

**A SMOTE-Optimized RoBERTa Stacking Framework for Multiclass Cardiovascular
Risk Prediction**

Mayuresh Sinha

Mtech Scholar, Department of Computer Science Engineering, Lakshmi Narain College of
Technology
mayureshsinha9292@gmail.com

Rajkumar Sharma

Assistant Professor, Department of Computer Science Engineering, Lakshmi Narain College
of Technology
rajkumars@lnct.ac.in

Abstract: Predicting multiclass cardiovascular risk from the diagnostic text data is complex because of class imbalance, weak feature representation, and reliance on individual classifiers. The current research aims to develop an improved system that combines RoBERTa-enriched embeddings, SMOTE, Optuna optimization, and stacking ensemble learning techniques. For the experiment, 21,430 samples from the PTB-XL dataset were used. Each record is labeled according to the five possible classes: NORM, MI, STTC, CD, and HYP, divided into train and test parts with 80/20 ratio. The diagnostic text is produced from SCP statements, and then it is represented with the concatenation of 768-dimensional CLS embedding and 768-dimensional mean-pooled RoBERTa embedding to form 1536-dimensional vectors. SMOTE is used for class balancing, and Optuna optimizes XGBoost and LightGBM models. The Results show that the framework provides 99.68% accuracy, precision, recall, and F1-score, along with 100% weighted ROC-AUC. In addition to this, the number of inaccurate predictions made by the existing WPSA-DRF framework was 243, while it became 1.

Keywords: Cardiovascular risk prediction, RoBERTa, PTB-XL, NLP, SMOTE, Optuna, stacking ensemble, XGBoost, LightGBM.

1. Introduction

CVD continues to pose a significant threat to global health, hence the need for accurate, efficient, and personal CVD risk evaluation. Recent developments in natural language processing (NLP) and LLMs have broadened the capacity of AI technologies to retrieve clinically relevant information from complex biomedical and cardiovascular data, allowing early diagnostics and personalized cardiovascular risk estimation [1]. AI has also found applications in hypertension by means of CVD risk prediction, phenotyping, blood pressure measurement, and precision management, proving itself to be a tool that may be used for personalized CVD treatment [2]. Apart from individual applications, combining deep learning, NLP, federated learning, and multimodal data has enabled the use of AI for prediction, diagnosis, treatment, rehabilitation, and continuous CVD monitoring [3]. Parallel developments have taken place in the domain of stroke care, especially in imaging, risk estimation, atrial fibrillation detection, and rehabilitation, though turning the performance of algorithms into clinical results proves to be difficult [4].

The advent of electronic health records (EHRs) has also contributed to the rise of data-driven medicine through provision of structured and unstructured data on diagnosis, medications, lab results, clinical notes, and patient history. Unstructured data in EHRs can be leveraged via NLP and deep learning in the creation of clinical representations that aid disease diagnosis and decision-making, while issues related to privacy, heterogeneity, bias, and robustness should also be considered [5]. AI-enhanced electronic medical record systems can also integrate physiological monitoring, lab data, predictive models, and language-based explanation for the purpose of identifying deteriorating patients early on [6]. Overall, AI-facilitated text mining is a great method of mining patterns of clinically relevant information from unstructured biomedical data [7].

Furthermore, artificial intelligence contributes to precision medicine through incorporation of more advanced biological and clinical data. In cardiology genomics, machine learning, deep learning, and generative artificial intelligence might help in genetic variant interpretation, effect prediction, and prioritization; however, explainability and integration in the clinical setting represent major issues [8]. Natural-language intelligence also represents a promising approach for chronic disease management due to its ability to assist with diagnostics, patient self-management, adherence to therapy, and personalized health care [9]. Regarding heart failure and cardiomyopathies, the AI-driven analysis of ECG, echocardiography, EHRs, wearables, and natural-language intelligence data is moving from model development to clinical use [10].

Moreover, transformer-based models and LLMs have additionally enhanced the analysis of unstructured medical data through the understanding of contextual and semantic relations in the clinical texts. They are used for EHR summarization, information extraction, documentation, answering medical questions, and biomedical knowledge processing; however, hallucination, bias, privacy issues, reproducibility, and the lack of prospective validation are major concerns here [11]. Additionally, conversational LLMs have been proven to hold great promise in personalized health communication, risk assessment, and patient engagement in maternal and neonatal healthcare [12]. In the area of cardiovascular applications, initial evidence on heart failure prognosis shows that LLMs are especially good at analyzing unstructured discharge summaries, but their reproducibility and performance should be validated [13].

