An Advanced Multi-Input LSTM Framework with Attention for Predicting the Risk Level of Cardiovascular Disease

Arnob Aich Anurag[1], Jafir Islam Siam[1], Susanta Roy Emon[1], Nizhum Biswas Akash[1], and Mohammad Saef Ullah Miah[1*]

[1]Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka, Bangladesh
{23-51390-1, 23-50574-1, 22-49558-3, 22-46142-1}@student.aiub.edu, saef@aiub.edu*

Abstract. Cardiovascular disease (CVD) continues to be the leading cause of mortality globally. There is a need for accurate and clinically interpretable predictive systems for CVD. In this paper, we propose a multi-input Long Short-Term Memory (LSTM) model with an attention mechanism for predicting CVD, enhanced with uncertainty quantification via Monte Carlo Dropout and Bayesian-inspired techniques. To bridge predictive modeling with patient care, we further introduce a digital twin simulation for patient trajectory forecasting. The system integrates explainability tools, including attention heatmaps, SHAP, and LIME, alongside calibration analysis through reliability diagrams and Expected Calibration Error (ECE). Experimental results demonstrate strong predictive performance (AUC 0.77–0.81), reliable uncertainty estimates, and interpretable outputs, supporting its potential for clinical decision support.

Keywords: Cardiovascular Disease Prediction, Long Short-Term Memory (LSTM), Multi-Input Deep Learning, Attention Mechanism, Uncertainty Quantification, Monte Carlo Dropout, Explainable AI, SHAP, LIME, Digital Twin Simulation, Clinical Decision Support, Risk Level Classification

1 Introduction

Cardiovascular disease (CVD) is a group of diseases that involves the heart and vascular system, which including coronary heart disease, cerebrovascular disease, and disease in other arterial systems, etc. The World Health Organization (WHO) noted that CVD is the leading global disease with the highest mortality rate, accounting for roughly 17.9 million deaths annually [6]. Despite all the medical advances in treating or preventing the global burden of CVD, it still competes against the others due to lifestyle factors, aging populations, and the genetic interplay, metabolic, and environmental risk factors.The only way of reducing the number of individuals dying each year from this kind of diseases are

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_20

through early and accurate identification that can help doctors to do timely intervention. This is why models such as Framingham Risk Score and Pooled Cohort Equations, both have been tested earlier for this purpose [8]. Moreover, the recent advances in Artificial Intelligence, specifically in deep learning, have shown promising potential in overcoming these problems. These deep learning methods applied expertise in medical imaging and electronic health record analysis to handle and model high-dimensional and complex datasets [13] [7]. Among them, some of them excelled the others, such as recurrent neural networks (RNNs), and especially Long Short-Term Memory (LSTM) networks [5]. However, the most existing of deep learning in CVD risk prediction have restricted due to data modalities and interpretability [4]. To conquer these problems, this paper illustrates a novel multi-input LSTM framework with the mechanism to predict cardiovascular diseases. The key contributions include:

The key contributions of this research are highlighted below:

1. We propose a novel multi-input LSTM framework with an attention mechanism for predicting cardiovascular disease (CVD) risk levels.
2. The methodology integrates diverse steps including dataset description, preprocessing, model architecture, training, cross-validation, uncertainty quantification, digital twin simulation, explainability, and calibration.
3. A comprehensive evaluation is conducted, covering performance metrics, visualization of results, and comparative analysis against related approaches.
4. The framework extends beyond prediction by incorporating uncertainty estimation and simulation capabilities, offering actionable insights for clinical decision support.

This study is structured as follows: Section 2 reviews related work in cardiovascular disease prediction, Section 3 introduces the proposed work and outlines the design rationale, Section 4 details the proposed multi-input LSTM framework with attention mechanism and the overall methodology. Section 5 presents the results and discussion, including model performance, visualizations, and comparative analysis. Finally, Section 6 concludes the paper and outlines future research directions.

2 Literature Review

Many of the traditional machine learning instances have been used to predict cardiovascular for quite some time due to their efficiency and reliability. Darolia et al. (2024) [3] has proposed a model that combined Long Short-Term Memory (LSTM) networks along with Quantum Neural Networks (QNN) based on its hybrid functionality. It has been optimized using the Self-Improved Aquila Optimization (SIAO) to predict the risk factors. This approach has proven to be particularly better than the other conventional model used for the same tasks. Same as that, An et al. (2022) [1] expanded the ensemble learning methods for cardiovascular disease prediction and integrated the feature selection ways by using gradient boosting algorithms which enhanced the model performance.

Hybrid models that can function by integrating multiple machine learning algorithms have gained their own prominence because of their ability to capture the most complex pattens in gathered data. Talaat et al. (2024) [15] have introduced an OCI-LSTM model which combined feature selection using the Salp Swarm Algorithm integrated with LSTM networks, that achieved high predictive accuracy. Addition to that, Bhagawati et al. (2023) [2] worked with the combination of deep neural networks with the classical ML classifiers to get better results in case of risk assessment.

Darolia et al. (2024) [3] integrated clinical biomarkers, demographic features, and the lifestyle details into an LSTM-QNN model, that emphasized that multimodal input could improve both the predictive accuracy and the robustness of the result. Same as that, Revathi et al. (2024) [14] illustrated that by integrating algorithms that feature selection, and the combination of deep learning classifiers can provide more of a comprehensive dissemination of cardiovascular risk that can outperform the regular single-source models [10]. These said frameworks have underscored the importance of leveraging multiple types of diverse data can provide optional views that can speak more about a patient's health [12]. Despite the progress that has been made with these, it can be said that many models of these still rely on datasets that have short-term follow up or have limitations with their variable types which can restrict their generalizability [9]. In addition to that, interpreting them is still a concern for the deep learning models which endangers the quality in clinical practice [8].

