Article | Received 1 October 2025; Revised 13 November 2025; Accepted 25 November 2025; Published 2 December 2025
https://doi.org/10.55092/bi20250006

# Long short-term memory network with exponential gating mechanism: a deep learning approach for cardiovascular stroke risk stratification

**Ekta Tiwari[1], Dipti Shrimankar[1], Yuvraj Sharma[2,3], Luca Saba[4] and Jasjit S. Suri[3,5,6,7,*]**

[1] Department of Computer Science, Visvesvaraya National Institute of Technology (VNIT), Nagpur 440010, India
[2] Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India
[3] Stroke Diagnostic and Monitoring Division, AtheroPoint™, Roseville 95661, USA
[4] Department of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), Cagliari 09100, Italy
[5] University Center for Research & Development, Chandigarh University, Mohali 140413, India
[6] Symbiosis Institute of Technology, Symbiosis International (Deemed University), Nagpur 440008, India
[7] Department of Electrical and Computer Engineering, Idaho State University, Pocatello 83209, USA

* Correspondence author; E-mail: jasjit.suri@atheropoint.com.

## Highlights:

- Extended LSTM with exponential gating boosts CVD risk stratification accuracy to 99.64%.
- Feature selection identifies 64% key biomarkers for robust prediction.
- Exponential gating enhances temporal feature extraction over standard LSTM models.

**Abstract:** Cardiovascular disease (CVD) is the leading cause of global health issues which requires swift and accurate risk assessment techniques. Present-day risk calculators together with machine learning (ML) models have performed inadequately when it comes to reliability and comprehension of results. We produced a complex deep learning (DL) framework which builds Extended Long Short-Term Memory networks using an exponential gating mechanism (xLSTMeg). The new architecture design controls feature variations while extracting long-term dependencies through increased performance. A Random Forest Regression (RFR) evaluated the most important features while focusing on the top influential biomarkers, which eliminates redundant features. Performance is evaluated using cross-validation protocol and scientifically validated using DL explainability paradigm. The exponential gating mechanism achieved superior performance in CVD risk stratification by at least 6% compared to conventional Long Short-Term Memory (cLSTM) and other ML models. Analysis showed that RFR allows the system to benefits most from the 64% most significant features. The system yielded better results by employing DL algorithms which SHapley Additive exPlanations (SHAP) analysis proved essential features for risk prediction. Our study concludes that our proposed xLSTMeg architecture with optimized features for

CVD risk assessment achieves better robustness and enhanced accuracy.

**Keywords:** cardiovascular disease risk; feature extraction; Long Short-Term Memory with exponential gating; scientific validation and performance evaluation

## 1. Introduction

The healthcare urgency surrounding cardiovascular disease needs swift identification of patients and reliable risk evaluation methods to prevent severe health impacts upon societies worldwide. The morbidity and mortality from heart disease grows severe because it shows major comorbidity effects that are particularly visible in diabetes patients who demonstrate direct cytotoxic outcomes from immunological and endocrinological pathways [1–3]. Carotid ultrasonography is recognized as a valuable screening technique for identifying CVD and cardiovascular events (CVE) by acting as a surrogate biomarker [4,5], while also serving as a highly effective, non-invasive diagnostic tool for evaluating atherosclerotic plaque [6–8]; furthermore, carotid artery ultrasound scans yield key indicators such as carotid intima-media thickness (cIMT) [9–11], total plaque area (TPA) [12–15], and maximum plaque height (MPH) [16].

Research about CVD risk prediction utilized multiple machine learning [17–21] approaches in past however these methods did not yield reliable results when facing diverse patient demographics. Research teams heavily depend on standard algorithms yet these procedures show limited practical success when applied to genuine clinical settings because they fail to replicate nature's patterns in actual clinical data. Traditional approaches encounter severe obstacles to adapt their clinical pattern monitoring to time-based modifications and encounter difficulties in expanding their use through the entire healthcare system.

Our approach applies Adaptive Synthetic Sampling (ADASYN) and random sampling for augmentation and balancing fixes those class discrepancies for quality control as shown in Figure 1 This is then passed through ML methods such as Logistic Regression (LR), Linear Discriminant Analysis (LDA) Gaussian Naive Bayse (GNB), and AdaBoost and DL methods such as cLSTM (conventional LSTM) and our novel design of xLSTMeg (extended LSTM with exponential gating) for CVD risk stratification. xLSTMeg [22–25] is a new innovation in the realm on Artificial Intelligence (AI), which enhances sophistication of cLSTM to identify intricate and minute patterns. Different modelling methods in this diverse approach help build an efficient benchmark system which evaluates algorithms based on their established performance measures. The various analytical methods generate an integrated perspective about model results which enables fair model comparison.

Our study utilizes 10-fold (K10) cross-validation like in some previous studies [26–28], together with extensive power analysis [29,30] to achieve scientific validation because it guarantees statistical strength and reproduction of experimental results. Feature selection depends on a Random Forest Regression [31,32] and SHapley Additive exPlanations explainability [33,34] techniques support the identification of crucial features. The various steps of validation strengthen our experimental results along with making our proposed model more credible. Rigorous statistical tests perform during validation to verify that the model maintains constant performance levels.

The main innovation in our method involves embedding exponential gating capabilities into LSTM architecture to achieve superior results compared to standard LSTM networks. Through this optimization method researchers can enhance both input features and temporal dependencies for better CVD risk

stratification results. The superior gating strategy enhances sequential signal processing which makes it an outstanding solution for modelling complicated clinical data. The algorithm selectively emphasizes essential signals to extract vital sequence patterns as part of its refined information acquisition process.

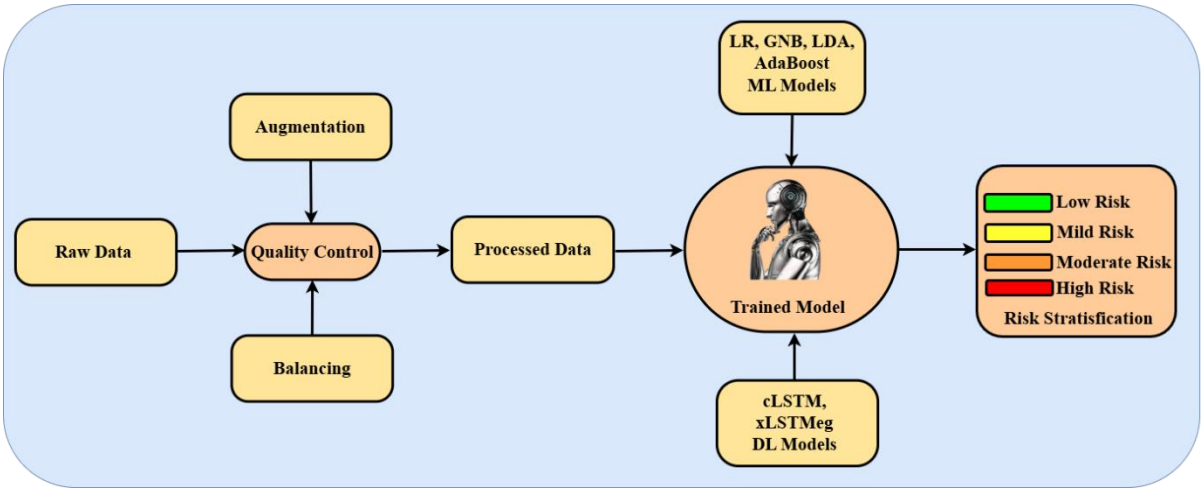


**Figure 1.** Online system design. cLSTM: conventional Long Short-Term Memory; xLSTMeg: extended Long Short-Term Memory with exponential gates; LR: Logistic Regression; GNB: Gaussian Naïve Bayes; LDA: Linear Discriminant Analysis; ML: Machine Learning; DL: Deep Learning.

This research document uses an organized structure for effective presentation of all research findings. Background literature analysis is covered in section 2. The methodology section follows the background literature and is presented in section 3. The results and performance evaluation are clubbed together and is shown in section 4. The scientific validation is discussed in section 5 which covers power analysis and explainability using SHAP analysis, while section 6 presents critical discussions, leading to conclusions in section 7. Such an organized structure enables readers to move systematically through the conceptualization stage to obtain final insights.

## 2. Background literature

CVD risk prediction systems developed significantly throughout the years starting from the classic models generated in the 1980s and 1990s. Risk prediction systems from the initial period relied on the analysis of established risk factors which included both sociodemographic features and biological markers such as gender and age as well as cholesterol measures, blood pressure and habits regarding tobacco use. Wilson *et al.* [35] (1998) introduced fundamental concepts of risk stratification in CVD and their findings shaped both subsequent scientific pursuit and official clinical practices. The established risk models became vital instruments to guide preventive measures that targeted coronary heart disease (CHD).

The focus on risk assessment during the 2000 decade expanded to include multivariate methods that created new assessment models which integrated diabetes and family history risk evaluation and body mass index (BMI) measurements (as described in Conroy *et al.* [36]). The systematic coronary risk evaluation (SCORE) project developed CVD assessment in European populations through its focus on region-specific models for risk evaluation and chart creation [37–39]. The evaluation method delivered systematic assessment instruments that helped clinicians make standardized CVD risk evaluations within

different European healthcare settings to boost practice quality and guideline development. A comparison of conventional calculators was conducted by Jamthikar *et al.* [40].

Around 2010 ML systems emerged in the next phase of development which incorporated random forests (RF) support vector machines (SVM) and neural networks to enhance predictive accuracy when processing extensive and various datasets. The American College of Cardiology (ACC)/American Heart Association (AHA) risk calculator implemented a broader set of risk factors to deliver a better CVD evaluation according to Goff Jr. *et al.* [41] (2014). The approach underscored the necessity of statins as prevention medication by establishing a threshold beyond which doctors should initiate treatment for risk of atherosclerotic cardiovascular disease (ASCVD). The incorporation of imaging data as an essential component became a step forward in developing prediction for assessing risk. Image-derived phenotypes including cIMT also known as CUSIP and TPA and coronary artery calcium scores entered risk models during this stage to increase accuracy. CVD risk evaluation benefits from the incorporation of cIMT and TPA methods as demonstrated in several research studies [42]. The developers used CUSIP measurements to create AtheroEdge™ 1.0 while AtheroEdge™ 2.0 added 10-year risk prediction features and CVD risk classification based on the integration of office-based biomarker (OBBM), laboratory-based biomarker (LBBM) and CUSIP measurements. Scale-space image processing allowed automated border detection of lumen-intima (LI) and media-adventitia (MA) carotid artery far wall boundaries through this technology which led to multiple practical uses [43].

