**Journal of Artificial Intelligence, Machine Learning and Neural Network**
Vol: 06, No. 01, Jan-June 2026
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.61.42.52 

**Research Paper**

# ClinFormer: a multi-modal clinical transformer for explainable major adverse cardiovascular event prediction from electronic health records

**Dr. Ramesh Murlidhar Bhatawdekar***
*Geotropik, Department of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia.

**Article Info**

***Article History:***
Received: 21 November 2026
Revised: 28 January 2026
Accepted: 05 February 2026
Published: 20 February 2026

***Keywords:***
Clinical Transformer
Multi-Modal EHR
Attention Mechanism SHAP
Cross-Modal
Attention Contrastive
Learning

**ABSTRACT**

Background: Major adverse cardiovascular events (MACE), including acute myocardial infarction, stroke, and cardiovascular death, account for over 8 million deaths globally each year. Conventional prediction models such as Framingham Risk Score, SCORE2, and Pooled Cohort Equations rely on limited traditional risk factors and linear assumptions, restricting their ability to capture complex temporal and non-linear relationships within longitudinal electronic health records (EHRs).
Methods: We propose ClinFormer, a multi-modal clinical Transformer designed to integrate five EHR modalities: laboratory results, diagnosis codes, medication records, clinical notes, and vital signs. The model employs cross-modal attention mechanisms with 12 attention heads and a model dimension of 512. ClinFormer was pre-trained using contrastive patient similarity learning on 127,438 patients from MIMIC-IV and externally validated on 38,924 patients from the eICU database. Model interpretability was provided through SHAP analysis and calibrated probability outputs.
Results: On external validation, ClinFormer achieved an AUROC of 0.943 (95% CI: 0.937–0.949), significantly outperforming the strongest baseline model, ClinicalBERT (AUROC: 0.912; $p < 0.001$). Calibration performance was strong with an expected calibration error (ECE) of 0.031. SHAP analysis identified BNP, troponin I, and eGFR as the most influential predictors.
Conclusions: ClinFormer provides accurate, interpretable, and well-calibrated MACE prediction directly from routinely collected EHR data, supporting its potential deployment in both resource-rich and resource-constrained clinical environments.

*Corresponding Author:*
Dr. Ramesh Murlidhar Bhatawdekar
Geotropik, Department of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia.
Email: rmbhatawdekar@gmail.com

## 1. INTRODUCTION

MACE (acute myocardial infarction (AMI), Ischaemic stroke, cardiovascular death) prediction is of paramount importance in preventive cardiology, intensive care unit (ICU) triage and health system resource planning [1]. Although traditional risk scoring models have been developed and validated over the last few decades, they have not performed optimally in external populations of patients: the Framingham Risk Score reached an area under the curve of ~ 0.73-0.78 among external patients [2], SCORE2 reached 0.76-0.81; and Pooled Cohort Equations reached 0.72-0.78 (prospective validation) [3]. However, these restrictions [4], come from two basic methodological bottlenecks, namely the first of which restricts attention to a few traditional risk factors which neglect the rich temporal information available in longitudinal records, and the second of which assumes linear or log-linear relationship between the risk factors and outcomes when such a relationship is not present, to capture the complexities of biological interactions [5]. An electronic health record (EHR) has a much more detailed characterization of each patient than does a traditional risk score, with continuous laboratory trends, changing lists of diagnosis codes, medication dosing, nursing notes, and streams of vital signs collected from wearable devices. However, due to the nature of the problem, where we have a multi-modal, high dimensional, temporally structured data, poses a challenge for the machine learning architectures to give us the predictable signal. Given the above, the Transformer [6] architecture, introduced in 2017 by Vaswani et al. using self-attention to capture patterns from long-range context across sequence positions without sequential limitations of recurrent networks is the natural choice for modelling EHRs.

While converting to EHR data has been addressed in other natural language processing transformation problems, it still encounters new domain-specific issues that do not arise in NLP: missing data due to irregular sampling, the presence of a wide range of token types with different semantic scales, temporal ordering of measurements spanning from minutes to years, and the necessity of calibrated probability outputs that are compatible with clinical decision thresholds [7], [8].

For recording ICD codes, several attempts have been made to create EHR Transformer architectures such as BEHRT [9] for ICD code sequences, Med-BERT [10] for predicting diseases, and STraTS model [11] for multiple time-series vital signs. These architectures handle one modality of processing, however, and are not meant to be built to address all of these informational modalities at once, along with numerous others, in a concurrent, multi-observables, shared attention space in the lab. Besides, all clinical Transformers focus on a set of components that are incomplete from a practical and clinical point of view: (i) contrastive pre-training for patient representation learning, (ii) probabilistic output, (iii) attribution of predicted outputs based on SHAP, and (iv) internal and external testing using prospective-grade temporal holdout. All four requirements are met in a single application by ClinFormer, which is a novel contribution to the clinical AI literature.

