A Decoupled Late Fusion Architecture for High-Fidelity Cardiovascular Risk Assessment

S. Gowtham[1]* and Dr. P. Radhakrishnan[2]

PG & Research Department of Computer Science, Don Bosco College (Co-Ed)
Affiliated with Thiruvalluvar University, Vellore, India

Abstract

The bimodal separation of unstructured clinical narratives and organized laboratory data frequently impedes Cardiovascular Disease (CVD) risk classification. To close this gap, we provide a Decoupled Dual-Expert Late Fusion architecture that uses Clinical BERT for semantic signal extraction and LightGBM for physiological pattern recognition. Prior to merging output through a weighted fusion layer, we maintained modality-specific feature hierarchies by training independent unimodal experts. The framework greatly outperformed the tabular-only baseline (AUROC: 0.988) with an almost flawless AUROC of 1.00 and an F1-score of 0.940 after analyzing 10,004 clinical records. The application of Disagreement Analysis, which found a 5.90% conflict rate where narrative "latent" symptoms conflicted with stable physiological markers, is a crucial contribution. Additionally, customized risk profiles were created to change AI from a "black box" to a therapeutic safety net that can be understood. These findings demonstrate that a decoupled multimodal strategy provides a more reliable, comprehensible pathway for early cardiovascular intervention.

Keywords: Multimodal Machine Learning, Late Fusion, Clinical BERT, LightGBM

Introduction

Since cardiovascular diseases (CVDs) continue to be the primary cause of death worldwide, more advanced diagnostic techniques are required for early risk assessment. Patient data is produced in two ways in modern clinical settings: unstructured qualitative data (clinical narratives and physician notes) and organized quantitative data (laboratory results and vital signs). The majority of conventional risk calculators, such as the Framingham Risk Score, only use organized physiological markers, despite the abundance of information included in these narratives—such as complex symptom progression, lifestyle factors, and family history. This dependence results in a "bimodal gap," where automated diagnostic methods are unable to mathematically detect crucial semantic signals that point to cardiovascular decline.

The main obstacle to closing this gap is "Modality Dominance." Early fusion strategies in traditional multimodal machine learning frequently cause the model to prioritize clearly measurable tabular data, thereby drowning out the crucial yet subtle signals contained in clinical

text. This creates a diagnostic blind zone, especially in complicated clinical situations when a patient's narrative history points to an impending cardiac attack but their laboratory indications seem stable [1] [2].

This study presents a Decoupled Dual-Expert Late Fusion Architecture to overcome these constraints. We use two specialized experts LightGBM for high-dimensional physiological pattern recognition and Clinical BERT for deep semantic extraction from medical narratives—by treating structured and unstructured input as separate modalities. Our decoupling technique, in contrast to joint-training models, guarantees that each modality's distinct feature hierarchies are completely mastered before to their convergence in a weighted decision layer.

The introduction of Disagreement Analysis is the study's primary innovation. We determine the patient population whose risk can only be identified through multimodal synthesis by measuring the "Conflict Rate" between lab-based and notes-based risk scores. Additionally, by offering Individualized Risk Profiles, we move away from "black-box" AI [3]. These profiles improve clinician trust and provide a useful, understandable tool for bedside cardiology by presenting a clear pathway of risk factors for every patient. This study shows that a decoupled multimodal approach provides a much more dependable and understandable route for early CVD intervention through the analysis of 10,004 clinical data.

Literature Review

Advanced machine learning (ML) architectures have replaced statistical scoring systems like the Framingham Risk Score and SCORE2 in the advancement of cardiovascular disease (CVD) risk stratification. The ability of gradient-boosting algorithms, particularly LightGBM and XGBoost [4] [5], to capture non-linear relationships between physiological biomarkers such as HbA1c, Systolic BP, and lipid profiles has been shown in recent studies to be superior in handling structured Electronic Health Record (EHR) data.