AI-assisted medicine developments also show how necessary is personalized and multimodal prediction in AI. AI-assisted digital interventions have been explored in terms of increasing the physical activity of people suffering from multimorbidity; however, there is currently little evidence on their effectiveness and consistency [14]. Multi-omics, microbiome, clinical, and wearable data are used together for risk prediction and personalized intervention within the field of obesity research through using AI [15]. In addition, Medical LLMs are tested in clinical decision support, documentation, diagnostics, biomedical research, and patient interaction; however, it is crucial that they are trustworthy and explainable in this case [16].

Furthermore, generative AI can be helpful to healthcare professionals by providing automated documentation, summarization of patient information, as well as evidence-based decision

support; nevertheless, clear governance and the maintenance of human judgment are crucial [17]. Intelligent platforms that include medical data management, disease risk prediction, and natural language processing-based conversation support are good examples of the use of AI systems in order to provide proactive monitoring and early disease detection [18]. However, traditional machine learning and deep learning models are also very useful in their own ways: on one hand, Random Forest and Support Vector Machines can be used for efficient analysis of structured data, while deep neural networks and Transformers are perfect for unstructured and high-dimensional medical data [19]. Regarding cardiovascular diseases, ML, DL, and ECG analysis via AI have shown significant possibilities for risk prediction [20].

Despite these advancements, challenges like lack of features representation, class imbalanced, and dependence of each classifier on particular features make multiclass cardiovascular predictions using diagnostic text challenging. For this reason, the current paper introduces a framework for the purpose of making predictions on cardiovascular risks based on multiclass classification using enhanced RoBERTa embeddings, class imbalance solved by SMOTE, automated hyperparameter tuning, and stacking. Here, diagnostic statements are represented using a 1536-dimensional concatenated vector of CLS and mean pooled 768-dimensional RoBERTa embeddings obtained from standard SCP statements in the PTB-XL data set. Finally, SMOTE is used to balance the five classes, and the ensemble consists of XGBoost, LightGBM, and Random Forest with Logistic Regression as meta-classifier.

1.1 Objectives and Contributions

The main objective of this work is to enhance the representation of RoBERTa text with a combination of CLS and mean pooling representations in one high-dimensional feature vector. For solving the problem of class imbalance, the SMOTE technique is used for improving the representation of minority diagnostic classes. The model selection process involves the application of Optuna-based Bayesian hyperparameters optimization technique along with stratified five-fold cross-validation. Additionally, the stacking ensemble, which consists of XGBoost, LightGBM, and Random Forest algorithms together with logistic regression as the meta-learner, is created for increasing the accuracy of predictions.

2. Related Work

Li, A. et al introduced a framework of RoBERTa-WPSA-DRF that can predict CVDs with an accuracy of 92.5%, showing the possibility of NLP in early diagnosis and personalized medicine [1]. Varzideh, F. et al. studied the application of AI in hypertension management, mentioning NLP, wearables, and decision-making support tools, but pointing to the importance of validation and explainability [2]. Kong, M et al. discussed the AI applications in cardiovascular prediction, diagnostics, therapy, and monitoring and mentioned the problems associated with reliability, scalability, and privacy issues [3]. Urfy, M., & Mir, M. T. analysed the use of AI in stroke, presenting achievements in image processing, atrial fibrillation, and rehabilitation and noting clinical validation and scalability as critical challenges [4]. Chakraborty and Kamila reviewed the use of NLP- and DL-based EHR analysis for disease detection and clinical decision support in terms of privacy issues, biases, and data heterogeneity problems [5]. Boulanger proposed MedROAD V2 that combines ML models and an LLM to

