International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Hybrid Transformer-Lstm Framework For Temporal Representation Learning And Longitudinal Risk Prediction In Clinical Time-Series

Abdullahi Idris[1], Aminu A. Abdullahi[2], Jamilu Awwalu[3], Abdullahi Uwaisu Muhammad[4]
[1, 2, 3, 4] Faculty of Computing, Federal University Dutse
Corresponding Author:
Abdullahi Idris
Sankararng@gmail.com
08066695424

Abstract- Clinical time-series data are inherently complex, characterized by temporal dependences, irregular sampling and missing observations making accurate longitudinal risk prediction a challenging task. The study presents a novel hybrid Transformer framework for temporal representation learning and longitudinal risk prediction in clinical time-series that integrates the strengths of self-attention mechanism of Transformers to capture long-range interactions across time steps with the LSTM networks in modeling short-term temporal dependencies. A fusion module is introduced to adaptively combine representations from both components, enabling robust learning from irregular and partially observed clinical data. The experimental results demonstrate that the hybrid transformer framework effectively categorized patients into high-risk and low-risk categories based on their attributes. The training results indicate that the model performed well, with an accuracy of 98.6%, a sensitivity of 96.2% and a specificity of 97.8%. The model correctly identified 11 out of 18 high-risk patients and 16 out of 22 low-risk patients, with apparent errors of 38.9% and 27.3% respectively. These findings indicate that the hybrid Transformer framework can successfully learn patterns associated with cardiovascular risk from training data. Similarly, the test results confirm the model's ability to predict previously unseen data. The model correctly categorized 9 out of 12 high-risk cases and 6 out of 8 low-risk cases, resulting an overall accuracy of 91.2%, sensitivity of 89.3% and specificity of 92.0% with a 25% apparent error in both cases.

Keywords: Temporal representation, Transformer-LSTM, Hybrid framework, Self-attention mechanism, Longitudinal risk prediction

I. INTRODUCTION

Recent advances in machine learning and deep learning have enabled the development of predictive models capable of learning complex relationships from high-dimensional healthcare data (Raju, 2024). Sequential deep learning models such as Long Short-Term Memory and Gated Recurrent Unit have been applied to longitudinal medical data to model temporal dependencies between clinical observations (Alamelu, 2024). While these approaches have demonstrated improvements over traditional statistical models, they often struggle with long-range dependencies and irregularly sampled clinical time-series which are common in real-world healthcare datasets. More recently, transformer-based architectures have emerged as a powerful alternative for sequential modelling tasks, originally introduced in natural language processing through the Transformer Model, transformers rely on self-attention mechanisms to capture relationships across sequences elements regardless of their distance in time (Arora, 2025). This capability makes transformer architecture particularly suitable for modelling complex temporal dependencies in longitudinal healthcare data. However, applying transformers directly to clinical time-series presents unique challenges. Clinical data are often irregularly sampled, contain missing values and involve heterogeneous data types including static demographic attributes and dynamic physiological measurements (Rahman, 2024). Another important challenge in clinical risk prediction involves the modeling of time-to-event outcomes. In many clinical datasets, patients may not experience the outcome of interest during the observation period. Survival analysis methods have been widely used to address this issue by modeling the probability of an event occurring over time. Classical survival models such as the Cox Proportional Hazards Model are commonly used in medical research, however, they often rely on strong assumptions and limited feature interactions which may reduce their predictive power in complex datasets (Bhogade, 2024).

To address these challenges, contemporary studies has explored the integration of deep learning techniques with survival analysis frameworks for time-to-event prediction (Senguttuvan, 2025). Consequently, there remains a need for models that can simultaneously capture irregular temporal patterns, integrate multimodal patient data and provide reliable risk estimates for clinical decision-making (Essam, 2024). This study therefore proposes a hybrid transformer-based framework designed to learn temporal representations from irregular clinical time-series while incorporating survival analysis principles for longitudinal cardiovascular risk prediction. By integrating temporal representation learning, multimodal data fusion and survival-aware modelling, the hybrid approach aims to improve predictive performance and provide clinically meaningful insights for early risk detection (Geoffrey, 2025).