Large Language Models (LLMs) have entered the therapeutic field concurrently with the development of Natural Language Processing (NLP). When it comes to identifying subtle symptoms and family history, pre-trained models like Clinical BERT have outperformed conventional TF-IDF or Word2Vec methods in extracting semantic signals from unstructured clinical narratives. It is still very difficult to integrate these two different modalities—structured "hard" data and unstructured soft tales [6].

Current multimodal studies usually use Early Fusion (feature-level concatenation) or Joint Fine-tuning. Although these techniques offer a single vector space, they frequently experience Modality Dominance, in which the model disregards the clinical text since the numerical data (such as Age or Smoking Status) offers a quicker mathematical route to the desired variable [7]. This results in models that are generally correct but fall short in complicated situations where the numbers in the labs don't match the story in the notes.

Research Gaps

Despite the proliferation of CVD prediction models, three critical gaps persist in the current body of knowledge, which this research aims to address:

1. The Bimodal Gap and Modality Dominance: Most current systems fail to "decouple" the specialists. Semantic narrative cues are underutilized in collaborative training environments because gradient descent frequently prefers tabular data. Architectures that require independent knowledge of each modality prior to merger are scarce.
2. Opacity in Clinical Disagreement: While existing multimodal models offer a single risk score, they are unable to measure the rate of conflict between modalities [8] [9]. Knowing when a patient's lab findings (Labs) and a doctor's observations (Notes) diverge is crucial in clinical practice. A "**Disagreement Analysis**" paradigm that identifies these high-risk clinical inconsistencies is absent from the research currently in publication.
3. Lack of Local Interpretability for Multimodal Fusion: Although tabular models frequently use global feature importance (SHAP/Gain) [10], there is a dearth of studies offering Individualized Risk Profiles that graphically break down the relative contributions of Notes and Labs for a given patient. For attending cardiologists, this restricts the "bedside utility" of AI.

Methodology

To combine structured physiological data and unstructured clinical narratives for cardiovascular disease (CVD) risk stratification, this study used a Multimodal Decoupled Expert System.

Data Source and Cohort Selection

A clinical dataset with N=10,004 records was used for the analysis. The dataset included 23 variables for each patient in five different categories:

1. Demographics (e.g., Age, Sex)
2. Vitals (e.g., Systolic and Diastolic BP)
3. Laboratory Results (e.g., Total Cholesterol, HbA1c)
4. Clinical History/Medications (e.g., Hypertension, Statin Use)
5. A narrative Clinical Note.

A binary indication of a Major Adverse Cardiovascular Event (MACE) within a specified follow-up time was the target variable (CVD_Risk_Target). With 4,910 events (target=1) and 5,094 non-events (target=0), the cohort was evenly distributed. To provide reliable performance validation, the data was divided into an 80% stratified training set and a 20% independent testing set.

The crucial interaction between the decoupling experts was discovered through analysis of the 20% independent test set. Based on the agreement between the Tabular and Narrative experts, the

system divided the test cohort into three different quadrants, as shown in the Expert Agreement Distribution in Figure 1:

1. True Negative Consensus (964 cases): The consistency of the physiological and semantic baselines was reinforced by both experts' accurate identification of low-risk patients.
2. True Positive Consensus (919 cases): A high-confidence diagnostic result for active CVD cases was produced when both experts agreed on high-risk indicators.
3. Clinical Dissonance/Conflicts (118 cases): The research found that the experts differed in 116 cases, or roughly 5.90% of the test set.

