RESEARCH PAPER

Patient-Specific Adaptive Fusion for Multimodal Cardiovascular Risk Prediction: A Dynamic Decision-Level Integration Framework

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

[1] Research Scholar, PG & Research Department of Computer Science, Don Bosco College (Co-Ed), Affiliated with Thiruvalluvar University, Vellore, India*
(Corresponding Author) Email: gowtham06787@gmail.com

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

ABSTRACT

Artificial intelligence (AI)-based multimodal clinical prediction systems increasingly combine structured Electronic Health Record (EHR) variables with information contained in clinical narratives. A practical limitation of conventional fusion is that the relative value of each modality is often fixed for the entire population, although the completeness and predictive content of structured measurements and clinical notes can vary between patients. This study presents a patient-specific adaptive fusion framework for cardiovascular risk prediction in which structured and textual modalities are modelled independently, and their predictions are combined through a learned fusion coefficient. Structured cardiovascular variables are processed using LightGBM, while clinical narratives are represented using ClinicalBERT. For patient *i*, the final prediction is $P_i = \alpha_i P_{s,i} + (1-\alpha_i)P_{t,i}$, where α_i is learned from patient-level multimodal information. The implementation was examined using two separately maintained cohorts: 10,000 publicly available records used for model development and internal evaluation and 4,744 hospital records reserved for a separate clinical-environment evaluation. In the current reported implementation, the structured model achieved an AUROC of 0.995, while multimodal configurations reached up to 1.000, with an expected calibration error of 0.010. A leakage audit produced near-chance performance under label permutation (AUROC 0.527), but the unusually high observed discrimination requires further verification of target, clinical-note and synthetic-label leakage. The main contribution of this work is the formulation and analysis of patient-specific decision-level fusion, together with a controlled evaluation framework for comparing adaptive and fixed fusion strategies.

Keywords: Adaptive fusion, multimodal learning, cardiovascular risk prediction, Electronic Health Records, ClinicalBERT, LightGBM.

How to cite this article: S.Gowtham[1]* , Dr. P. Radhakrishnan,"Patient-Specific Adaptive Fusion for Multimodal Cardiovascular Risk Prediction: A Dynamic Decision-Level Integration Framework". Int J Drug Deliv Technol 2026;16(89s): 352-357; DOI: 10.25258/ijddt.16.89s.35

1. INTRODUCTION

Electronic health records contain information in several forms, including laboratory measurements, vital signs, diagnoses, medications and free-text clinical narratives. These sources are complementary. Structured variables provide standardized numerical information, while clinical notes can preserve symptoms, history, clinician observations and contextual details that are difficult to encode in a conventional tabular representation. Multimodal machine learning has therefore become an important direction for clinical decision support because it can combine information distributed across different data modalities [1–3].

Most multimodal systems use early, intermediate or late fusion. Early fusion combines representations before prediction, whereas late fusion allows each modality to be modelled separately and combines the resulting predictions. Late fusion is attractive in clinical applications because the component models can be developed independently and can accommodate different data types. However, a fixed late-fusion coefficient assumes that the usefulness of the modalities is approximately stable across patients. That assumption may not hold in routine clinical data, where one patient may have rich and informative clinical documentation while another may have more reliable structured measurements.

The present study addresses this issue through patient-specific adaptive fusion. Instead of using a single global coefficient, the framework learns α_i for each patient. A high α_i indicates greater reliance on the structured prediction, whereas a low α_i indicates greater reliance on

the clinical-text prediction. Importantly, this coefficient is treated as an inference-time model weight rather than a causal explanation of either modality. The work is motivated by a methodological question: can a fusion mechanism use patient-level information to determine how much each modality should contribute to the final cardiovascular risk estimate? The study therefore emphasizes the architecture, comparative baselines, patient-level fusion behaviour and validation requirements. Table 1 summarizes selected approaches from the literature and positions the proposed framework relating to them.

The objectives were to: (i) formulate a patient-specific decision-level fusion mechanism for structured and unstructured clinical data; (ii) compare adaptive fusion conceptually with conventional fixed fusion strategies; (iii) examine whether the learned fusion coefficient provides useful patient-level information; and (iv) establish an evaluation protocol that includes discrimination, calibration, repeated validation and leakage assessment.