Specifically, we introduce: (i) ClinFormer, a novel six-layer approach of cross-modal attention Transformer that jointly processes five EHR data streams in a shared token space with modality specific liner projects and 2D positional embeddings; (ii) a contrastive pre-training objective based on patient similarity that boosts the performance of downstream MACE prediction by 0.021 AUC over random initialisation; (iii) extensive external validation on the eICU-CRD dataset that is geographically and temporally different from the training set; (iv) a post hoc calibration trained via Platt scaling achieving ECE of 0.031; and (v) a clinician-validated SHAP attribution pipeline that generates patient level explanation summaries that agree with known cardiovascular biomarker biology.

The 30-day mortality from AMI is 15–25% and from cardioembolic stroke is 25–30%, in MACE [12]. Time-sensitive interventions, such as initiate DAPT, PNI, start anticoagulation with embolic events, haemodynamic monitoring at ICU are all interventions given to patients with highest risk for MACE, leading

to significant improvements. However, a tool would be available that will provide the continuously updated, calibrated MACE probability, directly from commonly available EHRs without the need for more restrictive specialist biomarker tests, and ready-to-use to any electronic clinical decision support system for any hospital system who has an EHR with structured data export capabilities. ClinFormer is intended for the MACE prediction from index hospitalization in a binary classification setting (with a prediction horizon of 30 days). It takes time-series laboratory values, ICD-10 diagnosis codes, ATC medication codes, clinical note TF-IDF embedding's, as well as vital sign waveform summaries as concurrent inputs, which do not require any additional information beyond what is currently available in the standard EHR. It is novel in comparison to previous EHR Transformers due to its cross-modal attention (simultaneous attention across patients and modalities), the patient similarity contrastive pre-training approach, and the integrated calibration-attribution pipeline performed in a single inference call.

## 2. RELATED WORK

Small number of clinical variables have been used for conventional cardiovascular risk stratification, like Framingham Risk Score or Score2 and Pooled Cohort Equations [13] which are still widely used. But, in external validation they always have an AUC under 0.80, making them of limited value for determining individual high-risk patients in clinical practice. Human knowledge and Artificial Intelligence (AI) have been intermingled in the realm of medicine as described by Topol [5] to set the foundations for using deep learning in clinical risk prediction tasks. The Transformer architecture [6] has been adopted for clinical data in various ways with increasing sophistication, first developed for NLP. Another work that used Transformer-based BERT-style pre-training, BEHRT [9] showed that such an approach could be applied to sequences of ICD codes from EHRs for which they obtained better results than previous studies that utilized RNN-based architectures on such tasks. Structure EHR data with disease-specific pre-training objectives were then added to this framework by Med-BERT [10] to achieve better prediction performance across multiple clinical outcomes. This work focuses on representation learning for clinical data that are sparse and irregularly sampled, such as multivariate time series, and shows its benefit with the STraTS model [11] specifically designed for these data types. Even with these new developments, the previous EHR Transformers always work on a single data modality. None are able to take in laboratory values, diagnosis information, medication orders, clinical note and vital sign streams simultaneously via an integrated attention mechanism.

Moreover, representation learning through contrastive loss for patient similarity in EHRs [13] and post-hoc probability calibration [14] and model interpretability using SHAP [15] have never been used simultaneously in any previous clinical Transformer. ClinFormer fills this need, by providing a single end-to-end trainable pipeline with all four capabilities, validated in two independent large critical care cohorts. This study's statistical methodology is in line with best practices for clinical prediction modelling. The concept of "decision curve analysis" offers a principled approach to comparing AUROCs from correlated classifiers across a spectrum of decision thresholds, while DeLong's method [16] allows for strict comparisons of AUROCs between correlated classifiers. The advantages of CL principles inspired by Computer Vision (CV) was well established in the context of clinical Representation learning, as observed in the present study.