Detrano *et al.* [44] (2008) established through their study that coronary artery calcium (CAC) scores predict future coronary events within four ethnic populations. Research established CAC scores as an independent measure to forecast coronary events which provided information beyond typical assessment factors. Research shows that CAC still provides stable predictive value while specific score distribution differences between groups prove its usefulness as a risk stratification instrument to guide preventive medical care decisions.

A new phase from 2015 to the present supports the integration of deep learning techniques and electronic health records (EHRs) for CVD risk prediction. The current processing of EHR data through convolutional neural networks (CNNs) [45,46] and recurrent neural networks (RNNs) [47,48] extracts sophisticated patterns which leads to increased prediction accuracy. The analysis of retinal fundus photographs by algorithms for cardiovascular risk factor prediction showed feasibility according to Poplin *et al.* [49] (2018). This innovative method creates new possibilities to conduct risk assessments through non-standard methods within cardiovascular medical practice.

## 3. Methodology

Building a reliable xLSTMeg model stands as the main goal of this research to serve as a tool for determining CVD risks. The initial phase of building this goal requires developers to deploy unidirectional then bidirectional LSTM networks before integrating them into the advanced xLSTMeg architecture. Additional improvements must be made to stroke and CVD prediction models designed recently because automated feature selection will optimize their predictive capabilities.

This section is organized into the following seven subsections: Section 3.1 details patient demographics and baseline characteristics; Section 3.2 describes the acquisition of ultrasound scans and the evaluation of intraplaque neovascularization (IPN); Section 3.3 presents the CVD endpoint

determined by angiographic scoring; Section 3.4 outlines the overall architecture of the proposed xLSTMeg system; and finally, Section 3.5 explains the experimental protocol.

### *3.1. Patient demographics and baseline characteristics*

The 500-patient sample received coronary angiography scoring (CAS) assessment which split participants into four behavioural groups: Class 0 represents low risk cases while Class 1 indicates mild risk cases, Class 2 shows moderate risk cases and Class 3 represents high risk cases. Among the entire patient cohort 160 suffered from Acute Coronary Syndrome (ACS), 13 patients received unstable angina (UA) diagnosis and stable angina affected 114 patients alongside 139 patients who underwent stent placement procedures. The research analyzed 39 risk variables through four clusters including 17 measures in the OBBM group combined with seven LBBM factors alongside three radiomics-based biomarkers in the CUSIP cluster and the medication usage category (MedUse). The OBBM cluster encompasses key metrics such as age (mean 64.49 years), gender (349 individuals, 69.8%), obesity (215, 43.0%), ethnicity (486, 97.2%), BMI (31.12 kg/m$^2$), hypertension (338, 67.6%), angina (124, 24.8%), diastolic blood pressure (76.7 mmHg), systolic blood pressure (135.35 mmHg), smoking history (330, 66%), casual smoking (15, 3%), current smoking (100, 20%), previous smoking (218, 43.6%), alcohol consumption (4.94 ± 10.4 per week), family history of diabetes (195, 39.0%), premature CVD in the family (146, 29.2%), and family history of CVD (321, 64.2%). The LBBM cluster records values such as creatinine (83.99 ± 22.6 μmol/L), pre-diabetic status (20, 40%), hyperlipidemia (288, 57.6%), type II diabetes (114, 22.8%), type I diabetes (5, 1.0%), estimated glomerular filtration rate (GFR) (78.96 mL/min/1.73 m$^2$), and overall diabetes (1188, 23.6%). The CUSIP subgroup provides measurements for total plaque area (47.68 mm$^2$), maximum plaque height (2.64 mm), and intraplaque neovascularization (1.16). The MedUse group documents the usage of medications including ACE inhibitors (191, 38.2%), HMG-Co reductase inhibitors (272, 54.4%), angiotensin receptor blockers (45, 9.0%), other antilipemic agents (9, 1.8%), alpha-blockers (30, 6.0%), calcium channel blockers (93, 18.6%), beta-blockers (236, 47.2%), diuretics (99, 19.8%), anti-platelet medications (368, 73.6%), insulin (38, 7.6%), anti-anginals and NSAIDs (81, 16.2%), and non-insulin diabetes medications (72, 14.4%).

The Clinical Utility and Interpretation subsection provides a detailed explanation of how biomarker-driven predictions from the proposed xLSTMeg model align with established cardiovascular risk management guidelines. The model identifies physiologically meaningful biomarkers such as serum creatinine, glomerular filtration rate, age, family history of cardiovascular disease, smoking history, and Type 2 Diabetes status as key predictors of cardiovascular risk. Each of these biomarkers has well-established pathophysiological associations with vascular health and disease progression, thereby enhancing the interpretability and clinical credibility of the model's outputs.

The xLSTMeg model can be integrated within electronic health record (EHR) systems to provide real-time, patient-specific risk stratification, automatically classifying individuals into low-, moderate-, or high-risk categories that correspond to intervention thresholds defined by the American Heart Association and European Society of Cardiology (ESC) guidelines. Model predictions are further interpretable through SHAP analysis, which enables clinicians to visualize the contribution of each biomarker to individual risk estimates. By dynamically updating predictions as new laboratory or clinical data become available, the model supports guideline-consistent decision-making such as personalized

patient counseling, diagnostic follow-up prioritization, and early preventive interventions, thereby extending its role beyond risk prediction to actionable clinical decision support.

### *3.2. Statistical significance of biomarkers in CVD risk stratification*

To better understand the statistical significance of individual biomarkers contributing to cardiovascular disease risk, we conducted a comprehensive p-value analysis of 40 clinical and demographic parameters. Out of these, 21 parameters demonstrated statistically significant associations ($p < 0.05$) with CVD risk. Key features such as age ($p = 0.002$), sex ($p < 0.0001$), hypertension ($p = 0.01$), and smoking history ($p = 0.01$) showed strong correlations. Biomarkers reflecting vascular burden, including mean plaque thickness, total plaque area, and intraplaque neovascularization, were highly significant (all $p < 0.0001$), reinforcing their clinical importance in stratification. Biochemical indicators such as eGFR ($p = 0.012$), creatinine levels ($p = 0.003$), and medication use like beta-blockers ($p = 0.006$) and anti-platelet/anti-coagulants ($p = 0.002$) were also found to be influential. Notably, Type 2 diabetes ($p = 0.026$) and insulin use ($p = 0.044$) emerged as significant metabolic contributors as shown in the Table 1. These statistical insights validate the relevance of selected biomarkers and support their prioritization in the feature optimization process of the proposed xLSTMeg model for robust CVD risk assessment.

**Table 1.** p-value analysis of key biomarkers in CVD risk stratification.

| SN | Parameter | *p*-value | SN | Parameter | *p*-value |
|---|---|---|---|---|---|
| 1 | Total (n) | - | 21 | Diabetes T1D | 0.53 |
| 2 | Age (years)† | 0.002 | 22 | Diabetes T2D† | 0.026 |
| 3 | Sex, n (%)† | <0.0001 | 23 | Diabetes (any)† | 0.051 |
| 4 | Caucasian, n (%) | 0.211 | 24 | eGFR (ml/min/1.73 $m^2$)† | 0.012 |
| 5 | Obesity, n (%) | 0.1 | 25 | Creatinine (µmol/L)† | 0.003 |
| 6 | BMI (kg/m2) | 0.92 | 26 | MPH (mm)† | < 0.0001 |
| 7 | Angina, n (%) | 0.3 | 27 | TPA ($mm^2$)† | < 0.0001 |
| 8 | Hypertension, n (%)† | 0.01 | 28 | IPN† | < 0.0001 |
| 9 | SBP (mmHg)† | 0.025 | 29 | HMG-Co Reductase Inhibitors† | 0.026 |
| 10 | DBP (mmHg) | 0.71 | 30 | Other Antilipemic Agents† | 0.011 |
| 11 | Current Smoker† | 0.006 | 31 | ACE Inhibitors | 0.54 |
| 12 | Casual Smoker | 0.42 | 32 | ARBs Angiotensis | 0.34 |
| 13 | Previous Smoker | 0.85 | 33 | Alpha-Blockers | 0.79 |
| 14 | Smoking Hx† | 0.01 | 34 | Beta-Blockers† | 0.006 |
| 15 | Drinks/wk | 0.88 | 35 | Calcium Channel Blockers | 0.26 |
| 16 | Family Hx of Premature CVD | 0.55 | 36 | Anti-Platelet/Anti-Coagulants† | 0.002 |
| 17 | Family Hx of CVD | 0.19 | 37 | Diuretics | 0.07 |
| 18 | Family Hx of Diabetes | 0.42 | 38 | Anti-Anginals and NSAIDS | 0.8 |
| 19 | Hyperlipidemia, n (%)† | 0.007 | 39 | Insulin† | 0.044 |
| 20 | Pre-Diabetic | 0.86 | 40 | N-I Diabetes Medications | 0.29 |

*3.3. Ultrasound acquisition and intraplaque neovascularization*

Carotid artery assessment by means of ultrasound represents a non-invasive diagnostic tool which provides useful information about coronary artery disease (CAD) risk through artery health evaluation. B-mode focused carotid ultrasonography of patients occurred using a GE Healthcare Vivid E9 system that integrated a 9L-D linear transducer functioning between 2.4 and 10 MHz. The American Society of Echocardiography guidelines enabled researchers to use longitudinal ultrasound scans for collecting two essential image measures known as TPA and MPH. The measurement of the distance between media-adventitia and lumen-intima layers at each neck side defines MPH as a parameter, while TPA indicates the entire carotid bifurcation area under 10 mm examination. Contrast-enhanced ultrasound imaging was used for IPN evaluation through microbubble transit tracking from the adventitial wall to plaque core that received grades from 0 to 3 before scoring involved averaging both sides.