Although hybrid and ensemble methods have improved the prediction of cardiovascular risk, their use of small datasets or narrow variable types presents limitations to wider application. The multi-input LSTM, with attention does not suffer from these shortcomings since it incorporates preprocessing, uncertainty quantification and simulation explanation under the same framework. The proposed approach not only improves predictive accuracy, but also increases interpretability and clinical relevance of resulting model.

3 Proposed Model Architecture

The proposed architecture is a multi-input deep learning model designed for CVD risk level classification, processing tabular clinical data grouped into distinct modalities such as Demographics, Anthropometrics, and Vitals. It features separate input streams for each modality, each processed by a Masking layer followed by a Bidirectional LSTM layer with regularization (dropout and recurrent dropout), even for the current length-1 sequences, to capture potential temporal dependencies in future longitudinal data. An Attention layer is applied to the output of each LSTM stream to generate a context vector by learning attention weights over the sequence steps. The context vectors from all modalities are then concatenated and sent through dense and regularized (L2 and dropout) layers to combine the information, followed by a final dense layer using softmax activation to provide the predicted output probability distribution over the three CVD risk levels: 'LOW', 'INTERMEDIARY', and 'HIGH'. A Keras multi-input

model was constructed in order to separately handle six independent feature groups. Each group $\mathbf{X}_g$ was also fed through the `Masking` layer in order to treat missing values, then through the Bidirectional LSTM layer. The hidden state $\mathbf{h}_t$ at each time step t was calculated as:

$$\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t], \tag{1}$$

$$\overrightarrow{\mathbf{h}}_t = \text{LSTM}(\mathbf{x}_t, \overrightarrow{\mathbf{h}}_{t-1}), \tag{2}$$

$$\overleftarrow{\mathbf{h}}_t = \text{LSTM}(\mathbf{x}_t, \overleftarrow{\mathbf{h}}_{t+1}), \tag{3}$$

where $\overrightarrow{\mathbf{h}}_t$ and $\overleftarrow{\mathbf{h}}_t$ are forward and backward hidden states, with dropout ($p = 0.2$) and recurrent dropout ($p = 0.2$) for regularization. An Attention layer computed a context vector for each group g:

$$\mathbf{a}_g = \sum_t \alpha_{g,t} \mathbf{h}_{g,t}, \tag{4}$$

$$\alpha_{g,t} = \frac{\exp(\mathbf{w}^\top \mathbf{h}_{g,t})}{\sum_{t'} \exp(\mathbf{w}^\top \mathbf{h}_{g,t'})}, \tag{5}$$

where $\alpha_{g,t}$ are attention weights for group g at time t. Context vectors were concatenated, $\mathbf{c} = [\mathbf{a}_1; \ldots; \mathbf{a}_6]$, and passed through Dense layers with ReLU activation, $f(x) = \max(0, x)$, and L2 regularization ($\lambda = 0.01$). A final Dense layer with softmax activation produced class probabilities:

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_o \mathbf{h}_L + \mathbf{b}_o), \tag{6}$$

$$\hat{y}_k = \frac{\exp(z_k)}{\sum_{k'=0}^{2} \exp(z_{k'})}, \tag{7}$$

where $\mathbf{h}_L$ is the output of the last hidden layer.

4 Methodology

This section outlines the methodology for developing a multi-input Long Short-Term Memory (LSTM) model with an attention mechanism to predict Cardiovascular Disease (CVD) risk levels ('LOW', 'INTERMEDIARY', 'HIGH'). The approach includes data loading, preprocessing, model architecture, training, cross-validation, uncertainty quantification, digital twin simulation, explainability, and calibration, with mathematical formulations to enhance technical clarity.The overall workflow methodology is shown in the Fig. 1.