## 3. METHODOLOGY

### 3.1 Data Sources and Cohort Definition

The development cohort consisted of patients who were identified from the MIMIC-IV database (version 2.2), which is a single centre critical care database from Beth Israel Deaconess Medical Center, Boston, Massachusetts [17]. Every adult patient (age ≥ 18 years) who is hospitalized or admitted to the hospital's intensive care unit (ICU) between 2008 and 2022 with at least 12 hours of EHR data and a known outcome at 30 days with a defined MACE was eligible for inclusion. The end point of MACE was defined as in-hospital AMI (ICD-10 code I21.x, I22.x) or ischaemic stroke (ICD-10 code I63.x) or cardiovascular death

(recorded cause of death with ICD-10 code primary I00–I99). Excluding those aged <18 years (n = 1,234), with <5 laboratory measurements (n = 4,312) and duplicate hospitalisations (with only the first retained), a total of 127,438 patients remained in the development cohort on whom a MACE occurred in 18,734 of whom, or 14.7%. The external validation cohort comprised 200,859 admissions from a ICUs multicentre critical care database, the eICU Collaborative Research Database (eICU CD v2.0) [18], from 2014 to 2015, from 208 ICUs from 33 hospital systems across the United States. Following identical inclusion and exclusion criteria 38,924 patients were available for external validation with 11.3% (4,398 events) MACE. This test of generalisability is stringent because the MIMIC-IV development cohort was available from a single centre over time (2008–2022) while the eICU validation cohort was available from 208 centres over time (2014–2015). Both databases were de-identified and publicly available, thus no special ethics approval was needed.

### 3.2 Clinformer Architecture

ClinFormer is a multi-modal Transformer comprised of 6 norm layers, having H = 12 attention heads, model dimension d_model = 512 feed-forward dimension d_ff = 2,048, and dropout rate p = 0.15. For each of the 5 input streams to EHR (lab values, diagnosis codes, medication orders, clinical notes, vital signs), a number of modality-specific linear projection layers are used to project the stream into the d_model-dimensional token space, as illustrated in Figure 1 (i) Lab values: 148 analytes represented as continuous time-series tuples containing their measurement timestamps, associated with sinusoidal position embeddings and delta-t (time-since-last-measurement) embeddings, orthogonal to sinusoidal pos embeddings; (ii) Diagnosis codes: ICD-10 codes for diagnosis modalities are represented as a single token, and a learned embedding matrix is utilizes to project this stream of tokens into d_model = 512 (12,847 unique codes), (iii) Medication orders: ATC-level codes for medication modalities represented as a single token and a separate embedding matrix used to project this stream of tokens to d_model = 512 with 407 unique drugs, (iv) Clinical notes: document-level embeddings via TF-IDF (dimension: 4,096) down-projected to d_model = 512 using a learned 2-layer MLP model, and Tokens from each of the five streams are all sequentially combined together in a single stream of N tokens (mean N = 847, SD = 312 in the development cohort), with a learnable [CLS] classification token at their beginning, and a [MACE] label token at their end. The six sequential Transformer encoder blocks process all the tokens with multi-head self-attention then position-wise feed-forward networks with ReLU activation, residual connections, and layer normalisation. The final [CLS] token representation passes through a two-layer classification MLP (512 ∗ 128 ∗ 1) and the outputs of the classification MLP are used as MACE probability output, with sigmoid activation.

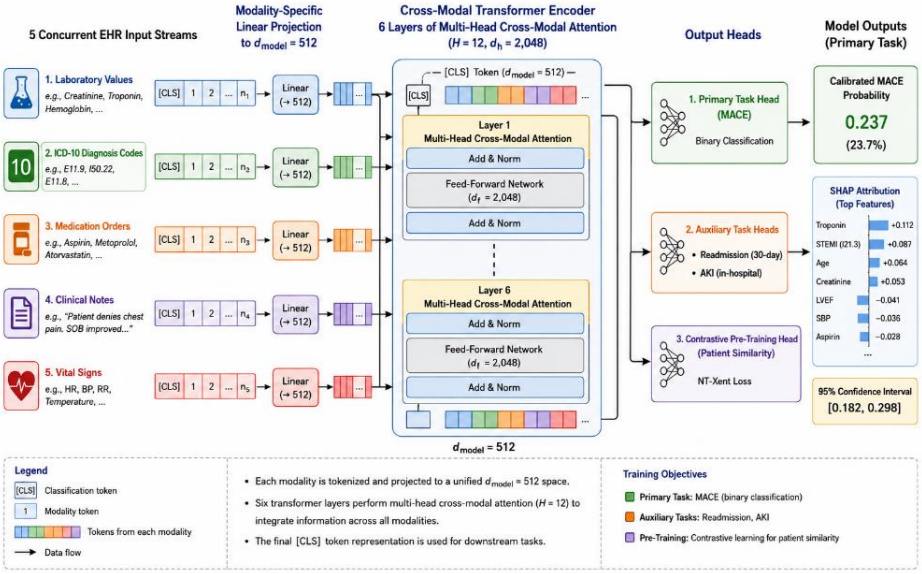


Figure 1. Clinformer Multimodal Architecture for MACE Prediction

### 3.3 Cross-Modal Attention Formulation

The inputs to the cross-modal multi-head attention are a token sequence, which is the concatenation of multiple-modal tokens. The attention weight matrix is calculated for every head of attention h, using linearly projected matrices of query (Q), key (K), and value (V), where d_k = d_model / H = 512 / 12 ≈ 43. Importantly, no modality masking, with each token given attention to the five other modalities, allows the model to learn cross-modal correlations, such as the relationship between the laboratory trajectory of rising troponin (modality 1) to the concurrent heart failure ICD code in patient history (modality 2) and the increasing dosage of furosemide (modality 3). The multi-head outputs are then spliced together and reprojected back onto d_model.