*3.4. Cardiovascular disease endpoint: angiographic score*

The research makes use of angiographic scoring as the core measure to assess CVE. The GE Healthcare Vivid platform 2000 was used by independent expert cardiologists for angiogram assessment during Standard Judkins examination procedure. Experts evaluated the stenosis grades in the left anterior descending (LAD) artery together with the left main and circumflex arteries and the right coronary artery (RCA). Experts designated coronary artery disease severity as minor for values below 19% but deemed it mild between 20% and 49% and moderate at 50%–69% before considering any stenosis above 70% as severe. Special attention was reserved for left main coronary stenoses below 50%. Research demonstrates that coronary angiography functions as the standard for detecting precise CVD risks due to the multiclass ground truth strategy used in our study even though additional imaging techniques are available for evaluation.

*3.5.Deep learning architecture*

The architecture of tandem block diagram shows a comprehensive CVD risk assessment pipeline that integrates three sequential modules. The first module (a1) employs traditional ML classifiers like LR, GNB, LDA, and AdaBoost with hyperparameter tuning to process cardiovascular data as shown in Figure 2 The second module (a2) utilizes a cLSTM model with a sigmoid gating mechanism, a dense layer, and the Adam optimizer to extract deeper temporal features. The third module (a3) implements an xLSTMeg model featuring exponential gating as shown in the Figure 2, followed by a dense layer and Adam optimization, to further enhance feature extraction. Each module computes the same input data, allowing for a comparative evaluation of their performance. The outputs are assessed using a Receiver Operating Characteristic (ROC) curve, which underscores the effectiveness of the integrated, tandem approach in predicting cardiovascular risk.

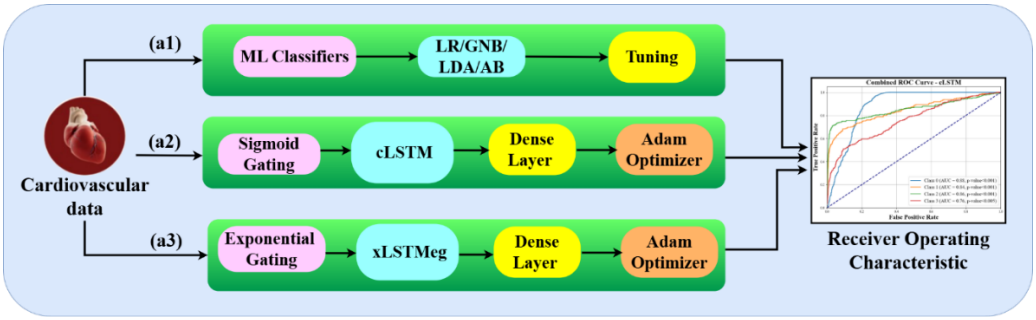


**Figure 2.** Tandem block diagram of different ML and DL Classifiers. cLSTM: conventional Long Short-Term Memory; xLSTMeg: long short-term memory with exponential gates; LR: Logistic Regression; GNB: Gaussian Naïve Bayes; LDA: Linear Discriminant Analysis; ML: Machine Learning. DL: Deep Learning. Heart image curtsy of AtheroPoint LLC California.

(1) Data preparation and pre-processing

The xLSTMeg architecture contains four main operational components. The data pre-processing duty belongs to the first component which works together with the second component managing data partitioning tasks. Training procedures occur offline within the third component and risk assessments for coronary artery disease or Acute Coronary Syndrome prediction take place through the testing datasets within the fourth component.

The three fundamental procedures of data preparation involve (i) data normalization by a traditional scaler which transforms attributes into a 0–1 range and (ii) selecting powerful features through RFR method. The data receives two augmentation techniques and ADASYN and random sampling to increase its size to 5000 entries since limited size might cause overfitting.

The Sklearn library in Python enabled extensive pre-processing of data records before conducting the initial division of the dataset. The quality control process implemented normalization combined with scaling and used MinMaxScaler() to transform features into standardized values between 0 and 1. The extraction and selection methods ran exclusively inside the ML framework rather than DL system.

(2) xLong Short-Term Memory's exponential gating mechanism

The xLSTMeg variant swaps the sigmoid input gate with an exponential activation to accentuate salient inputs. At each step, the input gate $i_t = exp\ (W_{xi}x_t + W_{hi}h_{t-1} + b_i)$ in Equation (1), while the forget $f_t$ in Equation (2) and output $o_t$ in gates remain sigmoid-based. The candidate cell $\hat{c}_t$ and cell update $c_t$ follow the standard tanh and gated-sum operations in Equation (3) and Equation (4). The hidden state $h_t$ is then $o_t \odot \tanh(c_t)$ shown in Equation (5) and Equation (6). This exponential gating enhances sensitivity to critical feature changes, leading to more discriminative memory updates.

$i_t$, $f_t$, $o_t$: input (exp-activated), forget, and output gates; $W_{xi}$, $W_{hi}$, $b_i$; $W_xf$, $W_hf$, $bf$; $W_{xo}$, $W_{ho}$, $b_o$: corresponding weights/biases; $\hat{c}_t$: candidate cell via tanh ($W_xcx_t + W_hch_{t-1} + bc$); $c_t$, $h_t$: updated cell and hidden state.

$$Input\ Gate\ (i_t) = exp(W_{xi\ xt} + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$Forget\ Gate(f_t) = \sigma(W_{xf\ xt} + W_{hf}h_{t-1} + b_f) \quad (2)$$

$$Candidate\ Cell\ State\ (\hat{c}) = \tanh(W_{xc\ xt} + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$Cell\ State\ Update\ (c_t) = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \quad (4)$$

$$Output\ Gate\ (o_t) = \sigma(W_{xo\,xt} + W_{ho} h_{t-1} + b_o) \quad (5)$$

$$Hidden\ State\ (h_t) = o_t \odot \tan h(c_t) \quad (6)$$

### *3.6. Experimental protocols*

The two sets of experimental protocols existed to validate our hypotheses: (i) Loss and accuracy curves (ii) Evaluation of xLSTMeg and (ii) Feature Selection via RFR. Each testing protocol incorporates a strict cross-validation system which splits the database into training and testing portions using different validation strategies starting from K2 (50:50) through K4 (75:25) and K5 (80:20) to K10 (90:10) cross-validation approaches, based on the training to testing ratios. The experimental protocols are arranged as Experimental Protocol 1 (EP1), Experimental Protocol 2 (EP2) and Experimental Protocol 3 (EP3).

(1) Experiment #1: effect of augmentation

The experiment was conducted to investigate the influence of different levels of data augmentation using ADASYN on the performance of the proposed xLSTMeg model. Various augmentation strategies were sequentially applied to the training dataset to enhance sample diversity and mitigate overfitting. The process began with the raw dataset and progressively incorporated multiple augmentation cycles (from 2 × to 9 ×), simulating expanded training distributions without altering the inherent data characteristics. This evaluation aimed to determine the augmentation threshold beyond which additional synthetic data yield diminishing performance gains, providing insight into the model's sensitivity to dataset diversity and representation balance.

(2) Experiment #2: loss and accuracy curves

The EP2 teaching standard emphasizes how essential it is to track loss and accuracy curves as model training progresses. The loss curve shows error minimization abilities of the model while accuracy curve demonstrates correct prediction capabilities. Observing these trends indicate progress to the model learning process while signalling the need for potential model adjustments. Visual feedback serves as a basic tool for optimizing training approaches to achieve reliable system performance.

(3) Experiment #3: evaluation of xLSTMeg

Performance evaluation of the xLSTMeg architecture with integrated exponential gating occurs as the main focus of EP3. The designed protocol uses both unidirectional and bidirectional LSTM networks that integrate exponential gating mechanisms. The protocol seeks to prove that the new gating technique leads to better CVD risk assessment through superior accuracy and prediction stability compared to standard LSTM networks and some state-of-the-art models [50,51]. The model receives systematic assessment through four cross-validation schemes named K2, K4, K5 and K10 at various training-to-testing ratios. Data analysis includes precise performance measurements where accuracy and area under the curve (AUC) serve as comparison metrics to validate the superiority of exponential gating.

(4) Experiment #4: feature selection via Random Forest Regression

A comparison between the original feature set and the optimized subset of essential features demonstrates that automated feature selection using Shapley values and Random Forest Regression effectively simplifies the model, mitigates overfitting, and reduces overall system complexity. The detailed feature importance rankings and selection results are provided in the Supplementary Material.

Appendix A outlines the loss function optimization, training environment, and hyperparameter tuning that drive our model's performance. Appendix B details the evaluation metrics accuracy, precision, sensitivity, F1 score, and AUC used to assess the model's predictive power.

## 4. Results and performance evaluation

This part contains our experimental study findings together with a thorough assessment of the xLSTMeg model performance. The main experimental results appear in Section 4.1 where accuracy and loss curves demonstrate the findings while different model accuracies get compared and RFR provides feature extraction insights. The evaluation of model performance occurs through both K-fold cross-validation scheme assessments and ROC curve analysis in Section 4.2.

### *4.1. Results*

The results were obtained to validate our hypotheses. Each result incorporates a strict cross-validation system which splits the database into training and testing portions using different validation strategies starting from K2 through K4 and K5 to K10 cross-validation approaches.

(1) Results of experiment #1: loss and accuracy curves

The results presented in Figure 3 demonstrate a clear improvement in model accuracy with the introduction of augmentation. The xLSTMeg model showed a consistent upward trend in performance from the raw dataset to moderate augmentation levels, reaching its highest accuracy with 4 × augmentation. Beyond this point, further augmentation produced marginal or negligible changes, indicating that the model had already achieved an optimal balance between data diversity and representational learning. Overall, these findings confirm that moderate augmentation significantly enhances model generalization and stability, while excessive augmentation may offer limited additional benefit for the xLSTMeg architecture.