predict patient deterioration and provide explainable clinical decision support [6]. Manoharan and Sehgal explored the applicability of NLP, Transformers, and LLMs to text mining in biomedicine while emphasizing issues related to biases, privacy, and clinical translation [7]. De Paoli et al. discussed the application of ML, DL, and generative AI in cardiovascular genomics for variants interpretation and precision medicine while pointing out the problems related to data heterogeneity and explainability [8]. Zeng et al. stated that NLIT could be helpful for chronic diseases management, self-management, diagnosis, and medication adherence, though heterogeneity and bias were to be confirmed [9]. Croon et al. stressed the use of AI together with ECG, echocardiography, EHRs, wearables, and LLMs in detection and management of heart failure, while generalizability, interoperability, bias, and adoption issues were indicated [10]. Hornback et al. outlined applications of LLM in EHRs summarization, documentation, medical QA, genomics, and drug discovery, while pointing out hallucinations, bias, privacy issues, reproducibility problems, and lack of clinical validation as major limitations [11]. Rasoli et al. pointed out that LLM/GPT-based chatbots could assist in maternal and neonatal care due to psychoeducation, mental health assistance, engagement, and pregnancy risk prediction [12]. Shang et al. found that LLMs were more promising for unstructured clinical text than for structured data in predicting heart failure, though emphasized the necessity for better reproducibility and broader validation [13]. Pinn et al. reviewed the application of AI-based digital interventions in physical activity for people with multimorbid conditions, pointing to possible advantages and poor evidence [14]. Wang discussed the utilization of machine learning, deep learning, federated learning, and language models in managing obesity for predicting risks and personalizing treatment approaches, yet mentioning the issues of explainability, fairness, data quality, and ethics [15]. Saha et al. reviewed Medical Language Models in medicine for clinical support, documentation, diagnoses, drug development, and interactions with patients, pointing out the questions of trustworthiness, explainability, bias, reliability, and ethics [16]. Dhavale et al. considered the possibility of using generative AI in supporting clinical decision-making, documentation, and summarization of patient data, focusing on the issues of fairness, transparency, governance, and human supervision [17]. Bhardwaj et al. introduced HealthCare.AI combining patient data management, multi-disease risks assessment, and NLP chatbot [18]. Li compared ML and deep learning techniques in the context of healthcare, pointing out that the traditional ML technique has better interpretability while deep learning is better suited for unstructured data, with hybrid and XAI being prospective techniques [19]. Tiwari et al. emphasized the possibilities of using ML, DL, and AI-powered ECG analysis in cardiovascular risk prediction despite the lack of understanding among health professionals [20]. Table 1 presented below gives an account of the comparative review of AI, NLP, and large language model techniques in cardiovascular risk prediction and health applications.

Table 1. Comparative Review of AI, NLP, and Large Language Model Approaches for Cardiovascular Risk Prediction and Healthcare Applications

Ref.	Title	Method/Technique	Key Findings	Limitations
------	-------	------------------	--------------	-------------

Li, A. et al [1] (2025)	AI driven cardiovascular risk prediction using NLP and LLMs	RoBERTa, NLP, WPSA-DRF	Improved cardiovascular risk prediction with high diagnostic accuracy	Limited explainability and dependency on large clinical datasets
Varzideh, F. et al. [2] (2026)	AI in Cardiovascular Medicine: Focus on Hypertension	AI and ML models	Enhanced hypertension prediction and precision healthcare	Limited real-time clinical validation
Kong, M et al. [3] (2026)	Intelligent cardiovascular system platform	AI-integrated healthcare platform	Improved cardiovascular diagnostics and decision-making	High computational and infrastructure requirements
Urffy, M., & Mir, M. T. [4] (2026)	AI in Stroke Care	AI-based precision medicine	Better stroke prediction and rehabilitation analysis	Limited generalization across diverse populations
Chakraborty, P., & Kamila, S. [5] (2026)	AI Integration in Electronic Health Records	AI-enabled EHR systems	Improved healthcare management and patient monitoring	Privacy and data security concerns
Boulanger, Pierre. [6] (2026)	MedROAD V2	AI-integrated EMR system	Enhanced clinical decision support and analytics	Requires large-scale healthcare integration
Manoharan, J., & Sehgal, Y. [7] (2026)	AI-Enabled Text Mining	NLP and text mining	Improved disease prediction and drug discovery	Limited accuracy with unstructured noisy data
De Paoli, F. [8] (2026)	AI Approaches for Cardiovascular Genomics	ML and genomics analysis	Enhanced biomarker identification and genomic prediction	High dependency on genomic data quality
Zeng, J., et al. [9] (2026)	NLP Technology in Chronic	Natural language intelligence	Improved patient care and nursing efficiency	Limited multilingual support