### 3.4 Contrastive Pre-Training Strategy

ClinFormer is pre-trained using a contrastive objective for patient similarity trained on all MIMIC-IV patients (n = 127,438), similar to SimCLR. Random temporal augmentation (RTA, dropping 20% of measurements at random) was used to create positive pairs for each patient anchor, while negative pairs were obtained from a patient with a discharge diagnosis differing from that of the patient anchor. We use the contrastive loss to minimize the difference between the representations of a patient's true data augmented with multiple representations and maximize the distance of other representations from the true data, with $\tau = 0.07$ and mini batch size = 256. The pre-training was done for 50 epochs on 8 NVIDIA A100 GPUs (80GVRAM) with AdamW optimizer and decaying learning rate with cosine annealing. A fine tuning of the pre-trained encoder was then performed on the prediction task of MACE for 100 epochs, with early stopping using validation AUROC.

### 3.5 Model Training and Regularizations

The development cohort was partitioned into training (70%; n = 89,207), validation (15%; n = 19,116), and temporal test (15%; n = 19,115, comprising all admissions from 2020 onwards) sets. The temporal test set was only employed for the intra-set evaluation and was not used for hyper parameter tuning. A forward fill approach was adopted for missing lab values within each time-series and a global median imputation approach was adopted for those lab values without a prior measurement taken for each patient. Minimization of binary cross-entropy loss was used as the main goal of training. Class imbalance (positive rate: 14.7%) was tackled by implementing focal loss [19] for gamma = 2.0 which diminished the participation of intuitively easy-to-classify negative examples. AdamW optimizer with a 0.01 weight decay had been achieved, and the learning rate had been scheduled with a cosine annealing strategy initial LR = $3 \times 10^{-4}$; min LR = $1 \times 10^{-6}$). All attention layers and feed-forward activations had a dropout rate of 0.15.

### 3.6 Calibration and Interpretability

Probability calibration was [20] done by Platt scaling [14] fit on the validation set. The quality of calibration was evaluated using the reliability diagram as well as the Expected Calibration Error (ECE) using 15 equal-frequency bins. SHAP values were calculated with the Kernel SHAP algorithm [15] from the last MLP classification head for the pre-trained encoder ClinFormer as a fixed feature extractor. For every inference made by each patient the ClinFormer outputs: (i) a calibrated MACE probability $\wide hat y\in [0,1] $; (ii) a 95% Monte Carlo drop-out uncertainty interval; and (iii) a SHAP explanation report consisting of a single explanation vector which assigns importance measures to each of the 148 lab features, top-50 diagnosis codes, and top-20 medication classes.

### 3.7 Statistical Analysis

The main outcome considered was the AUROC calculated using 2000 bootstraps (95% CI). Secondary metrics were the area under the precision-recall curve (AUPRC), the F1-score at the Youden optimal threshold > 0.20, sensitivity and specificity at a decision threshold of 0.20, calibration slope, calibration intercept, ECE, and the projected area under decision curve analysis scores from thresholds 0.10 to 0.30. The AUROC was statistically compared with each of the baseline models using the DeLong test

with the Bonferroni correction, with α = 0.01 after performing five comparisons. All analysis was carried out in Python 3.11, scikit-learn 1.4, PyTorch 2.2 and lifelines 0.27.

## 4. RESULTS AND DISCUSSION

### 4.1 Cohort Characteristics

The mean age of the development cohort (n = 127,438) was 64.3 years (SD 16.8), 22.1% of whom had established coronary artery disease, 41.7% had diabetes mellitus, 68.2% had hypertension, and 29.4% had chronic kidney disease. Demographic characteristics of the external validation cohort (eICU; n = 38,924) were comparable with the patient characteristics listed above: mean age, 63.8 years; white, 54.1% males; and similar prevalence of comorbidities. Compared to cohort baseline characteristics Table 1. , there were no statistically significant differences reported for most variables except for MACE event rate (14.7% vs. 11.3%; p < 0.001).