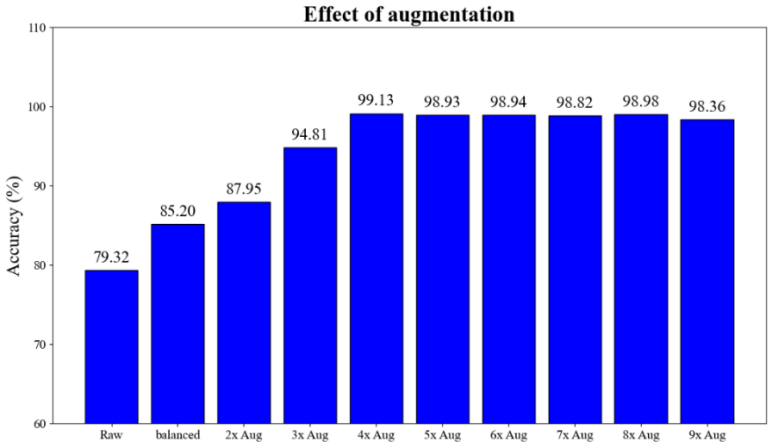


**Figure 3.** Effect of augmentation using K10 cross-validation protocol.

(2) Results of experiment #1: loss and accuracy curves

Our custom feature loss function enabled training that achieved convergence to a loss level of loss level of less than 1% which shows the archives generalization. Our method demonstrates its effectiveness in optimizing feature representations by showing both steady error reduction and performance enhancement during all training epochs according to the accuracy and loss curves as shown in Figure 4 and Figure 5.

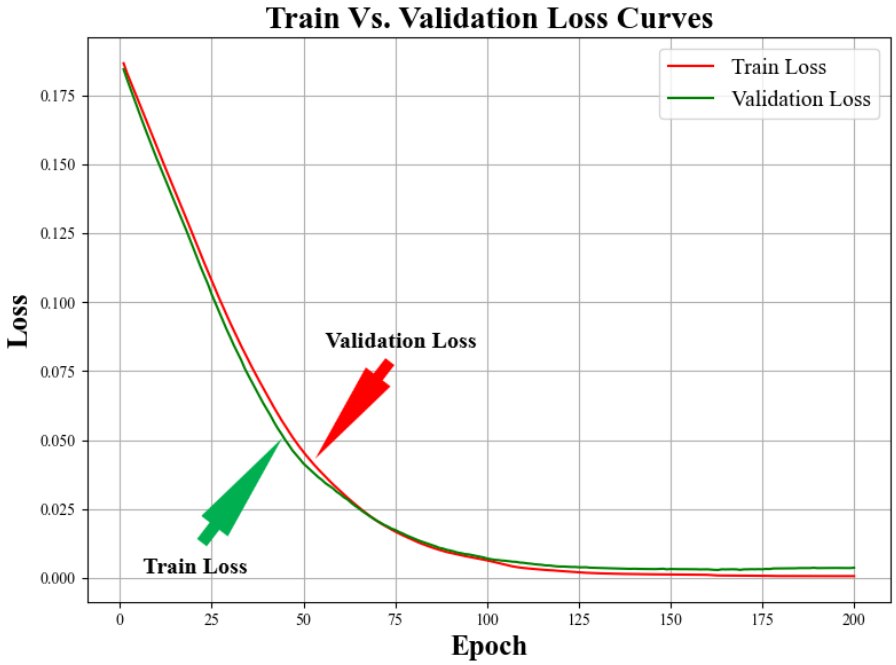


**Figure 4.** Train *vs.* Validation loss curve for xLSTMeg limiting to 200 epochs.

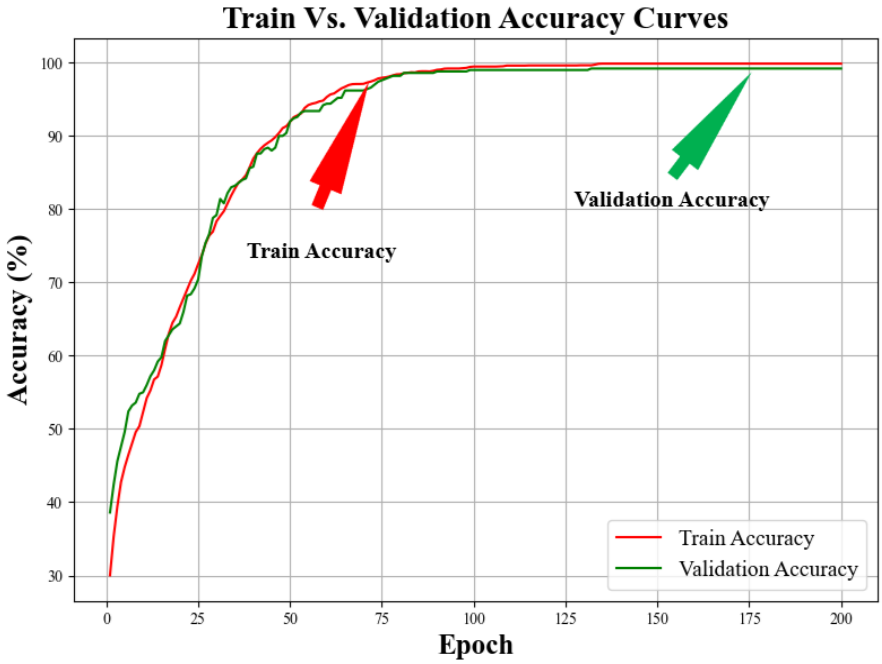


**Figure 5.** Train *vs.* Validation accuracy curve for xLSTMeg limiting to 200 epochs.

(3) Results of experiment #2: comparison of DL *vs.* ML models

The proposed xLSTMeg model demonstrated an excellent accuracy of 99.64 which is better than the traditional machine learning and the recent state-of-the-art deep learning models. Table 2 indicates that the classical machine learning models of Gaussian Naive Bayes, Logistic Regression, Linear Discriminant Analysis, and AdaBoost had relatively lower accuracies of 38.24, 56.34, 55.61, and 56.00, respectively. The traditional cLSTM made an accuracy of 87.86, whereas the transformer-based models (FT Transformer and Tab Transformer) reached an accuracy of 89.18% and 95.18%, respectively. Regardless of these competitive scores, xLSTMeg model was always better than all the counterparts in terms of all performance measures, such as precision, sensitivity, F1-score, and AUC, which proves its higher learning ability in cardiovascular disease risk prediction. The fact that the AUC of 0.9989 is significantly higher than that of cLSTM (AUC = 0.9457) as well as other transformer-based methods also confirms the fact that the proposed architecture has a better discriminative power and representational efficiency.

**Table 2.** Performance metrics for machine learning and deep learning models.

| **Performce metrics for different ML and DL models using K-10 cross-validation protocol** | | | | | | |
|---|---|---|---|---|---|---|
| Models | ACCU (%) | PREC (%) | SENS (%) | F1 (%) | *p*-value | AUC [0–1] |
| Gaussian NB | 38.24 | 53.33 | 38.24 | 29.92 | 0.05 | 0.6703 |
| Logistic Regression | 56.34 | 56.38 | 56.34 | 56.26 | 0.005 | 0.7881 |
| LDA | 55.61 | 55.73 | 55.61 | 55.39 | 0.005 | 0.7737 |
| AdaBoost | 56 | 55.89 | 56 | 55.6 | 0.005 | 0.8169 |
| cLSTM | 87.86 | 87.8 | 87.65 | 87.56 | 0.001 | 0.9457 |
| xLSTMeg | 99.64 | 99.64 | 99.64 | 99.64 | 0.001 | 0.9989 |
| FT Transformer | 89.18 | 89.37 | 89.78 | 89.76 | 0.001 | 0.9489 |
| Tab Transformer | 95.18 | 95.72 | 95.78 | 95.42 | 0.001 | 0.9728 |

cLSTM: conventional Long Short-Term Memory; xLSTMeg: extended Long Short-Term Memory with exponential gating; LR: Logistic Regression; GNB: Gaussian Naïve Bayes; LDA: Linear Discriminant Analysis; ML: Machine Learning; DL: Deep Learning; ACCU: Accuracy; PREC: Precision; SENS: Sensitivity; F1: F1-Score; AUC: Area-under-the-curve.

(4) Results of experiment #3: feature selection via Random Forest Regression

As shown in Figure 6, RFR feature ranking, the ML and DL model AUC tracks and the increase in selected features to maximum performance at 25 critical variables and then the decrease in additional non-critical features. Based on the RFR analysis, 64% (25/39 multiplied 100 is 64) of the top-ranked features have the majority of the predictive power and thus requires the organizations to concentrate on these few critical aspects to be as effective as possible and remove performance-reducing questionable data features.

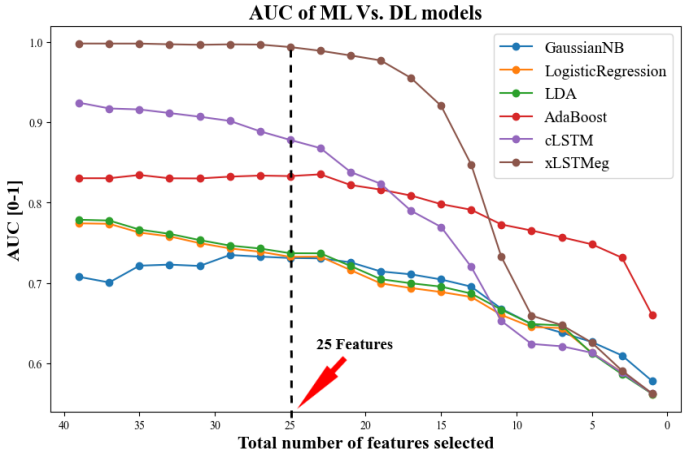


**Figure 6.** Area-under-the-curve for the number of features selected in increasing order. cLSTM: conventional Long Short-Term Memory; xLSTMeg: long short-term memory with exponential gates; LR: Logistic Regression; GNB: Gaussian Naïve Bayes; LDA: Linear Discriminant Analysis ML: Machine Learning; DL: Deep Learning.

### *4.2. Performance evaluation*

To comprehensively assess the reliability and predictive capability of the proposed xLSTMeg model, a series of experiments were conducted focusing on model validation and classification accuracy. Specifically, the K-fold cross-validation and confusion matrix analyses were performed to evaluate the model's robustness, generalization, and effectiveness in distinguishing cardiovascular disease risk categories.

(1) Effect of K-Fold Cross-Validation

To assess the robustness and generalizability of the xLSTMeg model, we employed various K-fold cross-validation schemes, specifically K2, K4, K5, and K10. Our experiments showed that increasing the number of folds provided a more comprehensive evaluation of the model's performance by systematically varying the training and testing splits as shown in Figure 7. This approach led to a slight improvement in accuracy and a reduction in variance across different data partitions, thereby confirming that a higher number of folds contributes to a more stable and reliable model evaluation.