	Diseases Nursing			
Croon et al. [10] (2026)	AI Tools for Heart Failure Detection	AI diagnostic systems	Improved detection of heart failure and cardiomyopathies	Clinical implementation challenges
Hornback et al. [11] (2026)	LLMs in Healthcare and Biomedical Informatics	Large Language Models	Advanced medical text analysis and clinical support	Risk of hallucination and biased outputs
Rasoli et al. [12] (2026)	GPT-driven Conversational Intelligence	Conversational AI and GPT	Enhanced maternal and newborn healthcare engagement	Ethical and privacy-related challenges
Shang, L., [13] (2026)	LLMs in Heart Failure Prediction	Large Language Models	Improved prognostic prediction compared to traditional models	Small-scale pilot study limitations
Pinn, C. K., [14] (2026)	AI Digital Tools for Physical Activity	AI-integrated monitoring tools	Promoted physical activity in multimorbidity patients	Limited long- term validation studies
Wang [15] (2026)	AI Across the Obesity Continuum	AI-driven healthcare analytics	Improved obesity prevention and personalized therapy	Requires continuous patient monitoring
Saha et al. [16] (2026)	Medical-LLMs Evaluation Framework	Medical Large Language Models	Evaluated effectiveness of LLMs in healthcare tasks	Benchmark datasets may not represent real-world scenarios

The existing approach is based on one single mean-pooled RoBERTa embedding, does not take into consideration class imbalance of the diagnoses, implements limited hyperparameter tuning, and uses only one single Dynamic Random Forest model. The novelty of this study lies in using enriched RoBERTa embeddings, class balance adjustment, automatic hyperparameter tuning, and stacking of multiple models for heterogeneous multiclass cardiovascular risk prediction. In particular, CLS and mean-pooled embeddings are combined into one single 1536-dimensional embedding, the class balance problem is addressed by applying SMOTE,

Optuna is utilized for hyperparameter tuning, and XGBoost, LightGBM, and Random Forest are stacked by means of Logistic Regression.

3. Proposed Methodology

3.1 Dataset Description

The proposed framework uses the PTB-XL ECG dataset from PhysioNet. SCP codes are transformed into natural language texts that are further processed for NLP and RoBERTa feature extraction. After excluding all entries with no diagnoses, 21,430 ECG samples are classified into five cardiovascular categories, as listed in Table 2 below.

Table 2. Diagnostic Class Distribution of the PTB-XL Dataset

Diagnostic Class	Description	Number of Records
NORM	Normal ECG	9,528
MI	Myocardial Infarction	5,485
STTC	ST/T Change	3,872
CD	Conduction Disturbance	2,009
HYP	Hypertrophy	536
Total	All Diagnostic Records	21,430

In order to avoid any kind of information leak, a patient-level split is done with a 80:20 ratio such that none of the records from the same patient appear in both splits. The data split includes 17,112 training instances and 4,318 testing instances with the diagnoses categories coded numerically.

3.2 Data Preprocessing

This phase involves the transformation of structured SCP codes into human-readable diagnoses ready for RoBERTa-based processing. Multiple diagnoses associated with one and the same ECG are concatenated and slightly normalized while preserving clinically-relevant information. The text is then tokenized with the RoBERTa tokenizer and padded/truncated to get the uniform sequence. Finally, an 80:20 patient-based splitting yields 17,112 train and 4,318 test ECGs, thus avoiding data leakage by preventing one patient from being represented in both datasets.

3.3 Existing WPSA-DRF Approach

The existing approach utilizes RoBERTa for obtaining contextual embeddings and mean pooling of the resulting sequence to get a 768-dimensional feature vector. These features are then ranked and filtered according to their importance scores calculated with the use of Random Forest. The WPSA tunes the number of trees, tree depth and number of features, after which the DRF predicts the diagnosis (NORM, MI, STTC, CD, HYP).