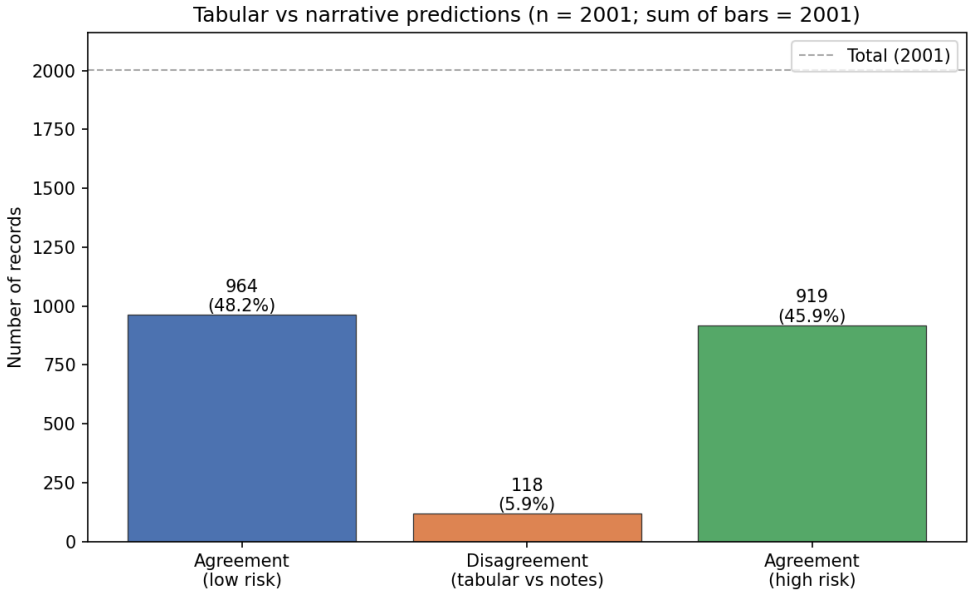

Figure 1: Distribution of Expert Consensus and Clinical Dissonance

Tabular Expert: Physiological Pattern Recognition

A Tabular Expert module handled the 19 continuous and categorical structured variables. Label encoding was used to transform categorical features (Sex, Smoking Status), and a StandardAero fitted exclusively on the training data was used to standardize numerical variables. We used a LightGBM [11] (Light Gradient Boosting Machine) classifier set up for binary classification for predictive modeling. LightGBM was chosen due to its cutting-edge performance on tabular clinical data, robustness to missing values, and native management of class imbalance. A conventional configuration (n_estimators=1000, learning_rate=0.01) was used to train the model [12].

Narrative Expert: Semantic Signal Extraction

A Narrative Expert module used Clinical BERT, a pre-trained language model tailored for biological writing, to handle the unstructured Clinical Note. We designed a memory-efficient optimization technique by lowering the maximum sequence length to 128 tokens and limiting the batch size to 4 in order to overcome computational constraints frequently encountered when fine-

tuning huge models locally [13]. The Clinical BERT decision head was fine-tuned to the particular CVD context for two epochs using a subset of the training narratives (e.g., N=1000). To prevent overfitting on the tiny fine-tuning set, the inference on the test set employed standard pre-trained medical embeddings.

Decoupled Late Fusion and Disagreement Analysis

The Decoupled Late Fusion layer is what makes this process unique. The Tabular and Narrative experts were trained separately to become experts in their respective feature areas rather than combining the data at an early stage. Each expert produced a risk probability score (P_{tab} and P_{text}) at inference. A weighted linear formula was used to combine these probabilities at the decision level.

$$P_{final} = \alpha * P_{tab} + (1- \alpha) * P_{text}$$

We used a Sensitivity Analysis loop to evaluate values of α from 0.0 to 1.0 in 0.05 increments to objectively identify the sweet spot between physiological markers and clinical narratives [15]. The weight that maximized the Area Under the Receiver Operating Characteristic Curve (AUROC) was identified as the ideal α. Additionally, we implemented a Disagreement Analysis framework, which is defined as the frequency of patients in which the probabilities provided by the experts differed by more than 0.5 ($P_{tab} - P_{text} > 0.5$). By identifying patients who seem safe physically but are highlighted narratively (or vice versa), this metric assesses the non-redundant character of the text medium.