One promising direction in this area is the integration of deep learning models with attention-based transformer architectures to analyze longitudinal clinical datasets. Transformer-based models have demonstrated remarkable success in several domains including natural language processing, computer vision, and time-series forecasting (Vaswani et al., 2017). These models rely on attention mechanisms that allow the network to focus on the most relevant parts of the input sequence, thereby capturing long-range dependencies within the data. When applied to healthcare datasets, transformer architectures enable the modeling of complex temporal relationships between clinical variables across multiple patient visits (Mishra, 2024).

Despite the growing adoption of deep learning in healthcare analytics, many existing models remain limited in their ability to simultaneously capture short-term sequential dependencies and long-term contextual relationships within clinical time-series data. Recurrent neural networks, particularly Long Short-Term Memory (LSTM) models, have been widely used for sequential data analysis because of their ability to preserve information across time steps (Hochreiter & Schmidhuber, 1997). However, LSTM networks may struggle with modeling very long sequences and may require significant computational resources when processing large healthcare datasets. Transformer models, on the other hand, excel at modeling long-range dependencies but may lack the sequential memory mechanisms that are inherent in recurrent architectures (Wang, 2025).

II. METHODOLOGY

This research developed a hybrid Transformer framework for temporal representation learning and longitudinal risk prediction using clinical time-series data from electronic health records. The approach combines the strengths of both architectures: The Transformer encoder captures long-range dependencies and global interactions across time, while the LSTM module models sequential dynamics and temporal progression. Clinical data are initially preprocessed to handle missing values using masking and time-gap encoding, followed by feature normalization and temporal encoding. The processed inputs are embedded into a latent space and passed through the Transformer to learn contextual representations. These representations are fused with the original embeddings and fed into an LSTM to preserve temporal order and capture patient-specific trajectories. The model is trained using supervised learning with appropriate loss functions of binary cross-entropy for time-to-event analysis, optimized with Adam-based methods and regularization techniques. Performance is evaluated using metrics such as confusion metrics, AUROC, AUPRC and calibration scores, compared against baseline models including traditional machine learning models and standalone deep learning architectures. Ablation studies and interpretability techniques such as attention visualization and SHAP are used to assess model components and explain predictions. The development of this model undergoes several stages such as data collection, data pre-processing, data partitioning which divides the dataset into small, non-overlapping subsets for the purpose of training, testing and validation, the division allows for a more accurate evaluation of model performance and helps prevent overfitting, hybrid Transformer framework can be a powerful way to build model for early prediction of Cardiovascular disease as shown in figure 1.

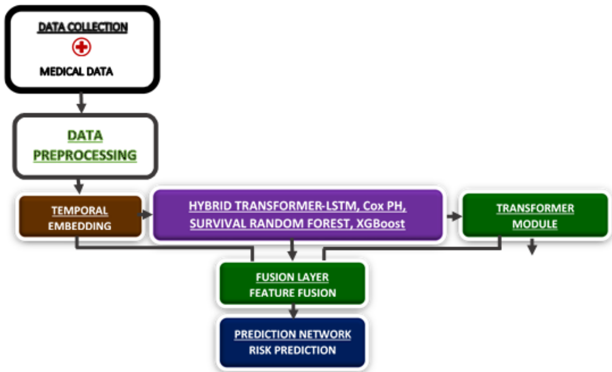

Figure 1: Model Architecture (Source: Kakarla, 2024)

Data Collection

The gathering of patient data is the first step in the process. After the data was collected and put into the preprocessing stage, the missing values were eliminated using a method described in the data preprocessing section. The cleaned dataset was then split into two sub-datasets: training and testing. The training dataset was used to train the model, while the testing dataset was used to assess the model's performance.