3.4 Proposed Model

The proposed framework upgrades the existing framework using both the CLS-token and the attention-masked mean pooled RoBERTa embeddings for obtaining a more elaborate 1536

dimensional feature representation. This is followed by standardization of features and class balance using SMOTE. The hyperparameter tuning of the models XGBoost and LightGBM is done using Optuna with TPE, and in the stacking framework, XGBoost, LightGBM, and Random Forest are used as base learners with Logistic Regression as the meta learner. Regularization techniques and five-fold out-of-fold predictions are applied to avoid overfitting. Figure 1 for the RoBERTa Three-model Stacking Ensemble Framework.

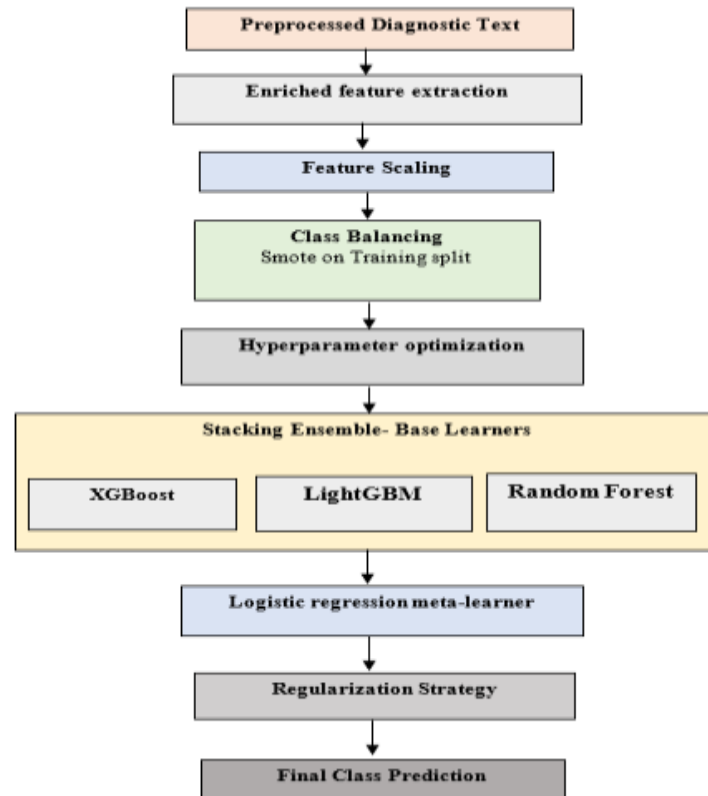


Figure 1 Proposed RoBERTa Three-model Stacking Ensemble Framework

3.5 Evaluation Metrics

Accuracy: Measures the percentage of correctly classified records.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: Represents the number of records predicted to be a particular class which belong to that particular class.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall / Sensitivity (True Positive Rate): Represents the ability to correctly classify records belonging to a particular cardiovascular class.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Specificity (True Negative Rate): Represents the ability to correctly classify records which do not belong to a considered class.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

F1-Score: Balances the measures of precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Confusion Matrix: Represents true and false predictions made by the classifier with respect to the five classes namely NORM, MI, STTC, CD, and HYP through TP, TN, FP, and FN.

AUC-ROC: Evaluates the classification model's performance to differentiate between each of the classes of cardiovascular disease from the rest of the other classes through different thresholds, where,

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN} \quad (6)$$

4. Results and Discussion

Table 3 below shows that the proposed model exhibits outstanding classification accuracy for all the five cardiac disease classes. MI showed a precision, recall, and F1 score of 100%, and the NORM and STTC class had an F1 score of above 99%. Although HYP and CD have relatively low support compared to other classes, they had F1 scores of 99.04% and 98.66%, respectively. This shows that this model is effective at recognizing minority classes.

Table 3 Class-wise Performance of the Proposed RoBERTa Three-model Stacking Ensemble Framework

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
CD	98.14	99.19	98.66	372
HYP	98.10	100.00	99.04	103
MI	100.00	100.00	100.00	1,124
NORM	99.90	99.74	99.82	1,939
STTC	99.61	99.23	99.42	780
Weighted Avg	99.68	99.68	99.68	4,318

The figure 2 below shows the performance of the classification of CD, HYP, MI, NORM, and STTC classes. The precision, recall, and F1-score are consistently high in all five diagnostic classes and also above 95% threshold value, which reflects the excellent and consistent multiclass classification performance without degradation of any diagnostic class.