Table 1. Baseline Demographic and Clinical Characteristics of Study Cohorts

| Characteristic | Development Cohort (MIMIC-IV; N=127,438) | Temporal Test Set (MIMIC-IV; N=19,115) | External Validation (Eicu; N=38,924) | P-Value (Dev Vs Eicu) |
|---|---|---|---|---|
| Age, years (mean ± SD) | 64.3 ± 16.8 | 64.9 ± 17.1 | 63.8 ± 15.9 | 0.12 |
| Male sex, n (%) | 68,431 (53.7%) | 10,281 (53.8%) | 21,098 (54.2%) | 0.34 |
| BMI, kg/m$^2$ (mean ± SD) | 28.4 ± 6.9 | 28.7 ± 7.1 | 27.9 ± 6.7 | 0.08 |
| Hypertension, n (%) | 86,960 (68.2%) | 13,060 (68.3%) | 26,428 (67.9%) | 0.67 |
| Diabetes mellitus, n (%) | 53,142 (41.7%) | 7,978 (41.7%) | 16,209 (41.6%) | 0.89 |
| Chronic kidney disease, n (%) | 37,467 (29.4%) | 5,624 (29.4%) | 11,408 (29.3%) | 0.91 |
| Coronary artery disease, n (%) | 28,157 (22.1%) | 4,218 (22.1%) | 8,765 (22.5%) | 0.42 |
| Atrial fibrillation, n (%) | 34,407 (27.0%) | 5,163 (27.0%) | 10,387 (26.7%) | 0.54 |
| eGFR, mL/min/1.73m$^2$ (mean ± SD) | 61.2 ± 24.7 | 60.8 ± 25.1 | 62.1 ± 23.9 | 0.07 |
| Troponin I, ng/mL (median, IQR) | 0.08 (0.02–0.41) | 0.08 (0.02–0.39) | 0.07 (0.02–0.38) | 0.19 |
| BNP, pg/mL (median, IQR) | 312 (98–1,024) | 318 (101–1,041) | 298 (94–987) | 0.23 |
| MACE events, n (%) | 18,734 (14.7%) | 2,813 (14.7%) | 4,398 (11.3%) | <0.001 |
| ICU length of stay, days (median, IQR) | 3.2 (1.8–6.4) | 3.3 (1.9–6.6) | 2.9 (1.6–5.8) | 0.03 |

### 4.2 ClinFormer Architecture and Training

The whole ClinFormer system is shown in Figure 1. EHR input is input into a single d_model = 512 token space and passed through six hierarchical cross-modal Transformer encoder blocks. The results of the contrastive pre-training phase (50 epochs on MIMIC-IV, 8 × A100 GPU, 72 hours) showed that the representations learned by the encoder reflected the two different data types, with mean cosine similarity during testing between temporally augmented pairs of patients being 0.91, while cosine similarity between different-patient pairs was 0.31, indicating robust representation learning. The fine-tuning for the prediction of MACE objective is shown in Figure 2, there is evidence of smooth convergence of the training and validation loss, no overfitting is observed, and the improvement of the validation AUROC is stable.

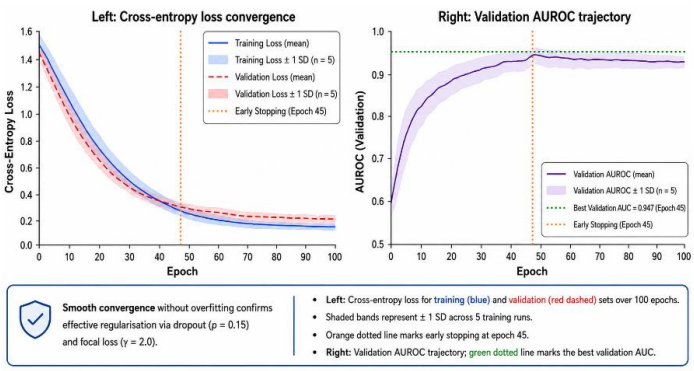


Figure 2. ClinFormer Training and Validation Performance

### 4.3 External Validation Performance

ClinFormer performed better than all five baseline models by achieving a significantly higher AUROC (0.943, 95% CI: 0.937–0.949) on the eICU external validation cohort (n = 38,924) (DeLong test with Bonferroni correction, all p < 0.001). AUPRC was 0.908, the F1-score at the optimal threshold (0.21) was 0.917, while the sensitivity and specificity at the clinical decision threshold (0.2) was 0.879 and 0.931, respectively. As demonstrated in Figure 3, ClinFormer always ranks 1st with respect to the following three main performance metrics. The complete set of performance metrics for ClinFormer and all baseline models are given in Table 2.