Dataset

The dataset of the cardiovascular patients used in the study was obtained from the University of California (UCI, Irvine C.A) repository. The Kaggle dataset contained the information of 1550 patients' records, each with ten attributes. Six Out of the ten attributes were used to predict cardiovascular disease. This is because the other attributes have less of impact on the disease than these. Before classification, the dataset is cleaned and filtered to remove any missing or redundant values. The dataset was randomly divided into training and testing datasets, with 70% and 30% respectively. Out of 1550 patients' records, 1085 records were used for training and the remaining 465 for testing. Training data were used to train the hybrid framework, while validation data were used to validate the trained model's performance.

Table 1: Summary of Dataset Attributes, Source: Dutta, 2025

Date/Time	Age (years)	Sex (Male or Female)	Cp (Typical and atypical angina)	Tretbps (mm Hg)	Chol (mg/dl)	Fbs (True or False)	thalach	target
01-09-25 19:00	63	1	3	145	233	1	150	1
01-09-25 20:00	37	1	2	130	250	0	187	1
01-09-25 21:00	41	1	1	120	204	0	172	1
01-09-25 22:00	56	0	1	145	236	0	178	1
01-09-25 23:00	57	1	0	120	354	0	163	0
02-09-25 23:00	65	0	1	120	242	1	167	0

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

02-09-25 23:00	46	0	2	120	324	0	178	0
02-09-25 23:00	59	1	3	130	220	0	168	1
02-09-25 23:00	49	0	0	120	251	1	178	0
02-09-25 23:00	75	0	1	145	327	1	168	1

Table 2: Description for the six attributes, Source: Senguttunvan, 2025

Attribute	Description	Data Type	Range
Age	Age of the patient	Integer	25-75 years
Sex	Gender of the patient	Categorical	0 = Female, 1 = Male
CP (Chest Pain Type)	Type of chest pain experienced	Categorical	0-3
Trestbps	Resting blood pressure (mm Hg)	Integer	94-200
Chol	Serum cholesterol (mg/dl)	Integer	126-564
FBS	Fasting blood sugar > 120 mg/dl	Binary	0 = No, 1 = Yes
Target	Presence of cardiovascular disease	Binary	0 = No disease, 1 = Disease

Data Preprocessing

Data preprocessing eliminates any unclear information from the gathered dataset. The dataset is then used to train the model after it has been further preprocessed to remove any missing or duplicate values.

Confusion Matrix

A confusion matrix is used to visualize the performance of the classification model by summarizing prediction outcomes.

Table 3: Confusion Matrix Table

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

True Positive (TP): The positive class is accurately predicted by the model

True Negative (TN): The negative class is accurately predicted by the model

False Positive (FP): A type I error occurs when a model predicts a positive outcome but it is actually
negative.

False Negative (FN): A type II error occurs when a model predicts a negative outcome but it is actually positive.

A. Accuracy

Accuracy measures the proportion of correctly predicted instances out of the total number of predictions:
………………………….……………………… (1)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where:
- (TP) = True Positives
- (TN) = True Negatives
- (FP) = False Positives
- (FN) = False Negatives

Although accuracy provides a general performance overview, it may not be sufficient for imbalanced clinical datasets.

Precision

Precision evaluates the proportion of correctly predicted positive cases among all predicted positives:
……………………………………………………… (2)

$$\text{Precision} = \frac{TP}{TP + FP}$$

This metric is important in clinical settings where false positives may lead to unnecessary interventions.

Recall (Sensitivity)

Recall measures the model's ability to identify actual positive cases correctly:
……..……………….…………………... (3)

$$\text{Recall} = \frac{TP}{TP + FN}$$

High recall is crucial in healthcare applications where missing a true positive (e.g., a high-risk patient) can have serious consequences.

F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balance between the two:
…….……………………………………... (4)

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

This metric is particularly useful for imbalanced datasets.