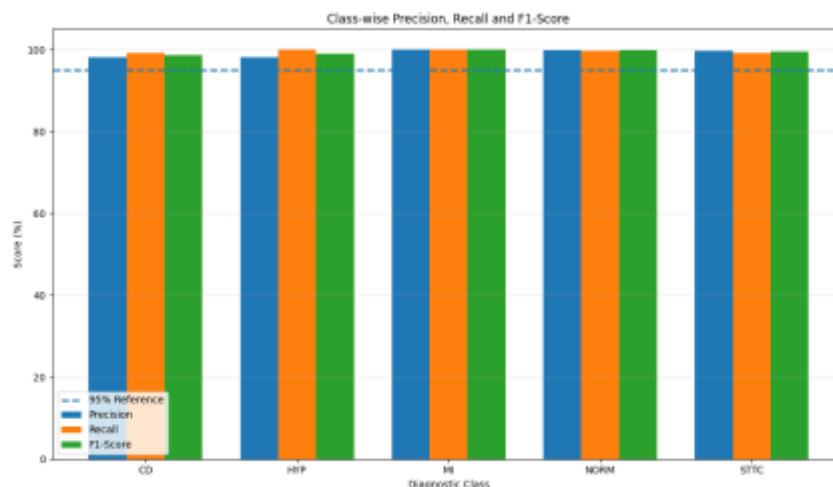


Figure 2 Class-wise Precision, Recall, and F1-Score of the Proposed Framework

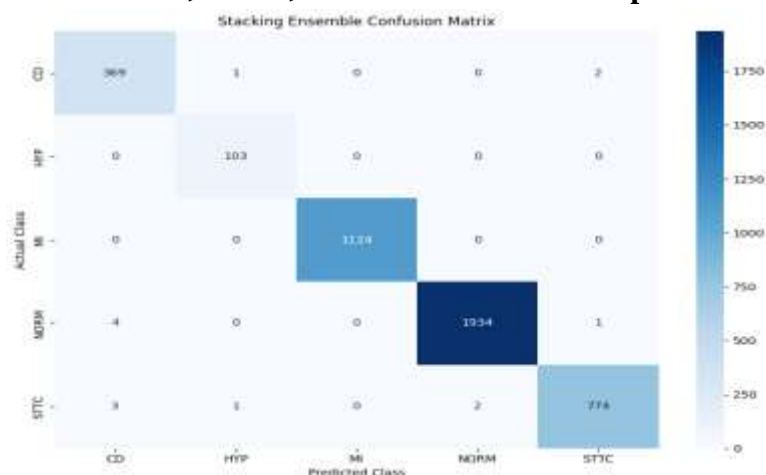


Figure 3 Confusion matrix of the proposed RoBERTa Three-model Stacking Ensemble Framework

The confusion matrix in Figure 3 shows the robust class-wise accuracy achieved by the proposed stacking ensemble on 4,318 test cases. The number of correctly identified CD, HYP, MI, NORM, and STTC cases was 369, 103, 1,124, 1,934, and 774, respectively, for a total of 4,304 correctly classified cases and just 14 incorrect classifications. It is important to note that there was no mistake made when classifying MI and HYP cases, whereas only minor confusion took place between CD, NORM, STTC, and HYP cases.

Table 4 Overall Comparison of the Existing and Proposed Models

Metric	Existing (%)	Proposed (%)
Accuracy	92.50	99.68
Precision	92.70	99.68
Recall (Sensitivity)	99.23	99.68
F1-Score	95.60	99.68
ROC-AUC	95.00	100.00

Table 4 shows that the proposed stacking method achieves higher performance than the existing model across all reported metrics. Accuracy increases from 92.50% to 99.68%, precision from 92.70% to 99.68%, recall from 99.23% to 99.68%, and F1-score from 95.60% to 99.68%. ROC-AUC also improves from 95.00% to 100.00%, indicating stronger class discrimination. These results demonstrate improved classification performance, with the largest gains observed in accuracy and precision.