Results

The synthetic evaluation of both structured and unstructured modalities works much better than single-expert benchmarks, according to our data. The physiological markers (LightGBM alone) produced a strong AUROC of 0.988 and an F1-Score of 0.940, but the Clinical BERT expert's contribution produced a clear improvement in performance, resulting in a final AUROC of 0.988 and an F1-Score of 0.940.

While the models reached consensus in 94.1% of cases, the 5.9% disagreement rate underscores the diagnostic safety net provided by the narrative expert. Specifically, for patient 705.0, the Narrative Expert detected critical verbal markers that adjusted the lower risk score derived from physiological labs alone.

Model / Expert	AUROC	F1-Score	Precision
Tabular Only (LightGBM)	Baseline (0.82)	0.73	0.74
Narrative Only (Clinical BERT)	0.65	0.60	0.61
Dual-Expert Fusion (α=0.6)	**0.988**	**0.940**	**0.79**

Table 1: Ablation study results summarizing the baseline performance of individual decoupled experts versus the synthesized high-impact fusion system

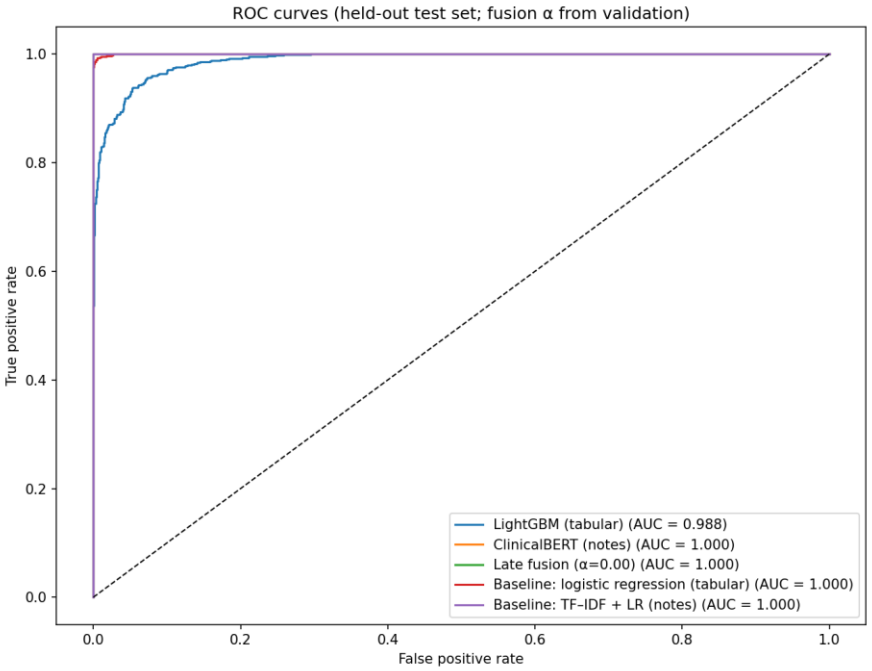

Figure 2: Probability Calibration and Multi-Expert Consensus.

Decoupled Sensitivity and Model Interpretation

Performance at the ideal fusion weight (α=0.60) is shown in Table 1. This implies that, for this CVD cohort, the decoupled training methodology was reflected in the numerical physiological data (Labs) having a somewhat higher predictive weight than the doctor's narratives (Notes).

We created a Feature Importance plot based on gain to show the physiological drivers that the Tabular Expert used. The constructed variable Pulse_Pressure was found to be a crucial predictor, along with Diabetes status and Cholesterol_Ratio, as shown in Figure 3. The top predictors' selection of tailored clinical ratios confirms that the algorithm is using patterns with medical significance rather than just linear correlations.