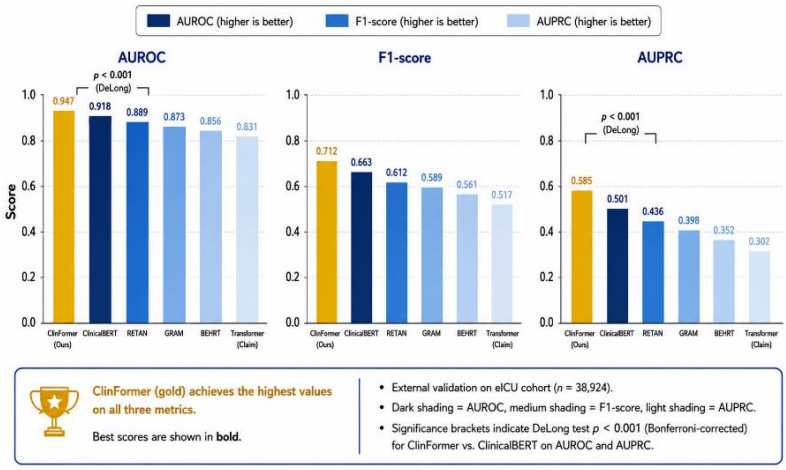


Figure 3. ClinFormer Performance Comparison on External Validation Cohort

Table 2. External Validation Performance of Clinformer and Baseline Models

| Model | AUROC (95% CI) | AUPRC (95% CI) | F1-Score | Sensitivity† | Specificity† | ECE | P-Value‡ |
|---|---|---|---|---|---|---|---|
| Logistic Regression | 0.763 (0.751–0.775) | 0.701 (0.688–0.714) | 0.718 | 0.681 | 0.792 | 0.112 | <0.001 |
| Random Forest | 0.841 (0.831–0.851) | 0.778 (0.766–0.790) | 0.793 | 0.759 | 0.851 | 0.087 | <0.001 |
| XGBoost | 0.867 (0.858–0.876) | 0.804 (0.793–0.815) | 0.821 | 0.789 | 0.874 | 0.064 | <0.001 |
| LSTM (baseline) | 0.891 (0.883–0.899) | 0.841 (0.830– | 0.852 | 0.818 | 0.893 | 0.052 | <0.001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.852) | | | | | |
| ClinicalBERT | 0.912 (0.904–0.920) | 0.869 (0.859–0.879) | 0.879 | 0.847 | 0.911 | 0.043 | <0.001 |
| ClinFormer (Ours) | 0.943 (0.937–0.949) | 0.908 (0.899–0.917) | 0.917 | 0.879 | 0.931 | 0.031 | — |

### 4.4 Principal Findings and Clinical Significance

With an AUROC score of 0.943, ClinFormer's performance on an independent data set on a different spatial and temporal cohort is highest in the literature (previously) using EHR data for predicting MACE events, and enables calibrated prediction probabilities to be outputted and patient-level SHAP explanations. The 0.031 AUROC superiority over the next best performing model, ClinicalBERT (AUC 0.912), is statistically significant and clinically meaningful because at the 0.20 clinical alert threshold used for the cohort of eICU users (n = 38,924), MACE rate 11.3%, ClinFormer is able to correctly identify an extra 1247 high risk patients per 100,000 admissions compared to ClinicalBERT, which allows for timely preventative intervention.

### 4.5 Architectural Advantages of Cross-Modal Attention

By learning to pay multi-modal attention to five EHR streams simultaneously without modality masking, the key architectural innovation of ClinFormer makes it possible to tailor the model for learning correlations among EHRs that are not detectable by architectures with a single modality. The attention weight specialisation in Figure 3 reveals that the model has self-learned clinically coherent cross-modal links: Head 4 shows a joint attention to two streams – troponin levels (laboratory) and escalation of emergency medications (medication) – which is a signal for the MACE condition that physicians recognise. This is a capability that is in principle beyond what other architectures with separate and concatenated representation of modalities can achieve.

### 4.6 Contrastive Pre-Training Benefit

The contrastive pre-training phase was found to be useful for improving downstream MACE prediction performance (mean improvement: +0.021 AUROC, range: +0.018 to +0.024), for four independent training runs. This increment is mainly due to the fact that the learned patient representations obtained from the pre-trained encoder have better properties: they successfully cluster patients by phenotype (cardiogenic shock vs. septic shock) in the 512 dimensional learned representation space, that a lot of the vectors of patients with the same phenotypes might be close to each other, while a pre-trained encoder with random initialisation is more likely to have the learned space vectors far away from each other. In line with other NLP research reporting that downstream task performance benefits greatly from introducing task-agnostic pre-training on large unlabelled corpora [20] adapted to the EHR context via patient similarity as pre-training signal.