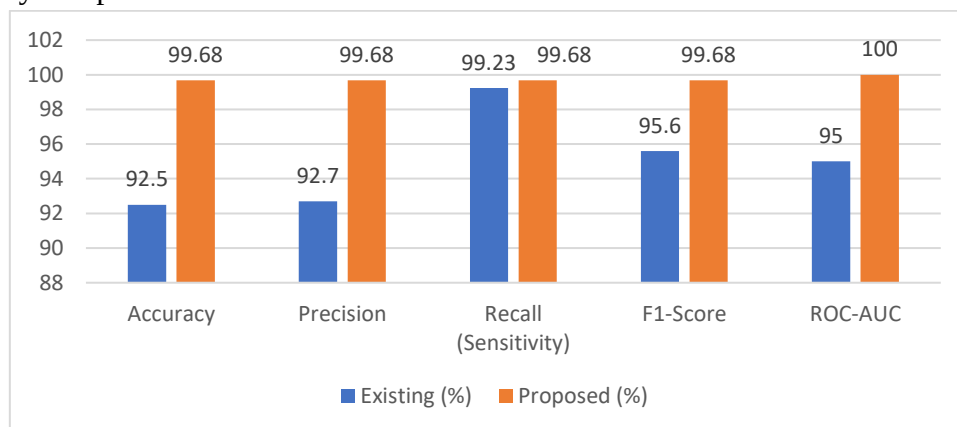


Figure 4 Comparison of Performance Metrics Existing WPSA–DRF vs Proposed RoBERTa Three-model Stacking Ensemble Framework

Figure 4 shows the comparison of both models in terms of their performance based on various performance criteria. The suggested framework performs better than the current model by getting 99.68% accuracy, precision, recall and F1 score with ROC-AUC score of 100%. The most important point is that there is a drastic reduction in the number of false predictions from 243 to 14.

5. Conclusion

The proposed study presented a novel stacked ensemble framework of RoBERTa with improvements for multiclass cardiovascular disease prediction using PTB-XL diagnostic text. The new framework addresses shortcomings of the existing framework through improved CLS and mean-pooling embeddings, SMOTE balancing, Optuna optimization, and heterogeneous stacking learning. It recorded 99.68% accuracy, precision, recall, and F1-score and 100% weighted ROC-AUC, whereas the existing approach obtained 92.50% accuracy and 95.60% F1-score. Moreover, misclassification reduced from 243 in the existing approach to 14 in the new framework. The results show that performance is robust in all the five classes, including CD and HYP. The above results prove that the combination of contextual textual representation, balancing, optimization, and stacking works effectively in cardiovascular diagnostic classification. These results clearly prove the efficiency of using contextual representation of texts, balancing, automatic optimization, and stacking in cardiovascular diagnostic classification. Future studies could consider validation of the framework in other independent and varied clinical datasets in order to examine generalization properties that extend beyond PTB-XL. The current framework can be further expanded through incorporating the diagnosis and the raw ECG signal along with other types of clinical information. Future research could involve utilization of explainable AI methods in order to understand model

predictions and computationally feasible implementation of the models in clinical settings. Clinical evaluation using real-life patient data will also add value prior to any practical applications.