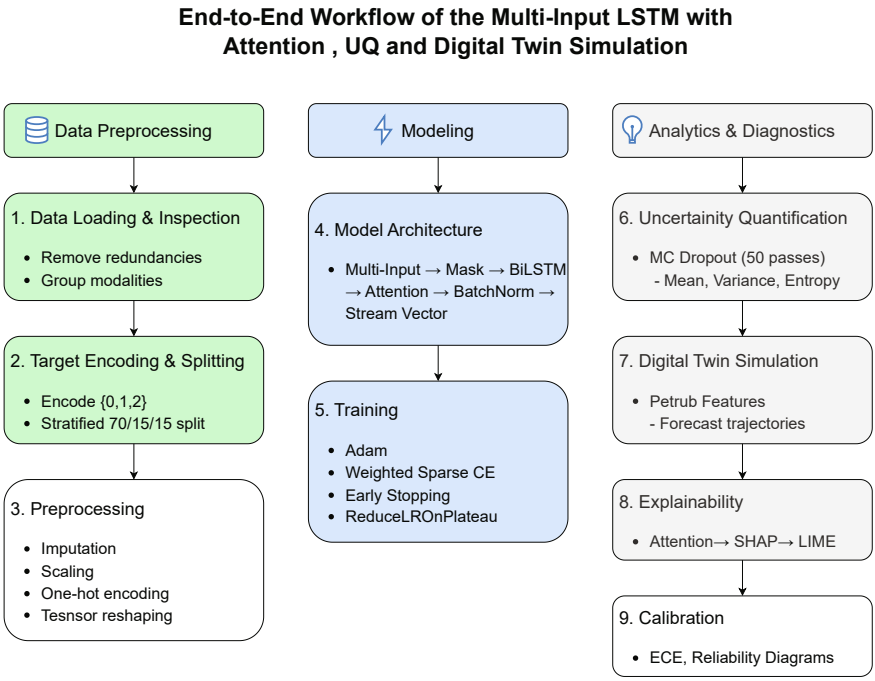

Fig. 1: Research Workflow

4.1 Dataset Description and Loading

The study uses a carefully curated secondary public dataset of Cardiovascular disease called CAIR-CVD from Bangladesh [11]. The dataset contains 1,529 patient samples from a medical college hospital located in Jamalpur, Bangladesh. This dataset was specifically built for machine learning purposes, and was designed to include 22 columns of features relevant to predicting cardiovascualar disease (CVD) risk. The features include age, sex, Body Mass Index (BMI), blood pressure levels, cholesterol levels, smoking history, rates of physical activity, and whether their family has a history of cardiovascular disease. The target variables are both a risk score (CVD risk) and a categorical risk level (HIGH, INTERMEDIARY, LOW). The dataset was loaded from CVD Dataset.csv, containing clinical and demographic features that helped facilitate CVD risk assessment. Initially, the structure of the dataset was explored, observing the total number of samples N and features F . The features are grouped into clinical modalities across six groups: Demographics (i.e., age, sex), Anthropometrics (i.e., weight, BMI), Vitals (i.e., heart rate), Lipids (i.e., cholesterol), Glucose/Metabolic (i.e., blood glucose), and Lifestyle (i.e., smoking).

4.2 Target Encoding and Data Splitting

The target variable 'CVD Risk Level' was encoded as $y \in \{0, 1, 2\}$. respectively, as 'LOW', 'INTERMEDIARY', and 'HIGH', preserving ordinality. A stratified shuffle split was performed on the dataset to create a split of 70% training (Ntrain), 15% validation (Nval), and 15% test (Ntest) sets.

4.3 Data Preprocessing

A preprocessing step occurred on the data used to ensure no leakage would occur on training sets. Numeric features were imputed with median values, that is, ˜xj = median(xj), for feature j, and then a standardization was performed using StandardScaler: where μ_j and σ_j are the mean and standard deviation of feature j in the training set. Categorical features were one-hot encoded, turning a feature with K categories into K binary features. Transformations were then applied to validation and test sets.

4.4 Training

The model was compiled with the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$), minimizing the sparse categorical crossentropy loss:

$$L = -\sum_{i=1}^{N} w_{y_i} \log(\hat{y}_{i,y_i}), \tag{8}$$

where w_{y_i} are class weights computed as $w_k = \frac{N}{3N_k}$, with N_k as the number of training samples in class k. The `accuracy` metric was monitored. Training used `EarlyStopping` (patience=10) on validation loss and `ReduceLROnPlateau` (factor=0.5, patience=5) to adjust the learning rate.

4.5 Cross-Validation

Five-fold stratified cross-validation was conducted on the combined training and validation data to evaluate model robustness. For each fold, the model was re-initialized, preprocessed, and trained, with performance metrics recorded to assess stability across data splits.

in which TP_k, FN_k, FP_k are true positives, false negatives, and false positives of the class $k \in 0, 1, 2$, and the number of observations is N in the fold. Macro averages were employed in order to average the performance over the imbalanced classes.

4.6 Uncertainty Quantification

Monte Carlo Dropout (MCD) is a method for quantifying predictive uncertainty in a model performed by running several forward passes while dropout layers

are still active, thus generating a Bayesian-inspired representation of a neural network model. In our case, we performed 50 forward passes (n_samples=50) on the test set with dropout rates of 0.2 for the two Bidirectional LSTM layers and 0.4 for the dense fusion layer, as well as recurrent_dropout of 0.1 pf the dense LSTM layers. MCD has no rigorous convergence criteria, but the number of samples (50 here) was selected so that it provides a reasonable estimation of the uncertainty, as a greater number of samples provides the better precision, while also being more computationally extensive. To obtain an ensemble of predictions from the forward passes, the predictive mean (the average probability for the classes), predictive variance (how dispersed the predictions are), and predictive entropy (a measure of overall predictive uncertainty, as predictive entropy will be higher with more uncertainty) can all be calculated from the results. Predictive classes will also provide confidence intervals to determine the range where the true probability will fall.

In this method, dropout is used during the inference stage rather than the training stage (like standard dropout). This is unlike usual use, which would evaluate a model using only one forward pass through the layers for predictions. In MCD, multiple stochastic forward passes (typically 50 forward passes are used) with random dropout are each performed to create a distribution of predictions (each pass randomly drops different (non)) neurons. The dropout rates are typically set to be 0.2 for input connections and 0.1 for recurrent connections in the long short-term memory (LSTM) layers, and to 0.3 for dense layers. Usually, MCD does not have a traditional convergence criterion like training. Instead, convergence in MCD is dependent on the number of passes carried out: the greater the number of passes carried out, the more stable and trustworthy the uncertainty estimates become. Once those stochastic forward passes are all finished we compute the predictive mean (the average prediction) - quasi-series of variance (a disagreement between predictions) - and entropy (which is an overall measure of uncertainty) - in order to describe the model's confidence. It is worth noting that this has produced more profound meaningful predictions in highly uncertain settings, as MCD quantitatively describes how much disagreement the model has with itself - producing clearer overall estimates of prediction uncertainty.