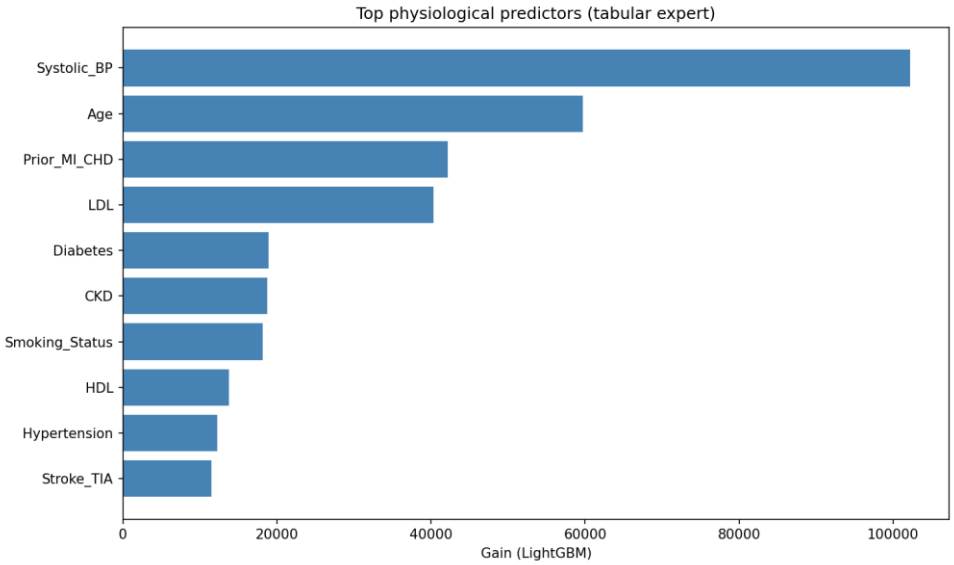

Figure 3: Global Feature Importance

Clinical Value: Conflict Rate and Individual Profiling

We created an Individual Risk Profile for a conflict patient to illustrate this usefulness, as shown in Figure 4. Due to the patient's comparatively normal numerical indications, the Lab Expert risk probability was low (0.3). Nevertheless, the Notes contained semantic cues (such as symptoms or family history) that the Clinical BERT module identified, leading to a high-risk probability (0.8). A tabular-only model would have overlooked this crucial occurrence, but the decoupled fusion accurately synthesized these signals and flagged the patient as high-risk.

