume 19, Issue 6, November 2018, Pages 1236–1246, https://doi.org/10.1093/bib/bbx044

[12] Ahmad, Waqar & Ramzan, Muhammad & Farooq, Aamir & Syed, Shahab & Satti, Zahra & Ud, Shahab & Syed, Din. (2025). Explainable AI-Enhanced Machine Learning Models for Early Detection of Cardiovascular Disease: Improving Predictive Performance and Clinical Transparency through Optimization. 10.63075/t83b6242.

[13] Rahmathullah, R., Bhavanam, S.N. and Midasala, V. 2025. Deep Learning-Powered Cardiovascular Disease Prediction: A Novel Approach to Early Diagnosis and Risk Assessment. Journal of Neonatal Surgery. 14, 4 (Mar. 2025), 21–31. DOI:https://doi.org/10.52783/jns.v14.2421.

[14] V K, Sudha & Kumar, D.. (2023). Hybrid CNN and LSTM Network For Heart Disease Prediction. SN Computer Science. 4. 10.1007/s42979-022-01598-9.

[15] Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. Sensors (Basel, Switzerland), 23(2), 634. https://doi.org/10.3390/s23020634