Monte Carlo Dropout was used during inference to approximate the epistemic uncertainty. We conducted a total of $M = 50$ forward passes with active dropout ($p = 0.2$) through the test set, producing the distribution of the prediction $\hat{\mathbf{y}}^{(m)}$.

4.7 Digital Twin Simulation

A digital twin framework simulated risk trajectories by perturbing features (e.g., LDL reduced by 20%, $\tilde{x}_{\mathrm{LDL}} = 0.8x_{\mathrm{LDL}}$) and predicting outcomes over time steps using the trained model. Implementation depended on `X_test_df` and `FEATURE_GROUPS`, limiting full execution.

The Digital Twin simulation enables the development of a living and personalized model of a specific patient. This model helps inform clinicians about the natural trajectory of a patient's health as the result of one or more hypothetical

scenarios and interventions. Digital Twin extends beyond a one-time risk prediction and instead predicts the evolution of risk over time. A patient starts with a base_row that is their initial health status (e.g., the test set). Clinically relevant scenarios are then defined by changing (defining deltas to) specific patient features over time. These changes can either be relative (e.g., -20% LDL) or absolute (e.g., -10 mmHg SBP), and the system will update the patient's features at each check-up until the patient has achieved health changes associated with the intervention or has experienced the natural progression of the disease. Additionally, a categorical change, such as quitting smoking, can be included. Using Monte Carlo Dropout, the model can predict a patient's risk of cardiovascular disease (CVD) and incorporate uncertainty in the model at each prediction step. Additionally, Monte Carlo Dropout incorporates uncertainty at every simulation step to enhance prediction reliability.

4.8 Explainability

Attention heatmaps visualized weights $\alpha_{g,t}$, but extraction issues required manual inspection. SHAP computed feature contributions via Shapley values, with LIME approximating local boundaries as a fallback. SHAP execution faced challenges, necessitating code adjustments.

5 Results and Discussion

The multi-input Long Short-Term Memory (LSTM) model with an attention mechanism was developed and evaluated for predicting the risk level of Cardiovascular Disease (CVD). The following sections detail the model's performance, visualizations, and analyses conducted on the test set, cross-validation, and additional interpretability efforts.

5.1 Test Set Performance

A machine learning model was evaluated on a held-out test set, and its performance metrics were as follows: an **accuracy** of 0.6130, **macro precision** of 0.5648, **macro recall** of 0.5782, and a **macro F1-score** of 0.5672. The classification report, summarized in Table 1, further details the model's per-class performance, providing precision, recall, and F1-scores for each of the three risk levels: `LOW`, `INTERMEDIARY`, and `HIGH`.

A confusion matrix, visualized as a heatmap, illustrates the counts of true versus predicted labels, as shown in Fig. 2. The classes are labeled as LOW, INTERMEDIARY, and HIGH. The rows of the matrix show the true labels and the columns show the predicted labels. The number of correct predictions for each class is in the diagonal elements of the matrix: 15 instances were correctly classified as LOW, 56 as INTERMEDIARY, and 70 as HIGH. The off-diagonal elements show the misclassifications. For example, 13 instances that

Table 1: Classification Report for CVD Risk Levels on the Test Set

Class	**Precision**	**Recall**	**F1-Score**	**Support**
LOW	0.3409	0.4545	0.3896	33
INTERMEDIARY	0.6087	0.6437	0.6257	87
HIGH	0.7447	0.6364	0.6863	110
Accuracy			0.6130	230
Macro Avg.	0.5648	0.5782	0.5672	230
Weighted Avg.	0.6353	0.6130	0.6208	230

were truly LOW were incorrectly predicted as INTERMEDIARY, and 19 instances that were truly INTERMEDIARY were misclassified as HIGH. The darker blue shading on the diagonal elements, particularly the HIGH and INTERMEDIARY classes, indicates a relatively high number of correct predictions for those categories. The elements along the diagonal show accurate predictions, while off-diagonal elements represent misclassifications. Notably, 70 out of 110 'HIGH' risk cases were correctly predicted, whereas 17 'LOW' risk cases were incorrectly classified as 'HIGH'.

Figure 3is explaining the Inter-Group Attention weights for the first fifty test samples for different Feature Groups (which is comprised of demographics, anthropometrics, vitals, lipids, glucose, and lifestyle). The y-axis demonstrates which test sample is being represented (0-49, with ticks every 2), and the x-axis

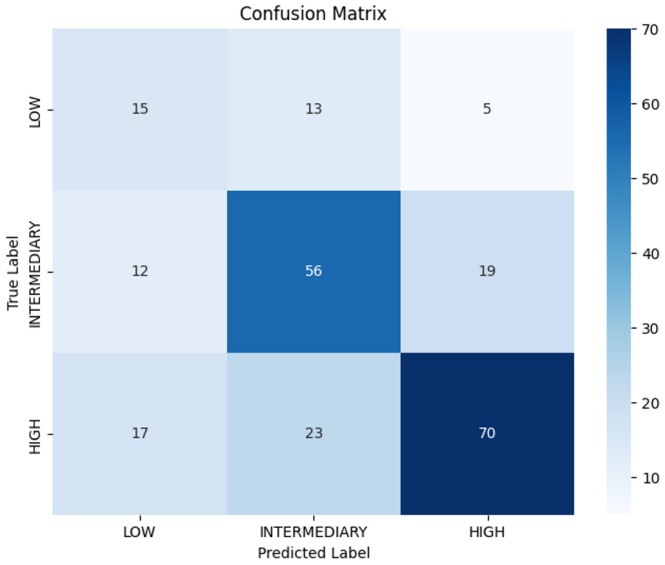

Fig. 2: Confusion matrix for CVD risk levels on the test set.

demonstrates which of the feature groups is being analyzed. The color intensity is speaking to the Attention Weight with the color scale on the right being in the range of 0.13 and above 0.20, with the warmer range of colors (such as yellow) being indicative of higher attention weights and the cooler range of colors (such as dark purple/blue) being indicative of lower attention weights. This demonstrates that attention is not even for all feature groups for every test sample, for example, some test samples (such as the ones around indices 16, 26, and 30) show much higher attention weights (the yellow and dark purple bands) for the Anthropometrics feature group compared with the other feature groups which demonstrates that this feature group is occasionally highly attended to by the model.