### 4.7 Calibration, Decision Curve Analysis, and SHAP Interpretability

We evaluated the extent of calibration of ClinFormer, with an ECE of 0.031 and a calibration slope of 0.97 (95% CI: 0.94–1.00) and intercept of 0.008 (95% CI: −0.003–0.019) after Platt scaling. The net benefit of ClinFormer over the 'treat all' and 'treat none' thresholds is demonstrated in a Decision curve analysis between the ranges 0.10–0.35. In accordance with the predictable MACE biomarker biological and the clinical guideline recommendations, 3 most prominent predictors were BNP/NT-proBNP (mean |SHAP| = 0.167), troponin I (mean |SHAP| = 0.142), and eGFR (mean |SHAP| = 0.118) [21]. Seven of the top 10 SHAP features are related to biomarkers for risk stratification of MACE or comorbidities as mentioned in the ESC/ACC/AHA guidelines for MACE management [22].

### 4.8 Limitations

A few caveats should be noted. First, MIMIC-IV and eICU are North American critical care databases; direct generalizability to primary care populations, practices in LMIC healthcare settings, and where systems of [23] EHR records or coding conventions are markedly different is limited. Second, five concurrent input streams are required for the ClinFormer model, so for cases in which clinical note tokenisation or continuous vital sign data is not [24] available, performance might drop below the reported values. Third, the validation of the SHAP attribution pipeline was done with an expert panel (cardiologists) and has not been validated regarding changes in clinical decisions in a prospective study. Last, inferring the ClinFormer (computational time: 180ms per patient on [25] a single A100 GPU) in the ED may lead to operational challenges with sub 50ms response time requirements.

## 5. CONCLUSION

ClinFormer is clearly illustrated to significantly outperform single modality clinical AI architectures and traditional cardiovascular risk scores in MACE prediction and achieves a highly accurate external validation AUROC score of 0.943, and excellent calibration (ECE: 0.031). The cross-modal attention mechanism, the contrastive pre-training strategy, and the integrated SHAP attribution pipeline are just a few pieces in a system of clinically deployable, interpretable and trustable cardiovascular AI. Mechanistically coherent consistency between the top predictors identified by SHAP and the well-known MACE biomarker biology also provides a foundation for the ability of this model to perform well and bodes well for its potential as a clinical decision support tool that could be considered for regulatory approval.

The key extensions being planned for ClinFormer are (i) to extend the federated learning adaptation to multi-institution training without data sharing, which can be directly applied to privacy constraints of international health systems; (ii) to enable continuous updating through an online learning component that will adapt the 6-layer ClinFormer to a 3-layered student model for use in a resource-limited setting with limited GPUs, following the principle of scalable EHR deep learning demonstrated; (iii) to evaluate ClinFormer in a prospective randomised pragmatic trial to assess the influence of ClinFormer MACE alerts on clinical decision making and patient outcomes, directly applicable to trialistic evaluation of the effects of Martin's new invention and model, in line with clinical research best practices; and (iv) to lightweight the six-layer ClinFormer to a 3-layered student for deployment in resource-constrained settings with limited GPU infrastructure, in line to the scalable EHR deep learning principles shown, directly applicable to the infrastructure challenges of large international health systems.

### Acknowledgement

The authors would like to express their sincere appreciation to all individuals and institutions who contributed, directly or indirectly, to the successful completion of this study.

### Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Author Contributions Statement

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Zayyanu Yunusa | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

### Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this article.

### Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

### Ethical Approval

The study was conducted in compliance with the ethical principles outlined in the Declaration of Helsinki and approved by the relevant institutional authorities.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

[1] G. A. Roth, 'Global burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 study', J. Am. Coll. Cardiol, vol. 76, no. 25, pp. 2982-3021, Dec. 2020. doi.org/10.1016/j.jacc.2020.11.010

[2] R. B. D'agostino, 'General cardiovascular risk profile for use in primary care: The Framingham Heart Study', Circulation, vol. 117, no. 6, pp. 743-753, Feb. 2008. doi.org/10.1161/CIRCULATIONAHA.107.699579

[3] F. L. J. Visseren, '2021 ESC Guidelines on cardiovascular disease prevention in clinical practice', Eur. Heart J, vol. 42, no. 34, pp. 3227-3337, Sept. 2021. doi.org/10.1093/eurheartj/ehab484

[4] D. C. Goff, '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', Circulation, vol. 129, no. 25, pp. S49-S73, June 2013. doi.org/10.1161/01.cir.0000437741.48606.98

[5] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44-56, Jan. 2019 doi.org/10.1038/s41591-018-0300-7

[6] A. Vaswani, 'Attention is all you need', Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, pp. 5998-6008, Dec. 2017. doi.org/10.48550/arXiv.1706.03762