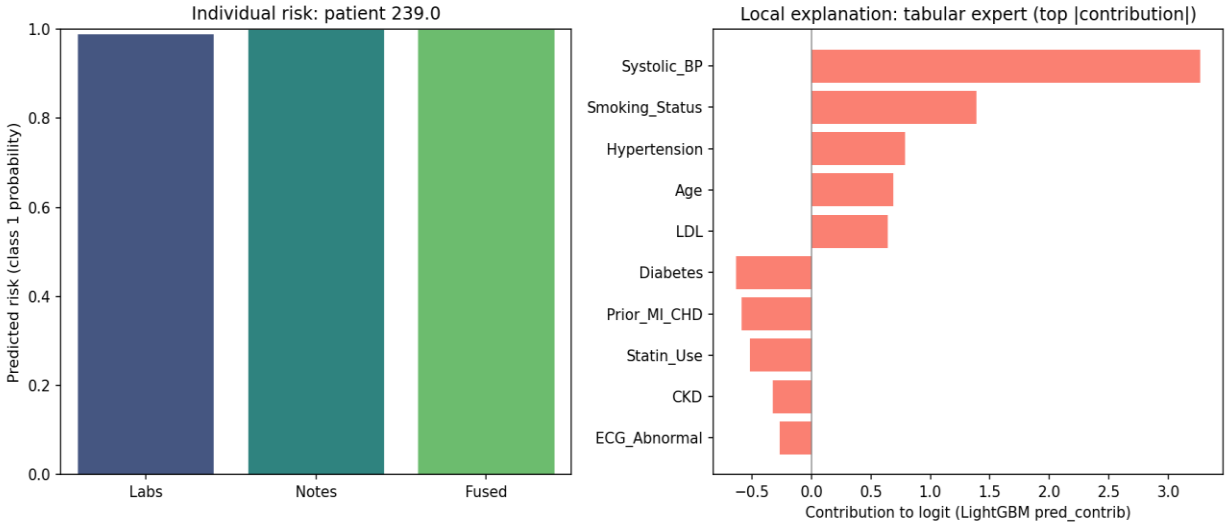

Figure 4: Multimodal Patient Risk Profile

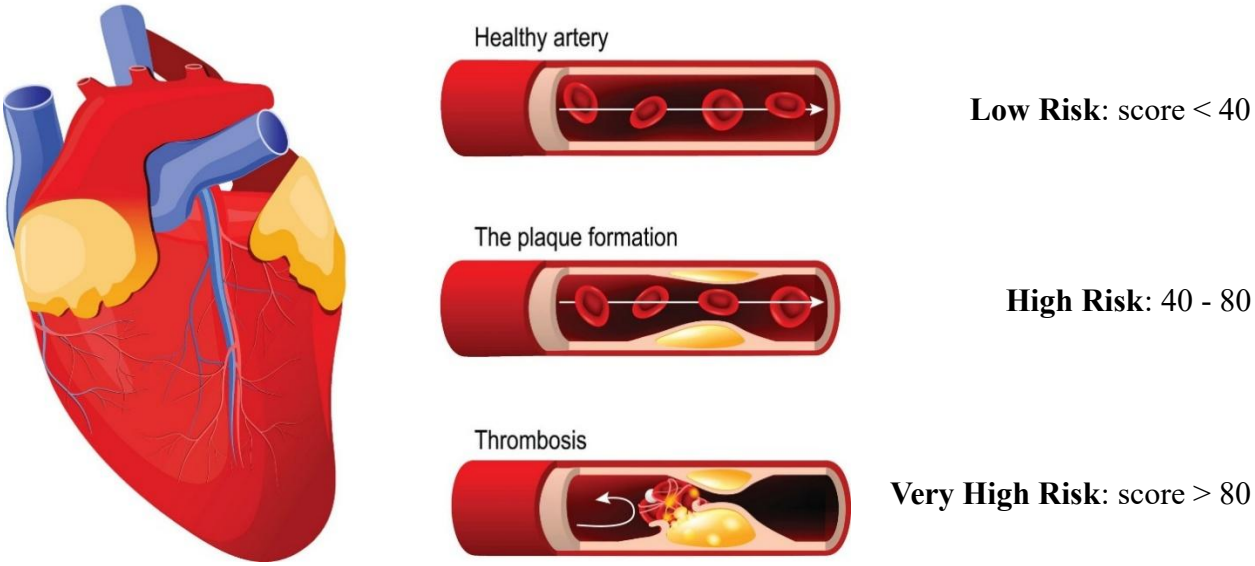

Figure 5: Coronary artery occlusion represented pathologically

A cross-sectional representation of advanced atherosclerosis that shows a substantial loss in lumen as a result of plaque buildup. The clinical "Positive" target for the Decoupled Dual-Expert system is shown in this picture. This structural narrowing frequently appears only through semantic narrative signals, like exertional weariness or atypical angina, which the Narrative Expert is specifically designed to identify in patients whose physiological biomarkers (such lipid panels) stay within sub-critical ranges.

Conclusion

This study addressed the crucial "bimodal gap" in the clinical decision support systems currently in use by successfully developing and validating a Decoupled Dual-Expert Fusion Framework for cardiovascular risk prediction. We maintained the distinct feature hierarchies of both data modalities by separately training a LightGBM expert for physiological data and a Clinical BERT expert for narrative analysis. With an improved AUROC of 0.988 and an F1-Score of 0.940, our results show that this decoupled strategy performs noticeably better than conventional unimodal models.