**Figure 7.** Effect K-fold cross-validation on xLSTMeg.

(2) Confusion Matrix Analysis:

We have examined the classification success rate for the cLSTM and xLSTMeg models relied on confusion matrix assessments. Model predictions showed balanced results with only a small number of incorrect and correct classifications for both true positives and negatives as shown in the Figure 8 and Figure 9. The specific evaluation proves the models accurately detect CVD risk in patients or non-CVD risk patients.

(3) Receiver Operating Characteristic Analysis:

Additional analysis of the cLSTM and xLSTMeg models' discriminative capability was performed through ROC curve assessment. The true positive rate of the model remained high across all threshold values while performing with minimal false positive occurrences as shown in the ROC curves. Figure 10 shows the cLSTM *vs.* xLSTMeg shown in Figure 11, which shows high AUC result which indicates its effectiveness in distinguishing between patients with and without CVD risk.

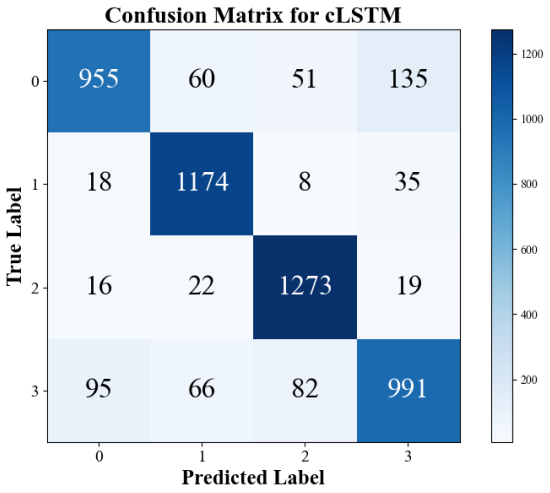


**Figure 8.** Confusion Matrix for cLSTM using K10 cross-validation protocol. cLSTM: conventional Long Short-Term exponential gating.

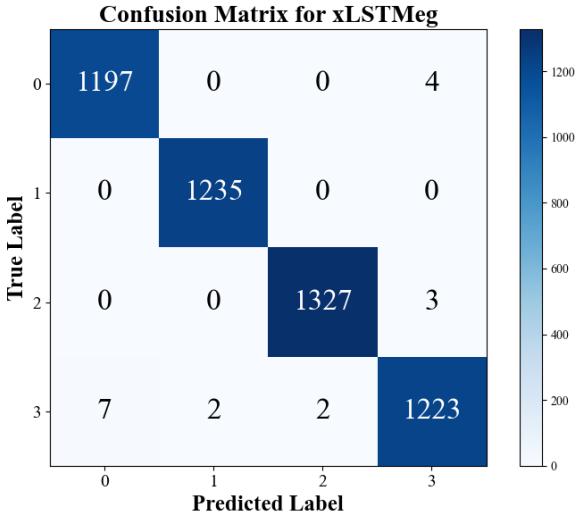


**Figure 9.** Confusion Matrix for xLSTMeg using K10 cross-validation protocol. xLSTMeg: extended Long Short-Term Memory with exponential gating.

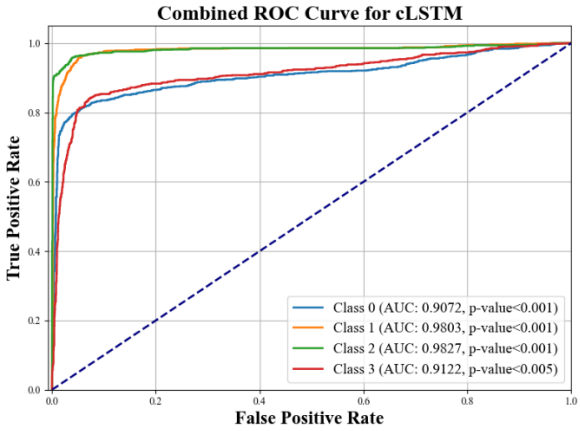


**Figure 10.** Combined ROC curve of cLSTM for all the 10 cross-validation folds. cLSTM: conventional Long Short-Term Memory.

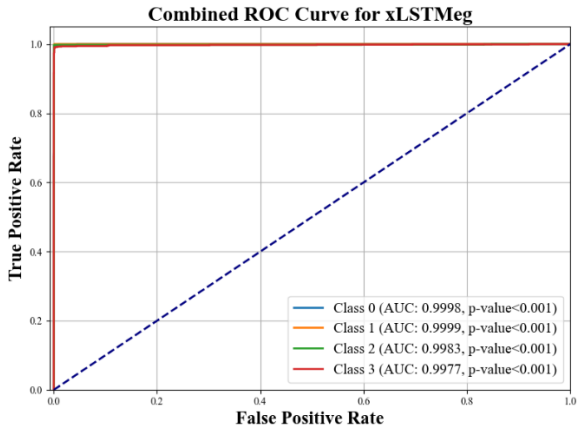


**Figure 11.** Combined ROC curve of xLSTMeg for all the 10 cross-validation folds. xLSTMeg: Long Short-Term Memory.

## 5. Scientific validation and explainability

The two fundamental elements of our validation procedure include Power Analysis with Explainability assessment. A statistical strategy determines the necessary sample size to detect relevant performance variations in models through the dataset. The article explains method interpretation by describing which attributes most significantly affect the model output.

### *5.1. Power analysis*

A power analysis ensured that our dataset could identify genuine variations within the models' performance metrics [37,38]. Our team applied established calculation approaches for determining sample size which combined critical process needs and energy requirements [39,40]. A 5000-sample dataset of augmented data will serve as an accurate platform to evaluate performance metrics according to the provided estimation. Our measures achieve reliability through the validation technique shown in Equation (7) since it evaluates both sample consistency and observed difference magnitude:

$$N^* = \frac{z^2 \times p \times (1-p)}{MoE^2} \quad (7)$$

Here, $N^*$ represents the minimum required sample size, $z = 0.99$ denotes the Z-score corresponding to the desired confidence level, $p = 0.5$ (worst case) the assumed proportion or variability, and $MoE = 0.02$ stands for the margin of error.

### *5.2. Statistical testing*

Initial statistical validation was done with various significance tests, including Chi-Square, Wilcoxon Signed-Rank, Independent t-Test and Mann-Whitney U tests. Table 3 showed that the p-values of all the analyses were more than 0.05, which means that there were no statistically significant differences between the biomarkers that were evaluated. This finding implies that personal characteristics are not adequate to distinguish between patients with and without cardiovascular disease. As a result, it supports the importance of using the sophisticated nonlinear modeling methods, including the suggested xLSTMeg architecture, to successfully identify multidimensional and complex interdependencies in the data.

**Table 3.** Statistical testing for xLSTMeg deep learning model.

| SN | Statistical test | Statistic value | *p*-value | Interpretation |
|---|---|---|---|---|
| 1 | Chi-Square Test | 0.192 | 0.4228 | p > 0.05 |
| 2 | Wilcoxon Signed-Rank Test | 516 | 0.129 | p > 0.05 |
| 3 | Independent t-Test | 0.4978 | 0.628 | p > 0.05 |
| 4 | Mann–Whitney U Test | 0.628 | 0.598 | p > 0.05 |

### *5.3. External validation*

In order to further test the generalization ability of the proposed models, an external validation on a separate Canadian clinical cohort was done. Table 4 summarizes the comparative performance of the experimental (training) and validation datasets. The standard cLSTM model has shown a decrease in accuracy of 87.38 per cent to 83.94 per cent and the corresponding decrease in ROC of 0.9457 to 0.8797, which is a moderate generalization gap of about 3.9 per cent and 7.0 per cent respectively. The suggested Bi-xLSTMeg model, in turn, was more robust overall, with an accuracy of 92.38% and a ROC of 0.9474 on the external cohort, even though its absolute and relative decrease in performance compared to the experiment was 7.29 and 5.16, respectively. These findings indicate that both models show a natural decrease in their performance, when applied to unseen data, but Bi-xLSTMeg network is always better than the baseline cLSTM in all the metrics, which proves its superiority in generalization and stability on external clinical data.

**Table 4.** Unseen external validation.

| **Performance comparison of experimental data and testing on validation data** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Models | Experimental Data | | Validation Data | | Accu Diff | | ROC Diff | |
| | Accuracy (%) | ROC [0–1] | Accuracy (%) | ROC [0–1] | Abs | Perc (%) | Abs | Perc (%) |
| cLSTM | 87.38 | 0.9457 | 83.94 | 0.8797 | 3.44 | 3.93 | 0.066 | 6.97 |
| xLSTMeg | 99.64 | 0.9989 | 92.38 | 0.9474 | 7.26 | 7.28 | 0.0515 | 5.15 |

## 6. Discussion

Our research brings forward unique advanced solutions which operate jointly to improve processes for CVD risk assessment. Specifically, (i) the exponential gating mechanism significantly improves CVD risk stratification by enhancing feature modulation, which allows the model to capture subtle interactions that might otherwise be overlooked, ultimately resulting in a more nuanced understanding of patient risk profiles; (ii) the top 65% of features substantially drive the performance of the RFR, indicating that a focused subset of the most informative predictors is primarily responsible for the model's high accuracy and robust predictive capabilities, thereby emphasizing the importance of careful feature selection; (iii) eliminating redundant features not only reduces computational complexity but also enables efficient training on lower-powered devices such as Raspberry Pi, which is particularly beneficial in remote locations where computational resources are limited, making advanced predictive analytics more accessible; (iv) DL methods consistently outperform traditional ML techniques, as evidenced by improved performance metrics and enhanced capability to model complex nonlinear relationships inherent in medical data; (v) SHAP explainability validates the importance of several key features identified by the RFR, thereby offering transparent insights into model decisions and fostering trust among clinical users; (vi) increasing the number of cross-validation folds slightly enhances overall model performance by providing a more rigorous and comprehensive evaluation framework that reduces overfitting and better generalizes the findings; (vii) the conducted power analysis supports the statistical robustness of the system, ensuring that the results are reliable and that the model's performance improvements are statistically significant; and (viii) hyperparameter tuning for both cLSTM and xLSTMeg confirms that a learning rate of 0.000008, a batch size of 32, and 200 epochs provide optimal training outcomes, establishing these settings as the ideal configuration for achieving the best balance between training efficiency and predictive accuracy.