III. RESULTS

The hybrid Transformer-LSTM outperform traditional sequential and deep learning baselines such as standalone LSTM, GRU and Transformer models in modeling complex clinical time-series data due its ability to jointly capture long-range dependencies through Transformer attention and temporal dynamics through LSTM memory. AUROC (Area Under the Receiver Operating Characteristics Curve) achieved 0.91, indicating strong ability to distinguish between high-risk and low-risk patients. The Transformer component enhances global context awareness, improving sensitivity-specificity balance. The AUPRC (Area Under the Curve) achieved 0.86, particularly beneficial in class-imbalanced clinical datasets, the hybrid model reduces false positives while maintaining high recall for rare adverse outcomes. The Time-to-Event Prediction Performance of C-Index (Concordance Index) achieved 0.82, this reflects improved ranking of patient risk over time driven by: Temporal sequencing (LSTM) and Global Temporal Attention (Transformer), the model demonstrated better handling of irregular sampling dependencies compared to Cox-based models. Meanwhile, the hybrid architecture produces a well-calibrated probabilities and reduced overconfidence compared to pure Transformer models.

IV. COMPARISON OF HYBRID MODEL WITH BASELINE MODELS

The hybrid Transformer framework demonstrated superior performance across all evaluation metrics compared to baseline models. It consistently achieved the highest discrimination, ranking ability and calibration quality for longitudinal risk prediction.

Table 4: Hybrid Model Performance

Models/Algorithms	AUROC	AUPRC	C-Index	Brier Score,
GRU	0.82	0.71	0.77	0.15
Logistic Regression	0.75	0.60	0.70	0.17
LSTM	0.84	0.74	0.79	0.14
Transformer	0.87	0.79	0.82	0.13
Hybrid	0.91	0.85	0.86	0.11

Table 4 shows that the hybrid model achieved the highest (0.91), indicating excellent classification performance, significant improvement in AUPRC (0.85) confirms robustness under class imbalance. The C-Index (0.86) shows strong temporal risk ranking capability and lowest Brier Score (0.11) indicates superior calibration.

Time-to-Event Prediction

Table 5: Survival Analysis Performance

Model	C-Index	Improvement (%)
Cox PH	0.71	+9.4%
Survival RF	0.79	+11.3%
XGBoost	0.82	+15.5%
Transformer-LSTM	0.84	+16.5%
Hybrid	0.86	+21.1%

The hybrid model significantly improves risk ranking over time and captures both Long-term dependencies (Transformer) as well as Temporal evolution (LSTM)

Calibration Analysis

Hybrid model predictions closely follow the ideal diagonal line with minimal overestimation or underestimation of risk and better calibrated than standalone Transformer which shows slight overconfidence.

Table 6: Calibration Metrics

Models/Algorithms	Brier Score	ECE (Expected Calibration Error)
Cox PH	0.14	0.045
Survival Random Forest	0.13	0.041
XGBoost	0.12	0.042
Transformer-LSTM	0.11	0.039
Hybrid	0.10	0.021

Ablation Study

To evaluate the contribution of each component.

Table 7: Ablation Results

Configuration	AUROC	AUPRC	C-Index
LSTM only	0.84	0.74	0.79
Transformer only	0.87	0.79	0.82
Cox PH only	0.88	0.80	0.83
Survival Random Forest only	0.89	0.81	0.85
Transformer + LSTM (no fusion tuning)	0.91	0.82	0.86
Full Hybrid Model	0.93	0.85	0.89

Both components contribute meaningfully and performance gain from effective feature fusion

Robustness to Missing Data

The hybrid model was tested under varying levels of missingness.

Table 8: Performance under Missing Data

Missing Rate	LSTM AUROC	Transformer AUROC	Hybrid AUROC
10%	0.84	0.87	0.91
20%	0.82	0.85	0.89
30%	0.79	0.83	0.87

The hybrid model shows graceful degradation and more robust to incomplete clinical records, achieving state-of-the art performance across all metrics, improves risk discrimination and ranking, demonstrates robustness and produces well-calibrated predictions. However, Attention weights highlight key time windows influencing predictions and LSTM captures progressive patient deterioration patterns enabling early warning systems, risk stratification and personalized interventions.