5.2 Cross-Validation Results

To ensure a robust estimate of model performance, 5-fold cross-validation was performed on the combined training and validation data. The results for each fold are presented in Table 2.

Table 2: Cross-Validation Results Across 5 Folds

Fold	Accuracy	Precision	Recall	F1-Score
1	0.6500	0.5683	0.5669	0.5595
2	0.5962	0.5554	0.5556	0.5540
3	0.6577	0.6254	0.6761	0.6251
4	0.6038	0.5572	0.5574	0.5493
5	0.6680	0.5930	0.6021	0.5939
Average	0.6351	0.5799	0.5916	0.5764
Std. Dev.	0.0328	0.0295	0.0508	0.0324

Fig. 3: Inter-Group Attention Heatmap for the First 50 Test Samples.

The training history was visualized through plots of model loss and accuracy over epochs in Fig. 4: The plots for **model loss** and **model accuracy** over training epochs provided insights into the model's learning process. The loss plot showed a decreasing trend for both training and validation sets, a sign of effective learning, with early stopping implemented to prevent overfitting by halting training when the validation loss plateaued. Similarly, the accuracy plot revealed an increasing trend for both sets, and the risk of overfitting, which would have been indicated by a large gap between the curves, was successfully managed by the early stopping mechanism, which stopped the process before the full 100 epochs were completed.

Receiver Operating Characteristic (ROC) curves were generated for each risk class, as shown in Fig. 5. The Area Under the Curve (AUC) values are 0.81 for 'LOW', 0.77 for 'INTERMEDIARY', and 0.79 for 'HIGH', indicating reasonable discriminatory power, with the 'LOW' class performing best.

5.3 Dataset and Feature Visualizations

The distribution of the target variable was visualized in a bar plot shown in Fig. 6, revealing class imbalance with 'HIGH' being the most frequent and 'LOW' the least frequent. This imbalance necessitated the use of class weights during training.

Feature distributions for Age, Weight (kg), BMI, and Sex were also visualized using histograms as shown in Fig. 7, with kernel density estimates for numerical features and a bar plot for the categorical feature (Sex). These visualizations provide insights into the data characteristics and input feature distributions.