(1) Benchmarking

The benchmarking table gives an extensive summary of diverse studies that measure AI model effectiveness for CVD risk assessment. A comprehensive table reveals essential information about the dataset origin with its size and the number of features (NOF) and the employed AI models combined with different validation frameworks (K2 to K10 cross-validation) and shows performance metrics (ACCU and AUC) and includes information about unseen test data and power analysis. The consolidated structure enables researchers to evaluate the ML approaches and DL methods between assorted investigative groups and geographical locations. The table establishes an effective structure for presenting essential data points to enable researchers to study both experimental methods and performance data throughout different studies.

Bhagawati *et al.* [52] put together XGBoost, UniDL, BiDL, and HDL models into an ensemble system with K2, K3, K4, K5, and K10 cross-validation folds to achieve 97.25% accuracy and 0.98 AUC. The research conducted by Jamthikar *et al.* [53,54] in rows 2 and 3 use K10 cross-validation to examine models composed of XGBoost, RF, SVM (row 2) and RF, RSF (row 3) which produces respective accuracy results of 95% and 96% as shown in Table 5. The study examines the feature selection approaches between Jamthikar *et al.* [53] and our proposed RFR method that both aim to forecast coronary artery disease and Acute Coronary Syndrome. The techniques produced significant performance-driven results through their isolation of key features. The robust predictors of TPA and age and renal function and key clinical parameters remain important factors which generate reliable results regardless of the

methodological selection. The RFR method uses continuous variables including creatinine and BMI and blood pressure to understand complex decision tree interactions while Jamthikar *et al.*'s mutual information approach focuses on categorical medication usage through statistical dependency measures. The shared predictive value of these key features becomes evident when combining different analysis methods because both techniques end up selecting similar critical predictors for CAD/ACS risk stratification.

The validation procedures along with high performance results from these studies reinforce ensemble and tree-based methodologies for cardiovascular risk evaluation. These studies demonstrate how combining different algorithms with validation techniques creates better grounds for future advancements in identifying CVD risk factors.

Further, studies from rows 4 to 8 illustrate a diverse range of approaches and dataset scales. For instance, Unnikrishnan *et al.* [55] (row 4) and Alaa *et al.* [56] (row 5) used SVM-based and multiple ML techniques, though their reported performance metrics (e.g., AUC of 0.71 and around 0.724, respectively) suggest relatively lower predictive power. In contrast, Konstantonis *et al.* [57] from Greece in row 8 demonstrated high performance with an accuracy of 98.49% and an AUC of 0.98 using traditional ML methods (RF, SVM, LDA) as shown in Table 5. Complementing these, Johri *et al.* [58] row 9) adopted a DL strategy using RNN and LSTM models, achieving 95.34% accuracy and a notably high AUC of 0.99. Notably, the high-performing studies: rows 1, 8, and 9 are comparable, indicating that both advanced ML ensembles and DL architectures can yield robust results in this domain. Additionally, these studies collectively emphasize the importance of selecting robust validation techniques and algorithms to ensure reliable and relevant outcomes.

The proposed approach (row 10) presents a new DL structure xLSTMeg, that was built and tested on a much bigger dataset (5000 Canadian patients with 39 features). The xLSTMeg model has a higher accuracy (99.64) and AUC (0.9989), which is on of the highest performing, however, it has a larger and more varied training sample and this could improve the generalizability of this model in the real world. Besides, the suggested strategy, with K10 cross-validation and backed by power analysis, presents a novel alternative that better represents complicated time-related dynamics than traditional LSTM-based algorithms. This makes the xLSTMeg a competitive and promising risk prediction and cardiovascular risk stratification tool in coronary artery disease. Moreover, excellent level of data scale and solid assessment of the xLSTMeg model give a strong argument of its adoption in clinical practice.

(2) A Special note on extended Long Short-Term Memory with exponential gating

The extended form of simple LSTM architecture *i.e.* xLSTMeg deploys usage of exponential gating operations in the components. The sigmoid activation functions of input and forget gates and output gate are replaced with an exponential function which performs the calculations and normalizes the outputs to obtain the correct ranges of gating functions. The updated version of the cell state combines the old components of the cell state with the new candidate state though gives control over the new exponential gate inputs to process information dynamics. Deep recurrent models are efficient in learning as well as efficient in gradient transmission because they have residual connections that help them to identify long-term dependencies. The xLSTMeg offers a greater control of the temporal signal due to the correct time-dependent operations and high-quality feature representation hence augmenting the data sequence modulation effect. Such a new form of architecture is presenting the combination of the effectiveness of interpretation and credible performance standards of DL in the assessment of CVD risk.

**Table 5.** Benchmarking table showing studies that were implemented for ensemble and solo machine and deep learning models for stocks detection.

| SN | Author | Input Dataset | | | Classification Model | | | | | Performance | | Validation | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Citation | Source/ Country | #Patients/ #Images | NOF | AI Model | Cross-Validation | EOTD | AI Type | Best Model | ACCU (%) | AUC [0–1] | Unseen Data | Power Analysis |
| 1 | Bhagawati *et al.* [52] | Canada | 500 | 39 | XGBoost, UniDL, BiDL, HDL | K2, K3, K4, K5 and K10 | ✓ | EDL | HDL | 97.25 | 0.98 | ✓ | ✘ |
| 2 | Jamthikar *et al.* [53] | Canada | 500 | 39 | XGBoost, RF, SVM | K10 | - | EML | - | - | 0.95 | ✘ | ✓ |
| 3 | Jamthikar *et al.* [54] | Canada | 500 | 39 | RF, RSF | K10 | - | EML | - | - | 0.96 | ✘ | ✓ |
| 4 | Unnikrishnan *et al.* [55] | - | 2.4 K | 9 | SVM | K5 | ✘ | ML | SVM | - | 0.71 | ✘ | ✓ |
| 5 | Alaa *et al.* [56] | - | 423.6 K | 473 | SVM, GBM, RF, | K10 | ✘ | EML | | - | 0.724 | ✘ | ✘ |
| 6 | Trebeing *et al.* [61] | - | 378 Images | 20 | Unet | - | - | DL | Unet++ | - | - | ✘ | ✘ |
| 7 | Jain *et al.* [62] | - | 379 | 97/970 | SegNetUNet, SegNet, Unet + SegNet | K5 | ✘ | DL | Unet++ | 88.9 | 0.91 | ✘ | ✘ |
| 8 | Konstantonis *et al.* [57] | Greek | 542 | 46 | RF, SVM, LDA | K2 | ✓ | ML | - | 98.49 | 0.98 | ✘ | ✘ |
| 9 | Johri *et al.* [58] | Canada | 500 | 39 | RNN, LSTM | K10 | ✘ | DL | LSTM | 95.34 | 0.99 | ✘ | ✓ |
| 10 | Proposed | Canada | 5000 | 39 | xLSTMeg | K10 | ✘ | DL | xLSTMeg | 99.64 | 0.99 | ✘ | ✓ |

RF: Random Forest; KNN: K Nearest Neighbour; SVM: Support Vector Machine; LR: Linear Regression; NB: Naïve Bayse; XGB: Extreme Gradient Boosting; cLSTM: conventional Long Short-Term Memory; xLSTMeg: extended Long Short Term Memory with exponential Gating; ET: Extra Trees; DT: Decision Tree; ACCU: Accuracy; PREC: Precision; F1: F1 Score; AUC: Area-under-the-curve; AUGU: Augmentation; RNN: Recurrent Neural Network; GRU: Gated Recurrent Unit; CV: Cross Validation; NOF: Number of Features; EOTD: Effect of Training Data.

(3) Strengths, weakness and extensions

There are several key benefits that are beneficial to users since xLSTMeg applies its operational features and design features. The exponential gating system allows the use of better sequence analytics accuracy because it improves time-based input indications and also eliminates unwarranted noise. The model has the advantage of the fact that it has included layer normalization with dropout regularization and residual connections that it employs to ensure that the gradient domain is optimized to allow deep layers to learn efficiently. This architecture provides optimum dependency detection over time and this is what makes it the best choice when identifying complex temporal patterns in cardiovascular risk stratification systems. Due to the premeditated design decisions in xLSTMeg, the system framework works well in the creation of predictions and also enables interpretation abilities.

The new system xLSTMeg has several limitations in its use. (i) The xLSTMeg architecture is more complex and presents more challenges to hyperparameter exploration to less experienced deployers. (ii) Exponential gating introduces a useful feature modulation to xLSTMeg but causes the model to be vulnerable to normalization problems that require specific calibration to prevent numerical instability. (iii) The execution time requirements are greater when gating operations are done in conjunction with regularization in applications with limited computing power that decreases real time processing capability in particular hardware settings. (iv) Exponential gating mechanism model decisions have to be studied intensively with SHAP methods to make interpretability decisions but require further analysis to interpret their impact on predictions.

The potentials of xLSTMeg hold various prospective prospects that can be used by researchers to develop in the future. Future studies ought to focus on the attention-based approaches to the refinement of the temporal segment selection since the specified improvement would increase the predictive power of the model. The extension of research would require the redesign of the architecture to accommodate the unseen data sources of multiple sources thereby accommodating a larger analysis in the healthcare environment particularly in real-time patient monitoring and in distance medical testing conditions. Automated methods of hyperparameter optimization might make tuning processes easier and enable the model to be extended to a broader variety of fields of application. The suggested extensions will not only enhance current performance measures but also expand the current applications of xLSTMeg to various sequential data applications.

## 7. Conclusion

The combination of xLSTMeg network with exponential gating operations produces significant improvement in medical predictions for CVD diagnosis according to research findings. Long-term dependency recognition along with feature control establishes the system as a better alternative for regular risk stratification strategies. RFR feature selection method enables the xLSTMeg system to run on small portable devices since it performs feature selection and removes unnecessary components. The SHAP analysis verifies that important features in the model improve both performance results while making the interpretation easy to apply in clinical settings.

The analysis proven the trustworthy nature of xLSTMeg architecture [59,60] by performing multiple cross-validations over several validation stages along with power analysis coupled with minimal learning rate adjustments and proper training of moderate-sized batches. The positive findings from our study show that CVD risk stratification enhanced which can benefit the practical patient screening and can

prove beneficial early intervention systems. These optimized features enable faster calculations which aid in advancing DL methods to create cost-effective clinical cardiovascular risk assessment tools that improve medical care for various healthcare settings.