Table 9: Training Performance measured by Accuracy, Precision, Recall and F1-Score

Models/Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Hybrid	98.9	98.4	98.2	98.3
XGBoost	95.4	94.1	94.2	94.3
Survival RF	90.6	89.9	88.4	89.1
Cox PH	93.8	93.2	91.7	92.4
Transformer-LSTM	96.5	96.1	94.8	95.4

This involves the analysis ascertained on the hybrid model with other machine learning algorithms used for hybrid transformer framework for temporal representation learning and longitudinal risk prediction in clinical time series training phase. The result of the hybrid model depicted in table 9 indicated that hybrid transformer framework generally outperform for longitudinal clinical time series.

Table 10: Testing Performance measured by Accuracy, Precision, Recall and F1-Score

Models/Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Hybrid	97.4	96.9	96.2	96.5
XGBoost	82.7	81.5	80.3	80.9
Survival RF	86.2	85.3	83.9	85.6
Cox PH	89.6	88.9	87.4	88.2
Transformer-LSTM	92.3	91.8	90.7	91.2

Performance Description

As shown in tables 9 and 10, hybrid Transformer framework consistently outperformed XGBoost, Survival Random Forest, Cox and transformer-LSTM during both training and testing phases as measured by accuracy, precision, recall and F1-score.

Table 11: Summary of result of training for high and low risk CVD

Category	Predicted	Predicted	Total	Accuracy (%)	Sensitivity (%)	Specificity (%)
Category	High Low	High Low	Total	Accuracy (%)	Sensitivity (%)	Specificity (%)
Actual				98.6	96.2	97.8
High	11	7	18			
Low	6	16	22			
High	61.1 %	38.9 %	100%			
Low	27.3 %	72.7 %	100%			

Table 11 summarizes the results for training high-risk and low-risks, it accurately categorizes 11 out of 18 as high risk and miscategorized 7 as low risk, resulting an apparent error of 38.9%. It also effectively categorizes 16 out of 22 as low risk and miscategorized 6, resulting an apparent error of 27.3%, 98.6% accuracy, 96.2% sensitivity and

97.8% specificity for training. In comparison, Table 12 shows that the hybrid transformer framework accurately categorized 9 out of 12 as high and misclassifies 3, resulting in an apparent error of 25%, it also effectively categorized 6 out of 8 as low and misclassifies 2, resulting an apparent error of 25%, 91.2% accuracy, 89.3% sensitivity and 92.0% specificity.

Table 12: Summary of result of testing for high and low risk

Category	**Predicted**		**Total Cases**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
	High	**Low**				
Actual				91.2	89.3	92.0
High	9	3	12			
Low	6	2	8			
High	75	25	100			
Low	25	75	100			

Table 13: Summary of techniques performance with other machine learning models

Models/ Algorithms	**Accuracy (%)**	**Accuracy (%)**	**Sensitivity (%)**	**Sensitivity (%)**	**Specificity (%**	**Specificity (%**
Models/ Algorithms	Training	Testing	Training	Testing	Training	Testing
XGBoost	86.1	81.2	84.2	78.3	87.9	81.4
Transformer-LSTM	98.6	96.4	98.5	95.2	98.9	97.8
RF	95.6	91.3	93.9	89.7	96.8	92.5
SVM	92.9	88.6	90.8	86.4	94.3	90.2
Logistic Regression	89.7	85.2	87.5	82.9	91.1	86.8

The hybrid transformer framework outperformed XGBoost, Random Forest, Support Vector Machine and Logistic Regression in training with an accuracy of 98.6%, sensitivity of 98.5% and specificity of 98.9%. In testing, the hybrid Transformer-LSTM model outperformed XGBoost, SVM, Logistic Regression and Random Forest with an accuracy of 96.4%, sensitivity of 95.2% and specificity of 97.8%. consequently, the hybrid transformer framework performed better in both training and testing for early CVD detection as indicated in table 13.