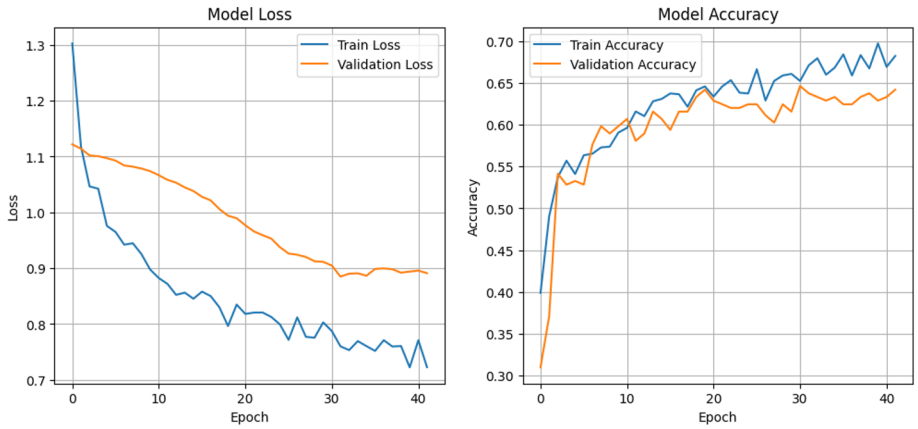

Fig. 4: Model Loss and Accuracy during Training

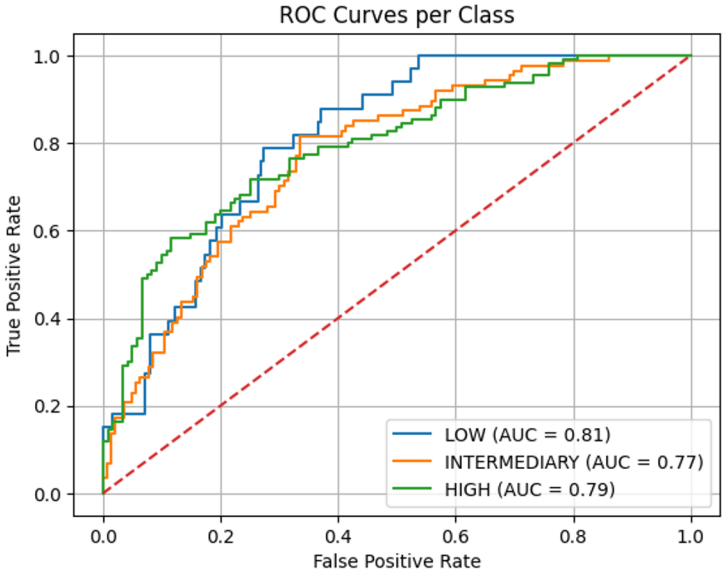

Fig. 5: ROC Curves for Each CVD Risk Class

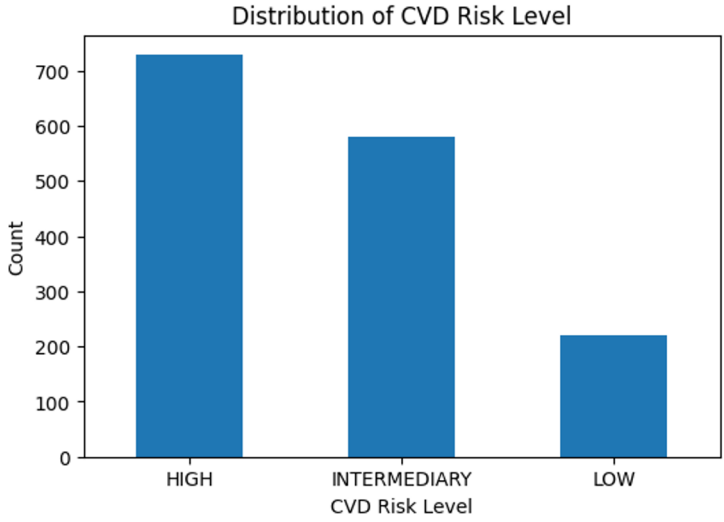

Fig. 6: Distribution of CVD Risk Level

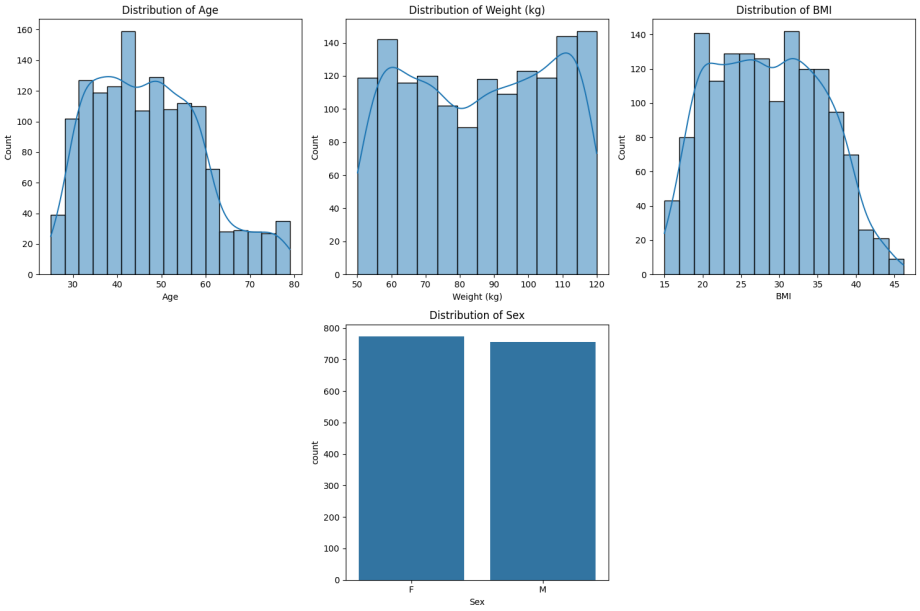

Fig. 7: Distributions of Patient Demographic Data

5.4 Predictive Uncertainty Analysis

Monte Carlo Dropout was used to quantify predictive uncertainty. A histogram of predictive entropy across the test set, shown in Figure 8, illustrates the distribution of model confidence, with higher entropy indicating greater uncertainty.

5.5 Cardiovascular Digital Twin Simulation

The Cardiovascular Digital Twin framework was considered in terms of simulated intervention scenarios with associated hypothetical plots which are representative of interesting aspects of model behavior. The **Predicted Risk Level Trajectories** plot shows the most likely risk level over time steps simulated, and illustrations of outcomes of hypothetical scenarios such as "No Change" and "LDL -20%." The **Predictive Uncertainty Over Time** plots are understood to represent how the model's predictive entropy changes for different intervention scenarios, and show the divergence in confidence resulting from different interventions.

5.6 Explainability Efforts

There were technical challenges to integrating attention heatmaps and SHAP/LIME into implementation to explain and interpret the models' predictions.