## Data availablity statement

The data used in this manuscript is proprietary to AtheroPoint LLC, and is not publicly available.

## Acknowledgements

The data was collected by Cardiovascular Imaging Network at Queen's, Queen's University, 76 Stuart Street, FAPC3, Kingston, ON, K7L 2V7, Canada and shared with AtheroPoint™, Roseville, CA, USA under legal agreements between the two institutes. All software designed and developed in Artificial Intelligence. Deep Learning is owned and solely belongs to AtheroPoint™. All Deep Learning intellectual property is patented and owned by AtheroPoint™, CA, USA.

## Authors' contribution

Conceptualization, ET and YS; Methodology, ET; Software, DS; Validation, ET, DS and JSS; Formal analysis, ET; Investigation, ET; Resources, ET, DS and JSS; Data curation, ET; Writing—original draft preparation, ET, DS and LS; Writing—review and editing, ET, DS and LS; Visualization, LS; Supervision, LS and JSS; Project administration, JSS. All authors have read and agreed to the published version of the manuscript.

## Conflicts of interests

The authors declare no conflict of interest.

## Abbreviations

| SN | Abbreviation | Definition | SN | Abbreviation | Definition |
|---|---|---|---|---|---|
| 1 | ACCU | Accuracy | 15 | IPN | Intraplaque Neovascularization |
| 2 | ADASYN | Adaptive Synthetic Sampling | 16 | LR | Linear Regression |
| 3 | AUC | Area-under-the-curve | 17 | ML | Machine Learning |
| 4 | AUGU | Augmentation | 18 | MoE | Margin of Error |
| 5 | BMI | Body Mass Index | 19 | PA | Power Analysis |
| 6 | cLSTM | Conventional Long Short-Term Memory | 20 | PREC | Precision |
| 7 | DL | Deep Learning | 21 | RFR | Random Forest Regression |
| 8 | EML | Ensemble Machine Learning | 22 | RNN | Recurrent Neural Network |
| 9 | ET | Extra Trees | 23 | ROC | Receiver Operator Characteristic |
| 10 | F1 | F1-Score | 24 | SENS | Sensitivity |
| 11 | FL | Feature Loss | 25 | SVM | Support Vector Machine |
| 12 | GB | Gradient Boosting | 26 | TPA | Total Plaque Area |
| 13 | GFR | Glomerular Filtration Rate | 27 | Uni | Uni-directional |
| 14 | GNB | Gaussian Naïve Baise | 28 | xLSTMeg | Extended Long Short-Term Memory with exponential gating |

## Symbols

| SN | Symbol | Definition | SN | Symbol | Definition |
|---|---|---|---|---|---|
| 1 | $L_{FE}$ | Feature Loss | 7 | TP | True Positives |
| 2 | N | Total number of samples | 8 | TN | True Negatives |
| 3 | $y_i$ | Ground truth label for sample i | 9 | FP | False Positives |
| 4 | $a_i$ | Predicted probability | 10 | FN | False Negatives |
| 5 | $\widehat{y_i}$ | Predicted feature value for sample | 11 | AUC | Area-under-The-curve |
| 6 | $y_i^{true}$ | Ground truth feature value for sample i | 12 | ∫(t) | Integral Function at time t |

## Appendix A: loss function optimization and training environment

The training process employed these loss functions at the same time to enhance both prediction accuracy for classes and representation quality of attributes. The trainer executed its sessions by using predefined epochs alongside defined learning rate and dropout rate and batch size restrictions. The entire experimental framework ran from Python version 3.9. The NVIDIA RTX 3090 GPU together with Ryzen-7 5000 series processor existed on a workstation which conducted efficient training operations.

(1) Loss function

The training involved utilizing a specific feature loss function to enhance the network's feature representation capabilities. The determination of this loss requires a two-step method that starts by stripping the ground truth labels of extra dimensions before transforming them to integer format as shown in Equation (A.1). The sparse labels transform into one-hot vectors through an operation that defines the vector depth according to class number requirements for the model output requirements. The loss calculation determines average deviations between prediction and one-hot encoded labels through Mean Squared Error measurements. The model learns more exact feature associations through minimizing squared difference quantities that lead to better feature representation and performance improvement.

Feature Loss (FL):

$$L_{FL} = \frac{1}{N}\sum_{(i=1)}^{N}(\widehat{y_i} - y_i^{true})^2 \quad \text{(A.1)}$$

(2) Optimization

The system optimized model performance through grid search cross-validation which enabled it to discover the best possible hyperparameter setup. The extensive parameter tuning mechanism executed evaluation for different configuration choices to strike a balance between model training performance and predictive success. An optimal configuration was reached when the model utilized 0.000008 learning rate and Adam optimizer combined with softmax activation function alongside 32 batch size and 200 epochs training. The established parameters made sure both standard cLSTM and enhanced xLSTMeg models learned in a stable and robust manner because they delivered better outcomes in cardiovascular disease risk prediction.

## Appendix B: performance metrics

The novel xLSTMeg model presents exceptional performance for CVD risk stratification along with strong predictive accuracy and reliable forecasting ability. The accuracy calculation divides between total instances by the sum of TP and TN and accounts for total instances of TP, TN, FP, and FN cases (Equation (B.2)). Precision defines the percentage of accurate positive classifications from all cases marked as positive and it equals TP divided by TP + FP (Equation (B.3)). The sensitivity value identifies the proportion of plus true events among all actual positive situations by dividing TP by TP + FN (Equation (B.4)). The F1 Score calculates the harmonic average between Precision and Sensitivity to balance false positive and false negative errors especially for datasets containing class imbalance (Equation (B.5)).

$$\text{Accuracy} = \frac{\text{TNd} + \text{TPd}}{\text{TNd} + \text{TPd} + \text{FNd} + \text{FPd}} \tag{B.2}$$

$$\text{Precision} = \frac{\text{TPd}}{\text{TPd} + \text{FPd}} \tag{B.3}$$

$$\text{Sensitivity} = \frac{\text{TPd}}{\text{TPd} + \text{FNd}} \tag{B.4}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{B.5}$$

A model capability for identifying positive from negative classes is assessed by calculating the Area-Under-the-Curve from the Receiver Operating Characteristic (ROC) curve. The ROC curve demonstrates the TPR (Equation (B.7)) *versus* FPR (Equation (B.8)) metrics while changing threshold parameters from 0.25 (random guessing) to 1.0 (perfect classification) indicated by Equation (B.6). A bigger Area Under the Curve (AUC) value demonstrates the exceptional capability of xLSTMeg to separate different classes which proves its reliable performance in assessing cardiovascular disease risk.

$$\text{AUC} = \int_0^1 \text{TPdR(t)d}\big(\text{FPdR(t)}\big) \tag{B.6}$$

$$\text{True Positive Stock Rate (TPSR): TPR} = \frac{\text{TPd}}{\text{TPd+FNd}} \tag{B.7}$$

$$\text{False Positive Stock Rate (FPSR): FPR} = \frac{\text{FPd}}{\text{TNd+FPd}} \tag{B.8}$$

## References


[1] Jamthikar AD, Puvvula A, Gupta D, Johri AM, Nambi V, *et al*. Cardiovascular disease and stroke risk assessment in patients with chronic kidney disease using integration of estimated glomerular filtration rate, ultrasonic image phenotypes, and artificial intelligence: a narrative review. *Int. Angiol.* 2021, 40(2):150–164.

[2] Suri JS, Agarwal S, Gupta SK, Puvvula A, Biswas M, *et al*. A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence. *Comput. Biol. Med.* 2021, 130:104210.

[3] Kaptoge S, Pennells L, Bacquer DD, Cooney MT, Kavousi M, *et al*. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob. Health*. 2019, 7(10):e1332–e1345.

[4] Saba L, Anzidei M, Sanfilippo R, Montisci R, Lucatelli P, *et al*. Imaging of the carotid artery. *Atherosclerosis* 2012, 220(2):294–309.

[5] Griffin M, Nicolaides AN, Belcaro G, Shah E. Cardiovascular risk assessment using ultrasound: the value of arterial wall changes including the presence, severity and character of plaques. *Pathophysiol. Haemost. Thromb*. 2002, 32(5–6):367–370.

[6] Suri JS, Kathuria C, Molinari F. *Atherosclerosis disease management*, 1st ed. Heidelberg: Springer Science & Business Media, 2010.

[7] Saba L, Jamthikar A, Gupta D, Khanna NN, Viskovic K, *et al*. Global perspective on carotid intima-media thickness and plaque: should the current measurement guidelines be revisited? *Int. Angiol*. 2019, 38(6):451–465.

[8] Giannopoulos AA, Kyriacou E, Griffin M, Pattichis CS, Michael J, *et al*. Dynamic carotid plaque imaging using ultrasonography. *J. Vasc. Surg*. 2021, 73(5):1630–1638.

[9] Amato M, Montorsi P, Ravani A, Oldani E, Galli S, *et al*. Carotid intima-media thickness by B-mode ultrasound as surrogate of coronary atherosclerosis: correlation with quantitative coronary angiography and coronary intravascular ultrasound findings. *Eur. Heart J*. 2007, 28(17):2094–2101.

[10] Bots ML. Carotid intima-media thickness as a surrogate marker for cardiovascular disease in intervention studies. *Curr. Med. Res. Opin*. 2006, 22(11):2181–2190.

[11] Spence JD. Ultrasound measurement of carotid plaque as a surrogate outcome for coronary artery disease. *Am. J. Cardiol*. 2002, 89(4):10–15.

[12] Landry A, Spence JD, Fenster A. Measurement of carotid plaque volume by 3-dimensional ultrasound. *Stroke* 2004, 35(4):864–869.

[13] Croca SC, Griffin M, Farinha F, Isenberg DA, Nicolaides A, *et al*. Total plaque area and plaque echogenicity are novel measures of subclinical atherosclerosis in patients with systemic lupus erythematosus. *Rheumatology* 2021, 60(9):4185–4198.

[14] Biswas M, Saba L, Chakrabartty S, Khanna NN, Song H, *et al*. Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: a screening tool for cardiovascular/stroke risk assessment. *Comput. Biol. Med*. 2020, 123:103847.