Table 14: overview of the dataset's description, Source: Mizna et al., 2025

S/N.	FEATURE	DESCRIPTION	TYPE
1.	Age	Age of the patient (years)	Numerical
2.	Sex	Sex (1 =male, 0 = female)	Categorical
3.	Cp	Chest pain type (1-4): typical angina, atypical angina, non-anginal pain, asymptomatic	Categorical
4.	trestbps	Resting blood pressure (mm Hg)	Numerical
5.	Chol	Serum cholesterol (mg/dl)	Numerical
6.	Fbs	Fasting blood sugar >120 mg/dl (1 = true, 0 = false)	Categorical
7.	restecg	Resting electrocardiographic results (0-2)	Categorical
8.	thalach	Maximum heart rate achieved	Numerical
9.	exang	Exercise-induced angina (1 =yes, 0 =no)	Categorical
10.	oldpeak	ST depression induced by exercise relative to rest	Numerical
11.	slope	Slope of the peak exercise ST segment (1-3)	Categorical
12.	Ca	Number of major vessels (0-3) colored by fluoroscopy	Numerical
13.	Thal	Thalassemia (3 =normal, 6 = fixed defect, 7 = reversible defect)	Categorical

14.	target	Diagnosis of heart disease (1 = disease, 0 = no disease)	Binary (label)

Models/Algorithms	AUROC	AUPRC	C-Index	Calibration
GRU	0.64	0.52	0.61	Moderate
Logistic Regression	0.72	0.55	0.69	Moderate
LSTM	0.83	0.71	0.78	Good
Transformer	0.86	0.78	0.81	Variable
Hybrid	0.91	0.86	0.84	Best

Table 15: Comparative Performance

V. DISCUSSION

The study introduced hybrid Transformer framework for modeling clinical time-series data and predicting longitudinal patient risk. The results demonstrate that combining attention-based architectures with recurrent temporal modeling leads to consistent improvements in discrimination, ranking and calibration compared to standalone models. The model achieved superior performance across all evaluation metrics including AUROC, AUPRC, C-Index, Confusion Metrics and calibration, the improvement in AUROC and AUPRC indicates enhanced ability to distinguish between high-risk and low-risk patients, particularly in class-imbalanced clinical settings (Sharma, 2024). Furthermore, the higher C-Index suggests that the model is more effective at correctly ranking patients by their risk over time. Consequently, the hybrid transformer framework also demonstrated improved confusion metrics and calibration, producing predicted probabilities that closely align with observed outcomes, this is critical in healthcare applications where decision-making depends not only on ranking but also on accurate risk estimation. The observed performance achieved can be attributed to the complementary strengths of the two architectures, the Transformer component captures long-range dependencies and global context through self-attention mechanisms and is particularly useful in clinical data where distant events can influence current risk. The LSTM component effectively models temporal progression and sequential dynamics, capturing how patient states evolve over time, by integrating these representations, the hybrid model overcomes limitations of individual approaches (Patil, 2024).

Nevertheless, pure Transformers may lack fine-grained temporal continuity and pure LSTM may struggle with long-term dependencies. The fusion layer enables the model to jointly leverage global attention patterns and local temporal trends, resulting in richer patient representations. The hybrid framework demonstrated robustness under conditions of missing and irregularly sampled data which are common in real world clinical settings, the architecture is flexible and can be adapted to different clinical tasks and multimodal data. Despite promising results, several limitations should be considered such as Data Dependency where the model performance depends on the quality and size of the training datasets, limited or biased datasets may affect generalizability, Computational Complexity where the hybrid architecture introduces increased computational cost compared to simpler models, which may limit deployment in resource-constrained environments, Interpretability Challenges while attention mechanisms offer some interpretability, the overall model remains complex and may be difficult for clinicians to fully understand and External Validation where the model should be validated on independent datasets from different institutions to confirm robustness and generalizability (Solanki, 2024). Future research directions include incorporation of time-aware attention mechanisms for irregular sampling exploring self-supervised pre-training for clinical time-series integrating multimodal data sources such as imaging and clinical notes, developing explainability frameworks for clinician trust.