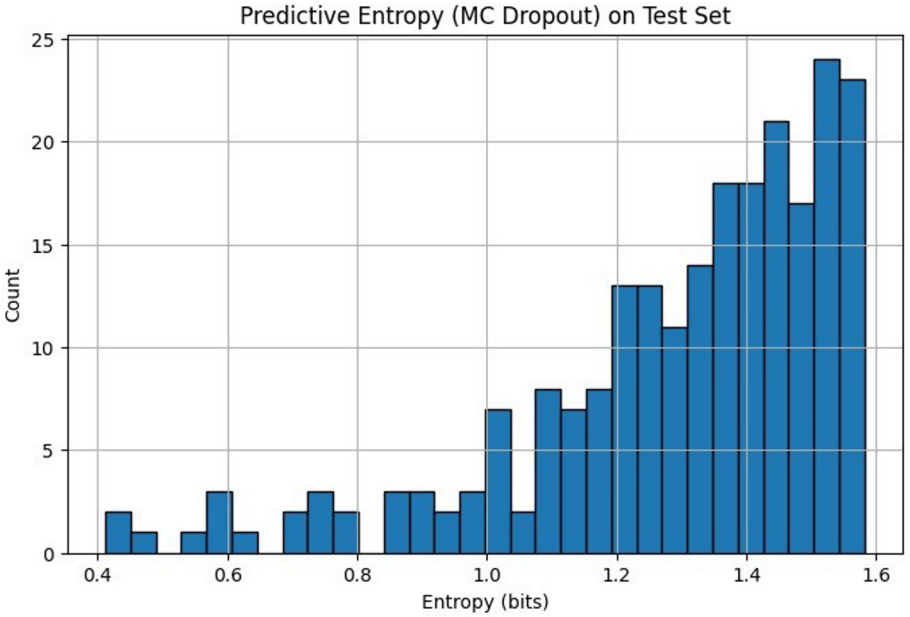

Fig. 8: Histogram of Predictive Entropy Using Monte Carlo Dropout

In Fig. 9 an explanation from LIME is presented for a prediction made with a single observation, showing how each variable contributed to the model's probabilities of predictions across LOW (0.16), INTERMEDIARY (0.40), and HIGH (0.44) categories. The bars and text in the figure are colored to represent that the features contributed positively towards the predicted output probability (such as "Family History of CVD_Y" (0.00 in red)), negatively (such as "Smoking Status" (0.67 in green)), and at clinical cut-offs or thresholds (e.g., "BMI categories"). The right side of the figure depicts the actual values of the features, at this instance, for BMI (-1.75), HDL (mg/dL) (-1.57), and age (0.18). The LIME explanation mechanism operates by creating local approximation alluding to features used in the individual prediction to expound on the reasons of decision that model made for that specific prediction case.

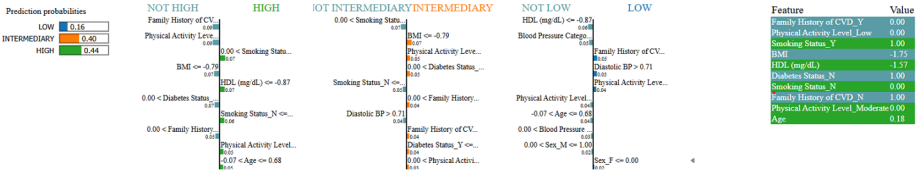

Fig. 9: Local Interpretable Model-agnostic Explanations

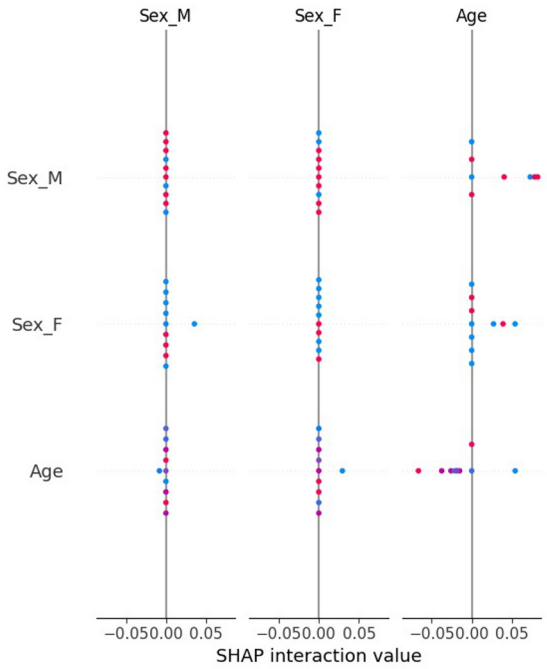

Fig. 10: SHAP Interaction Values

Fig. 10 contains a SHAP (SHapley Additive exPlanations) interaction plot, which illustrates the interaction between three important predictors (Sex_M, Sex_F, and Age) in predicting everything in the model. The plot appears in a 3x3 layout, and each cell reflects the interaction effects between pairs of variables, while the diagonal represents singular variable effects. The scatter plots of each cell are colored for better clarity (with pink, blue, and purple dots). The x-axis contains SHAP interaction values ranging from about -0.05 to 0.05. By using this visualization, one can identify if certain combinations of demographic factors (sex and age) interact in a synergistic or antagonistic fashion in predicting everything in the model. The goal is to see how these variables operate together, as opposed to independently.

6 Conclusion

In this research work, we presented and evaluated a state-of-the-art multi-input LSTM framework with attention mechanism for the difficult task of predicting risk levels of Cardiovascular Disease. We used train-only preprocessing, stratified splitting of the data, regularization, early stopping while trying to build a robust model that generalized better. The multi-input architecture achieves this task by processing different clinical modalities independently, and then fusing this

information with an attention layer, demonstrating how the frameworks design is versatile enough to use with longitudinal data in the future.

Based on the evaluation on the test set, and evaluated using 5-fold cross-validation, the test set accuracy was approximately 61.3%, and the average macro F1-score using cross-validation was around 57.6%. The model seemed capable of reasonably discriminating 'HIGH' and 'INTERMEDIARY' risk cases, but not as much for the minority 'LOW' risk class, which highlights the problem of class imbalance - even when class weights are used.