References

- [1] Li, A., Wang, Y., & Chen, H. (2025). AI driven cardiovascular risk prediction using NLP and large language models for personalized medicine in athletes. *SLAS technology*, 32, 100286.
- [2] Varzideh, F., Mone, P., Kansakar, U., Pande, S., Jankauskas, S. S., & Santulli, G. (2026). Artificial intelligence in cardiovascular medicine: Focus on hypertension. *Hypertension*, 83(6), e26094.
- [3] Kong, M., Zhao, X., Li, Q., Yu, Y., & Zhang, C. (2026). Building an intelligent cardiovascular system platform: Embedding artificial intelligence across all facets of cardiovascular medicine. *Advanced Intelligent Systems*, 8(3), e202501136.
- [4] Urfy, M., & Mir, M. T. (2026). A Decade of Artificial Intelligence in Stroke Care (2015–2025): Trends, Clinical Translation, and the Precision Medicine Frontier—A Narrative Review. *Journal of personalized medicine*, 16(4), 218.
- [5] Chakraborty, P., & Kamila, S. (2026). Transforming Healthcare with Electronic Health Records: AI Integration, Evolution, and Future. In *AI in Smart and Secure Healthcare: Research Trends and Future Opportunities* (pp. 257-275). Cham: Springer Nature Switzerland.
- [6] Boulanger, Pierre. "MedROAD V2: An AI-Integrated Electronic Medical Record System with Advanced Clinical Decision Support." *AI in Medicine* 1.1 (2026): 4.
- [7] Manoharan, J., & Sehgal, Y. (2026). AI-Enabled Text Mining: A Paradigm Shift in Disease Prediction, Drug Discovery, and Clinical Research. *Journal of Medico Informatics*, 2(02), 23-35.
- [8] De Paoli, F., Nicora, G., Galanti, K., Gonzales, C. S., Mangili, F., Smedley, D., ... & Chahal, C. A. A. (2026). Artificial Intelligence and Machine Learning Approaches for Cardiovascular Genomics: A State-of-the-Art Review. *Circulation: Genomic and Precision Medicine*, 19(4), e005447.
- [9] Zeng, J., Xu, Y., Chen, P., Hong, M., & Zhang, Q. (2026). Effectiveness of natural language intelligence technology in chronic diseases nursing: A systematic review and meta-analysis. *International Journal of Nursing Studies*, 105394.
- [10] Croon, P. M., Dhingra, L. S., Pedroso, A. F., & Khera, R. (2026). The Evolving Utility of Artificial Intelligence-Based Tools for the Detection of Heart Failure and Cardiomyopathies: From Potential to Implementation. *Current Heart Failure Reports*, 23(1), 25.
- [11] Hornback, A., Sathu, H., Kim, K., Wang, Y., Zhu, Y., Isgut, M., ... & Wang, M. D. (2026). Large language models in healthcare and biomedical informatics: A comprehensive review. *Innovation and Emerging Technologies*, 13, 2630001.
- [12] Rasoli, R., Ebrahimisadrabadi, F., Khedri, Z., & Sohrabei, S. (2026). Designing conversational intelligence: effect of large language models (GPT-driven) platforms for

- precision maternal and newborn health engagement: a systematic review. Oxford Open Digital Health, oqag001.
- [13] Shang, L., Chen, Y., Li, R., Zhang, X., Gao, M., Hou, Y., & Zhang, G. (2026). Exploring the prognostic utility of large language models versus traditional clinical models in heart failure: a pilot study. *International Journal of Surgery*, 112(3), 5778-5788.
 - [14] Pinn, C. K., Dahil, A., Keast, J., & Dambha-Miller, H. (2026). Artificial Intelligence-Integrated Digital Tools to Promote Physical Activity in People with Multimorbidity: A Rapid Review of Trials.
 - [15] Wang, M. H. (2026). Artificial intelligence across the obesity continuum: from mechanistic insights to global precision prevention and therapy. *Obesity*, 34(2), 294-316.
 - [16] Saha, H. N., Bhattacharya, D. C., Dutta, S., Bera, A., Basuray, S., Changdar, S., ... & Turdiev, J. (2025). Transforming healthcare with state-of-the-art medical-llms: A comprehensive evaluation of current advances using benchmarking framework. *Computers, Materials & Continua*, 86(2), 1-56.
 - [17] Dhavale, S. R., Hiremath, S. S., Khole, S. R., Lobo, L. M. R. J. R., Kakade, S. P., & Patil, S. S. (2026). Generative AI and Large Language Models (LLMs) for Personalized and Explainable Healthcare. In *AI Foundations, Technologies, and Future of Healthcare Systems* (pp. 75-112). IGI Global Scientific Publishing.
 - [18] Bhardwaj, A., Ali, M. M., Tayyab, M. S. U., Taher, M. W. U., & Yadala, P. K. (2026, January). Healthcare AI: An Integrated AI Platform for Proactive Patient Care and Medical Data Management. In *2026 IEEE International Conference on Emerging Computing and Intelligent Technologies (ICoECIT)* (pp. 1-5). IEEE.
 - [19] Li, E. (2026, January). A systematic review of traditional and deep learning algorithms in intelligent healthcare: applications, complementarity, and future prospects. In *Second International Conference on Communication, Information, and Digital Technologies (CIDT 2025)* (Vol. 14064, pp. 912-921). SPIE.
 - [20] Tiwari, A., Shah, P. C., Kumar, H., Borse, T., Arun, A. R., Chekragari, M., ... & Mylavarapu, M. (2025). The Role of Artificial Intelligence in Cardiovascular Disease Risk Prediction: An Updated Review on Current Understanding and Future Research. *Current Cardiology Reviews*, 21(6), e1573403X351048.