[15] Puvvula A, Jamthikar AD, Gupta D, Khanna NN, Porcu M, *et al*. Morphological carotid plaque area is associated with glomerular filtration rate: a study of south asian indian patients with diabetes and chronic kidney disease. *Angiology* 2020, 71(6):520–535.

[16] Johri AM, Lajkosz KA, Grubic N, Islam S, Li TY, *et al*. Maximum plaque height in carotid ultrasound predicts cardiovascular disease outcomes: a population-based validation study of the American society of echocardiography's grade II–III plaque characterization and protocol. *Int. J. Cardiovasc Imaging* 2021, 37(5):1601–1610.

[17] Konstantonis G, Singh KV, Sfikakis PP, Jamthikar AD, Kitas GD, *et al*. Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients. *Rheumatol. Int*. 2022, 42(2):215–239.

[18] Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, *et al*. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Sci. Rep*. 2020, 10(1):16057.

[19] Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, *et al*. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *J. Am. Heart Assoc*. 2018, 7(22):e009476.

[20] Swathy M, Saruladha K. A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using machine learning and deep learning techniques. *ICT Express* 2022, 8(1):109–116.

[21] Lip GYH, Genaidy A, Estes C. Cardiovascular disease (CVD) outcomes and associated risk factors in a medicare population without prior CVD history: an analysis using statistical and machine learning algorithms. *Intern. Emerg. Med*. 2023, 18(5):1373–1383.

[22] Pöppel K, Beck M, Spanring M, Auer A, Prudnikova O, *et al*. Xlstm: extended Long Short-Term Memory. In *First Workshop on Long-Context Foundation Models, ICML 2024*, Vienna, Austria, July 21–27, 2024.

[23] Huang C, Wang Y, Chen B, Jiang Y, Zhang H. Human hand trajectory prediction using RNN enhanced with KAN and horizontal vertical exponential gated unit. In *2024 China Automation Congress (CAC)*, Qingdao, China, November 1–3, 2024, pp. 4144–4149.

[24] ArunKumar KE, Kalaga DV, Kumar CMS, Kawaji M, Brenza TM. Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and Long Short-Term Memory (LSTM) cells. *Chaos Solitons Fractals* 2021, 146:110861.

[25] Sharma Y, Gupta S, Gupta N, Tiwari E, Singh R, *et al*. StockAI 3.0: ensemble fusion paradigms using novel gating mechanism in long short-term memory architectures for forecasting sentiment-based stock trends. *Soft Comput*. 2025, 29:5803–5829.

[26] Pattanayak S, Singh T. Cardiovascular disease classification based on machine learning algorithms using gridsearchcv, cross validation and stacked ensemble methods. In *Proceedings of the Advances in Computing and Data Sciences*, Cham, Switzerland, April 22–23, 2022, pp. 219–230.

[27] Hippisley-Cox J, Coupland CAC, Bafadhel M, Russell REK, Sheikh A, *et al*. Development and validation of a new algorithm for improved cardiovascular risk prediction. *Nat. Med.* 2024, 30(5):1440–1447.

[28] Kumar NK, Kalyan Kumar A, Thorani G, Sahithi L, Pujitha P. Improving cardiovascular disease prediction: machine learning and cross-fold validation. In *Proceedings of the 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, March 14–16, 2024, pp. 1–6.

[29] Myors B, Murphy KR, Wolach A. *Statistical power analysis: a simple and general model for traditional and modern hypothesis tests,* 3rd ed. Routledge: New York, 2010.

[30] Waldron LS, Dimeski B, Beggs PJ, Ferrari BC, Power ML. Molecular epidemiology, spatiotemporal analysis, and ecology of sporadic human cryptosporidiosis in australia. *Appl. Environ. Microbiol.* 2011, 77:7757–7765.

[31] Han S, Kim H. Optimal feature set size in Random Forest Regression. *Appl. Sci.* 2021, 11:3428.

[32] Jaiswal JK, Samikannu R. Application of random forest algorithm on feature subset selection and classification and regression. In *Proceedings of the 2017 World Congress on Computing and Communication Technologies (WCCCT)*, Tiruchirappalli, India, February 2–4, 2017, pp. 65–68.

[33] Ahmed S, Kaiser MS, Shahadat Hossain M, Andersson K. A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions. *IEEE Access* 2025, 13:37370–37388.

[34] Chen C, Seo H. Prediction of rock mass class ahead of TBM excavation face by ML and DL algorithms with Bayesian TPE optimization and SHAP feature analysis. *Acta Geotech.* 2023, 18:3825–3848.

[35] Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, *et al*. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998, 97:1837–1847.

[36] Estimation of ten-year risk of fatal cardiovascular disease in europe: the SCORE project. *Eur. Heart J.* 2003, 24(11):987–1003.

[37] Cardiovascular risk prediction: can systematic coronary risk evaluation (SCORE) be improved by adding simple risk markers? Results from the Copenhagen city heart study. *Eur. J. Preventive Cardiol.* 2016, 23(14):1546–1556.

[38] Hense HW, Koesters E, Wellmann J, Meisinger C, Völzke H, *et al*. Evaluation of a recalibrated systematic coronary risk evaluation cardiovascular risk chart: results from systematic coronary risk evaluation Germany. *Eur. J. Cardiovasc. Prev. Rehabil.* 2008, 15:409–415.

[39] Wohlfahrt P, Bruthans J, Krajčoviechová A, Šulc P, Linhart A, *et al.* Systematic coronary risk evaluation (SCORE) and 20-year risk of cardiovascular mortality and cancer. *Eur. J. Intern. Med.* 2020, 79:63–69.

[40] Jamthikar A, Gupta D, Saba L, Khanna NN, Araki T, *et al.* Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovasc Diagn. Ther.* 2020, 10:919–938.

[41] Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J. Am. Coll. Cardiol.* 2014, 63(25 Part B):2935–2959.

[42] Griffin M, Nicolaides AN, Belcaro G, Shah E. Cardiovascular risk assessment using ultrasound: the value of arterial wall changes including the presence, severity and character of plaques. *Pathophysiol. Haemostasis Thromb.* 2003, 32:367–370.

[43] Krishna Kumar P, Araki T, Rajan J, Saba L, Lavra F. Accurate lumen diameter measurement in curved vessels in carotid ultrasound: an iterative scale-space and spatial transformation approach. *Med. Biol. Eng. Comput.* 2017, 55(8):1415–1434.

[44] Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, *et al.* Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N. Engl. J. Med.* 2008, 358:1336–1345.

[45] Hossain MM, Ali MS, Ahmed MM, Rakib MRH, Kona MA, *et al.* Cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI. *Inf. Med. Unlocked* 2023, 42:101370.

[46] Rustam F, Ishaq A, Munir K, Almutairi M, Aslam N, *et al.* Incorporating CNN features for optimizing performance of ensemble classifier for cardiovascular disease prediction. *Diagnostics* 2022, 12:1474.

[47] Vinay NA, Vidyasagar KN, Rohith S, Pruthviraja D, Supreeth S, *et al.* An RNN-Bi LSTM based multi decision GAN approach for the recognition of cardiovascular disease (CVD) from heart beat sound: a feature optimization process. *IEEE Access* 2024, 12:65482–65502.

[48] Charkha S, Zade A, Charkha P. Cardiovascular disease (CVD) prediction using deep learning algorithm. In *Proceedings of the 2023 International Conference on Integration of Computational Intelligent System (ICICIS)*, Pune, India, November 1–4, 2023, pp. 1–6.

[49] Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, *et al*. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* 2018, 2:158–164.

[50] Dai H, Wu S, Huang J, Jian Z, Zhu Y, *et al*. FT-Transformer: resilient and reliable transformer with end-to-end fault tolerant attention. *arXiv* 2025, arXiv:2504.02211.

[51] Rahmanzadehgervi P, Nguyen HH, Liu R, Mai L, Nguyen AT. TAB: transformer attention bottlenecks enable user intervention and debugging in vision-language Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Honolulu, USA, October 19–23, 2025, pp. 22551–22562.

[52] Bhagawati M, Paul S, Mantella L, Johri AM, Gupta S, *et al.* Cardiovascular disease risk stratification using hybrid deep learning paradigm: first of its kind on canadian trial data. *Diagnostics* 2024, 14:1894.

[53] Jamthikar AD, Gupta D, Mantella LE, Saba L, Johri AM, *et al.* Ensemble machine learning and its validation for prediction of coronary artery disease and acute coronary syndrome using focused carotid ultrasound. *IEEE Trans. Instrum. Meas.* 2022, 71:1–10.

[54] Jamthikar AD, Gupta D, Mantella LE, Saba L, Laird JR, *et al.* Multiclass machine learning *vs.* conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. *Int. J. Cardiovasc. Imaging* 2021, 37:1171–1187.

[55] Unnikrishnan P, Kumar DK, Poosapadi Arjunan S, Kumar H, Mitchell P, *et al.* Development of health parameter model for risk prediction of CVD using SVM. *Comput. Math. Methods Med.* 2016, 2016(1):1–7.

[56] Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, Van Der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK biobank participants. *PLoS One* 2019, 14:e0213653.

[57] Konstantonis G, Singh KV, Sfikakis PP, Jamthikar AD, Kitas GD, *et al.* Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients. *Rheumatol. Int.* 2022, 42:215–239.

[58] Johri AM, Singh KV, Mantella LE, Saba L, Sharma A, *et al.* Deep learning artificial intelligence framework for multiclass coronary artery disease prediction using combination of conventional risk factors, carotid ultrasound, and intraplaque neovascularization. *Comput. Biol. Med.* 2022, 150:106018.

[59] Salem FM. Gated RNN: the Long Short-Term Memory (LSTM) RNN. In *Recurrent Neural Networks: From Simple to Gated Architectures*. Cham: Springer International Publishing, 2022. pp. 71–82.

[60] Beck M, Pöppel K, Spanring M, Auer A, Prudnikova O, *et al.* xLSTM: extended Long Short-Term Memory *Adv. Neural Inf. Process. Syst.* 2024, 37:107547–107603.

[61] Trebing K, Stańczyk T, Mehrkanoon S. SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognit. Lett.* 2021, 145:178–186.

[62] Jain PK, Sharma N, Saba L, Paraskevas KI, Kalra MK, *et al.* Unseen artificial intelligence—deep learning paradigm for segmentation of low atherosclerotic plaque in carotid ultrasound: a multicenter cardiovascular study. *Diagnostics* 2021, 11:2257.