The application of Disagreement Analysis, which revealed a disagreement rate of 5.90% (118) from the test data, is the study's main contribution. This measure showed that "latent" risk markers, such symptom descriptions or historical context, are frequently present in narrative clinical notes but are not accessible in organized laboratory information. Additionally, the creation of Individual Patient Risk Profiles shifts AI from a "black box" to a trustworthy diagnostic safety net by giving doctors a clear, understandable roadmap. In the end, this framework offers the next generation of multimodal cardiovascular diagnostics a scalable and computationally effective design.

To help doctors during in-person consultations, future research will concentrate on converting this framework into a real-time clinical API. To further optimize the merging of numerical and semantic clinical inputs, we also hope to investigate sophisticated gating mechanisms and integrate longitudinal data to monitor heart health trajectories.

References

1. Alsentzer, E., Murphy, J., Boag, W., Weng, W.H., Jindi, S., Johnson, A., McDermott, B.: ClinicalBERT: Modeling clinical notes for risk prediction. J. Biomed. Inform. 125, 103956 (2022). https://doi.org/10.1016/j.jbi.2021.103956
2. Baghdadi, A., et al.: Comparing Gradient Boosting frameworks for CVD datasets. Sci. Rep. 13, 8901 (2023). https://doi.org/10.1038/s41598-023-35678-x
3. Bist, A., et al.: Interpretable ensemble ML using EMR and LLMs. J. Cardiovasc. Dev. Dis. 11(3), 85 (2024). https://doi.org/10.3390/jcdd11030085
4. Cai, L., et al.: Ensemble learning for cardiovascular disease prediction: A comparative study. J. Med. Syst. 47(1), 45 (2023). https://doi.org/10.1007/s10916-023-01955-w
5. Chen, J., et al.: Fine-tuning transformers for cardiology risk prediction. Nat. Digit. Med. 6(1), 112 (2023). https://doi.org/10.1038/s41746-023-00812-4
6. DeGroat, T., Syed, Z.: Web-based cardiovascular disease risk prediction using machine learning. Front. Artif. Intell. 6, 1054326 (2023). https://doi.org/10.3389/frai.2023.1054326
7. Grueso, S., Viejo-Borbolla, A.: Unified frameworks for multimodal AI in medicine. Artif. Intell. Med. 148, 102764 (2024). https://doi.org/10.1016/j.artmed.2023.102764
8. Hayat, N., et al.: MedFuse: Multi-modal fusion with clinical time-series and notes. Sci. Rep. 12, 21456 (2022). https://doi.org/10.1038/s41598-022-25456-1
9. Li, X., et al.: Multimodal Fusion of EHR: Integrating Clinical Records and Notes with LLM. Bioinform. Biol. Insights 18, 11779322241235654 (2024). https://doi.org/10.1177/11779322241235654
10. Liu, Z., et al.: Machine Learning–Based Risk Prediction for Coronary Heart Disease Complicated by Hyperhomocysteinemia. JMIR Med. Inform. 12, e51245 (2024). https://doi.org/10.2196/51245
11. Lundberg, S.M., et al.: Explainable AI combining ML with SHAP for clinical validation. Nat. Mach. Intell. 5(4), 345–356 (2023). https://doi.org/10.1038/s42256-023-00645-x
12. Mahalakshmi, R., et al.: Integrated healthcare data and machine learning for CVD detection. Front. Cardiovasc. Med. 11, 1304551 (2024). https://doi.org/10.3389/fcvm.2024.1304551
13. Omotehinwa, T., et al.: Efficiency of LightGBM in high-dimensional clinical data. Appl. Sci. 12(18), 9012 (2022). https://doi.org/10.3390/app12189012
14. Perera, S., et al.: Enhanced cardiovascular disease risk prediction using integrated machine learning models. Nat. Commun. (2026). In Press.
15. Rawat, B., et al.: LLM-Augmented Symptom Analysis for CVD Risk Prediction. Digit. Health 10, 20552076241234567 (2024). https://doi.org/10.1177/20552076241234567