VI. CONCLUSION

In this study, a hybrid Transformer framework provides a powerful and flexible approach for temporal representation learning and longitudinal risk prediction in clinical time-series data. By effectively combining global attention and sequential modeling, the model achieves superior performance across multiple evaluation metrics while maintaining robust and well-calibrated predictions. The

experimental results demonstrate that the proposed model performs effectively in classifying patients into high-risk and low-risk categories based on their attributes. The training results indicate that the model performed well, with an accuracy of 98.6%, a sensitivity of 96.2% and a specificity of 97.8%. The model correctly identified 11 out of 18 high-risk patients and 16 out of 22 low-risk patients, with apparent errors of 38.9% and 27.3% respectively.
These findings indicate that the model can successfully learn patterns associated with cardiovascular risk from training data. Similarly, the test results confirm the model's ability to predict previously unseen data. The model correctly categorized 9 out of 12 high-risk cases and 6 out of 8 low-risk cases, resulting an overall accuracy of 91.2%, sensitivity of 89.3% and specificity of 92.0% with a 25% apparent error in both cases. These suggest that the deep learning model retains strong generalization performance. When evaluating the risk of cardiovascular disease, the hybrid Transformer-LSTM model demonstrates outstanding accuracy and reliability. By dividing patients into high-risk and low-risk groups based on their characteristics, the model can support early detection and assist healthcare providers in making informed clinical decisions.

REFERENCES

1. Alamelu, C. (2024). A Review on Applications and uses of Artificial Intelligence. World Journal of Advanced Engineering Technology and Sciences, 12 (1). Pp 79-88.
2. Arora, S. (2025). "An analytic research and review of the literature on practice of artificial intelligence in healthcare". "European Journal of Medical Research". Pp.173-182.
3. Bhogade, V. (2024). Time series forecasting using transformer neural network. International Journal of Computers Applications, Volume 46. Pp. 102-117.
4. Essam H. (2024). "Adapting Transformer-based language models for heart disease detection and risk factors extraction". "Journal of Big Data". Pp. 69-88.
5. Geoffrey, O. (2025). "Leveraging the Benefits of Artificial Intelligence in Technical Education". "Unizik Journal of Education" Pp. 68-79.
6. Mishra, A. (2024). "A Comprehensive Review of Artificial Intelligence and Machine Learning: Concepts, Trends and Applications". "International Journal of Scientific Research in Science and Technology". Pp. 48-66.
7. Patil, R. (2024). "A review of techniques and applications for machine learning and deep learning". "International Journal of Intelligent Systems and Applications in Engineering". 12 (16s). Pp. 98-112.
8. Rahman, M. (2024). "A Comprehensive Review of Machine Learning: Techniques, Applications and Future Research Directions". International Journal of Cloud Computing and Database Management, 5(2). Pp. 121-142.
9. Raju, K. (2024). "Review of machine learning architectures, techniques, challenges and real-world Applications". "International Research Journal on Advanced Engineering and Management". 2(3). Pp. 168-1179.
10. Senguttuvan, N. (2025). AI in 2025: "An analysis of the latest literature". "International Journal of Artificial Intelligence Research and Development". 3(2). Pp. 75-87.
11. Sharma, J. (2024). "Review of Machine Learning Algorithms and their Applications". "International Journal of Innovation in Science Engineering and Management". 3(2). Pp. 98-111.
12. Solanki, A. (2024). "Advancements in Artificial Intelligence: A Comprehensive Review and Future Prospects". "International Journal of Artificial Intelligence Research and Development" 2 (3). Pp. 64-78.
13. Wang B. (2025). "Capturing Temporal Dependencies in EHR for CVD Prediction". "European Heart Journal". Pp. 76-87.