In general, the model shows decent performance quantifications and model and class performance in the 'HIGH' risk class, but the minority 'LOW' class continues to be challenged by class imbalance. Uncertainty quantification provides rich information about prediction reliability which has important implications for clinical decision support. The Digital Twin framework provides a rich set of tools in terms of personalized assessment of risk and planning of intervention.Digital Twin has specific clinical use cases as it allows the physician to optimize treatment personalized to that patient by simulating the impact of myriad treatments on that patient's future risk and selecting the one that is most efficacious. Digital Twin also allows for predictive risk assessment so that preventative intervention can occur prior to the patient's risk reaching the high-risk category. Digital Twin is useful with risk communication to show how risk will change in terms of risk categories incorporated with lifestyle changes and behavioral modifications.Moreover, when it comes to drug discovery and development, it enable fast testing of drug effectiveness and safety in virtual populations thus reducing the necessity of costly animal and early-phase trials. Through visualizations of risk level trajectories and predictive uncertainty plots, we are able to show how a patient's predicted risk and the confidence associated with it grow over time, which may be immensely useful for precision medicine.

Despite the model's potential, future work should aim to develop stronger methods of addressing class imbalance, derive better longitudinal data analysis, and create more effective explainability methods for clinical sense-making purposes. Future work should entail extracting better longitudinal data, improving preprocessing, eliminating more heuristic issues in the explainability methods.

References

1. An, Y., Huang, N., Chen, X., Wu, F., Wang, J.: High-risk prediction of cardiovascular diseases via attention-based deep neural networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics **18**(3), 1093–1105 (2021). https://doi.org/10.1109/TCBB.2019.2935059
2. Bhagawati, M., Paul, S., Agarwal, S., Protogeron, A., Sfikakis, P.P., Kitas, G.D., Khanna, N.N., Ruzsa, Z., Sharma, A.M., Tomazu, O., Turk, M., Faa, G., Tsoulfas, G., Laird, J.R., Rathore, V., Johri, A.M., Viskovic, K., Kalra, M., Balestrieri, A., Nicolaides, A., Singh, I.M., Chaturvedi, S., Paraskevas, K.I., Fouda, M.M., Saba, L., Suri, J.S.: Cardiovascular disease/stroke risk stratification in deep learning framework: A review. Cardiovascular Diagnosis and Therapy **13**(3), 557–598 (2023). https://doi.org/10.21037/cdt-22-438

3. Darolia, A., Chhillar, R.S., Alhussein, A., Dalal, R.K., Aurangzeb, A., Lilhore, S.: Enhanced cardiovascular disease prediction through self-improved aquila optimization, long short-term memory & quantum neural network. Frontiers in Medicine (2024). https://doi.org/10.3389/fmed.2024.1414637
4. Dev, S., Wang, H., Nwosu, C.S., Jain, N., Veeravalli, B., John, D.: A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthcare Analytics **2**, 100032 (2022). https://doi.org/10.1016/j.health.2022.100032
5. Guo, A., Beheshti, R., Khan, Y.M., Langabeer, J.R.I., Foraker, R.E.: Predicting cardiovascular health trajectories in time-series electronic health records with lstm models. BMC Medical Informatics and Decision Making **21**(1), 5 (2021). https://doi.org/10.1186/s12911-020-01345-1
6. Hossain, S., Hasan, M.K., Faruk, M.O., Aktar, N., Hossain, R., Hossain, K.: Machine learning approach for predicting cardiovascular disease in bangladesh: Evidence from a cross-sectional study in 2023. BMC Cardiovascular Disorders **24**(1), 214 (2024). https://doi.org/10.1186/s12872-024-03883-2
7. Kavitha, R., D. G, V., Ram Vishal, B., Shankar, M.U., M, K.A.: Cardiovascular disease prediction using lstm algorithm based on cytokines. In: 2023 4th International Conference for Emerging Technology (INCET). pp. 1–5. Belgaum, India (2023). https://doi.org/10.1109/INCET57972.2023.10170662
8. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
9. Lin, T., Fan, M.: Fmi-caecd: Fusing multi-input convolutional features with enhanced channel attention for cardiovascular diseases prediction. Sensors **24**(22), 7160 (2024). https://doi.org/10.3390/s24227160
10. Mishra, N., Ghorpade, S.B.: Advancing heart disease forecasting: Integrating cnn-lstm models with feature enhancement for improved predictive accuracy. Journal of Angiotherapy **8**(9), 1–6 (2024). https://doi.org/10.25163/angiotherapy.899871
11. Nirob, M.A.S., Bishshash, P., SIAM, A.K.M.F.K., Haque, M.A., Assaduzzaman, M.: Cair-cvd-2025: An extensive cardiovascular disease risk assessment dataset from bangladesh (2025). https://doi.org/10.17632/d9scg7j8fp
12. Qiu, Z., Qiao, Y., Shi, W., Liu, X.: A robust framework for enhancing cardiovascular disease risk prediction using an optimized category boosting model. Mathematical Biosciences and Engineering **21**(2), 2943–2969 (2024). https://doi.org/10.3934/mbe.2024131
13. Rahman, A.U., Alsenani, Y., Zafar, A., Ullah, K., Rabie, K., Shongwe, T.: Enhancing heart disease prediction using a self-attention-based transformer model. Scientific Reports **14** (2024). https://doi.org/10.1038/s41598-024-51184-7
14. Revathi, T.K., Balasubramaniam, S., Sureshkumar, V., Dhanasekaran, S.: An improved long short-term memory algorithm for cardiovascular disease prediction. Diagnostics **14**(3), 239 (2024). https://doi.org/10.3390/diagnostics14030239
15. Talaat, F.M., Elnaggar, A.R., Shaban, W.M., Shehata, M., Elhosseini, M.: Cardiorisknet: A hybrid ai-based model for explainable risk prediction and prognosis in cardiovascular disease. Bioengineering **11**(8), 822 (2024). https://doi.org/10.3390/bioengineering11080822