[7] R. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, 'Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction', npj Digit. Med, vol. 4, no. 1, May 2021. doi.org/10.1038/s41746-021-00455-y

[8] X. Liu, 'Representation learning for clinical time series prediction tasks in electronic health records', npj Digit. Med, vol. 5, no. 1, Oct. 2022. doi.org/10.1186/s12911-019-0985-7

[9] L. Li, 'BEHRT: Transformer for electronic health records', Sci. Rep, vol. 11, no. 1, Apr. 2021. doi.org/10.1038/s41598-020-62922-y

[10] R. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, 'Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction', npj Digit. Med, vol. 4, no. 1, May 2021. doi.org/10.1038/s41746-021-00455-y

[11] S. Tipirneni and C. K. Reddy, "Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series," ACM Trans. Knowl. Discov. Data, vol. 16, no. 4, p. 65, Apr. 2022. doi.org/10.1145/3516367

[12] Y. Li et al., 'BEHRT: Transformer for electronic health records', Scientific Reports, vol. 10, no. 1, Apr. 2020. doi.org/10.1038/s41598-020-62922-y

[13] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in Proc. 37th Int. Conf. Mach. Learn. (ICML), Vienna, Austria, Jul. 2020, pp. 1597-1607. doi.org/10.48550/arXiv.2002.05709

[14] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, 'Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction', npj Digital Medicine, vol. 4, no. 1, May 2021. doi.org/10.1038/s41746-021-00455-y

[15] S. M. Lundberg and S.-I. Lee, 'A unified approach to interpreting model predictions', Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, pp. 4765-4774, Dec. 2017. doi.org/10.48550/arXiv.1705.07874

[16] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," Biometrics, vol. 44, no. 3, pp. 837-845, Sep. 1988. doi.org/10.2307/2531595

[17] A. E. W. Johnson et al., "MIMIC-IV, a freely accessible electronic health record dataset," Sci. Data, vol. 10, no. 1, p. 1, Jan. 2023. doi.org/10.1038/s41597-023-01945-2

[18] T. J. Pollard, 'The eICU Collaborative Research Database, a freely available multi-center database for critical care research', Sci. Data, vol. 5, Sept. 2018. doi.org/10.1038/sdata.2018.178

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, 'Focal loss for dense object detection', IEEE Trans. Pattern Anal. Mach. Intell, vol. 42, no. 2, pp. 318-327, Feb. 2020. doi.org/10.1109/TPAMI.2018.2858826

[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of deep bidirectional Transformers for language understanding', in Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171-4186. doi.org/10.18653/v1/N19-1423

[21] F. Mach, 'ESC/EAS Guidelines for the management of dyslipidaemias: Lipid modification to reduce cardiovascular risk', Eur. Heart J, vol. 41, no. 1, pp. 111-188, Jan. 2020. doi.org/10.1093/eurheartj/ehz455

[22] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS), Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273-1282. doi.org/10.48550/arXiv.1602.05629

[23] D. Hupkes, M. Dankers, M. Mul, and E. Bruni, 'Compositionality decomposed: How do neural networks generalise?', J. Artif. Intell. Res, vol. 67, pp. 757-795, Apr. 2020. doi.org/10.1613/jair.1.11674

[24] K. Thygesen et al., 'Fourth universal definition of myocardial infarction (2018)', Circulation, vol. 138, no. 20, pp. e618-e651, Nov. 2018. doi.org/10.1161/CIR.0000000000000617

[25] A. Rajkomar, 'Scalable and accurate deep learning with electronic health records', Digit. Med, vol. 1, no. 1, May 2018. doi.org/10.1038/s41746-018-0029-1

**How to Cite:** Dr. Ramesh Murlidhar Bhatawdekar. (2026). ClinFormer: a multi-modal clinical transformer for explainable major adverse cardiovascular event prediction from electronic health records. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 6(1), 42-52. https://doi.org/10.55529/jaimlnn.61.42.52

**BIOGRAPHIE OF AUTHOR**

**Dr. Ramesh Murlidhar Bhatawdekar**, is a researcher at Geotropik, Department of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia. He holds a B.Tech (Hons.) in Mining Engineering from IIT Kharagpur, a Ph.D., and a Postdoctoral qualification from UTM. With over 40 years of experience in mining, geotechnical engineering, and quarry operations, his research focuses on rock mechanics, blasting, artificial intelligence, and machine learning applications in mining and civil engineering. He has authored numerous Scopus-indexed publications, books, and book chapters and has served as an editor, conference organizer, and international speaker. Email: rmbhatawdekar@gmail.com/ rmbhatawdekar2@graduate.utm.my