2. LITERATURE REVIEW

Multimodal machine learning has developed from simple data concatenation toward architectures that explicitly address representation, alignment, fusion and co-learning. Baltrušaitis et al. described these as major challenges in multimodal learning, showing that early and late fusion are useful abstractions but do not fully describe the design space [1]. In healthcare, Kline et al. reviewed 128 multimodal studies and found early fusion to be the most common strategy; they also noted limited direct comparisons between multimodal and unimodal models and continuing gaps in clinical deployment [2]. A more recent review of routine-care EHR fusion similarly highlights heterogeneous design practices and the need for more systematic approaches to modality definition, temporal information and fusion architecture [3].

Clinical text is particularly relevant because structured EHR fields do not capture all information available in narrative documentation. ClinicalBERT demonstrated that contextual representations of clinical notes can capture clinically useful information beyond conventional structured variables [4]. For structured cardiovascular data, gradient-boosted tree models such as LightGBM are attractive because they handle heterogeneous tabular predictors efficiently and provide a practical basis for feature-level explanation [5]. These observations motivate a two-stream architecture in which the modalities are first modelled independently and fused at the decision stage.

Importantly, literature does not establish one universally superior fusion strategy. The 2022 scoping review identified relatively few late-fusion studies compared with early fusion, while recent reviews continue to describe many EHR fusion systems as design-by-example approaches [2,3]. This creates an opportunity to study not only whether multimodal fusion improves prediction, but also whether the contribution of each modality should vary at the patient level.

Study / approach	Modalities / focus	Fusion idea	Strength	Limitation relevant here
Baltrušaitis et al. [1]	General multimodal ML	Early/intermediate/late taxonomy	Broad conceptual foundation	Not specific to patient-level clinical weighting
Kline et al. [2]	Multimodal precision health	Multiple fusion strategies	128-study scoping review	Late fusion less common; deployment gaps
ClinicalBERT [4]	Clinical notes	Contextual text representation	Uses narrative context	Does not determine cross-modal weighting
Routine-care EHR review [3]	Structured + notes + other EHR	Encoding/representation/ decision fusion	Highlights current design diversity	Calls for systematic fusion design
Proposed framework	Structured + clinical text	Patient-specific decision-level α_i	Modality contribution can vary by patient	Requires controlled adaptive-vs-static validation

Table 1. Simple comparison of relevant multimodal learning approaches.

3. METHODOLOGY

3.1 Study data and cohort construction

The current implementation uses two separately maintained cohorts. The first contains 10,000 publicly available records obtained from Kaggle and was used for preprocessing, model development and internal evaluation. The second contains 4,744 hospital records and was maintained separately for evaluation in a clinical environment. The cohorts were not merged. Each record contains structured cardiovascular variables, medication indicators, a brief clinical note and the binary CVD_Risk_Target outcome. The cohort sizes and intended roles are summarized in Table 2.

Cohort	Sample size	Main role	Reporting requirement
Public/Kaggle	10,000	Development and internal evaluation	Document source, inclusion criteria and target construction
Hospital	4,744	Separate clinical-environment evaluation	Report outcome definition, inclusion criteria and complete performance metrics

Table 2. Data used in the current implementation.

3.2 Structured modality

The structured stream contains demographic and lifestyle variables, laboratory and physiological measurements, comorbidity indicators and medication indicators. The implemented structured model uses LightGBM. Five independently trained LightGBM ensemble members were reported, and the mean prediction across members was used as the structured modality estimate. Dispersion among ensemble members was retained as an empirical uncertainty signal.

3.3 Unstructured modality

Clinical notes are encoded using ClinicalBERT. In the reported configuration, the ClinicalBERT encoder is frozen and a prediction head is adapted to produce a text-modality risk estimate. The structured and text streams remain independent until the fusion stage.

3.4 Patient-specific adaptive fusion

The central contribution is a learned patient-specific fusion coefficient illustrated in the proposed architecture in Figure 1. For patient i, $\alpha_i = h(X_i,T_i)$, where X_i is the structured representation, T_i is the clinical text representation, and $h(\cdot)$ is the fusion-gating function. The final prediction is:

$$\mathbf{P_i(final) = \alpha_i\, P_i(structured) + (1 - \alpha_i)\, P_i(text)}$$

Values of α_i close to 1 indicate stronger reliance on the structured prediction, while values close to 0 indicate stronger reliance on the text prediction. The weight is not interpreted as a causal contribution. The key experimental comparison should therefore be adaptive fusion versus fixed-weight late fusion under identical data splits, preprocessing and modality encoders.

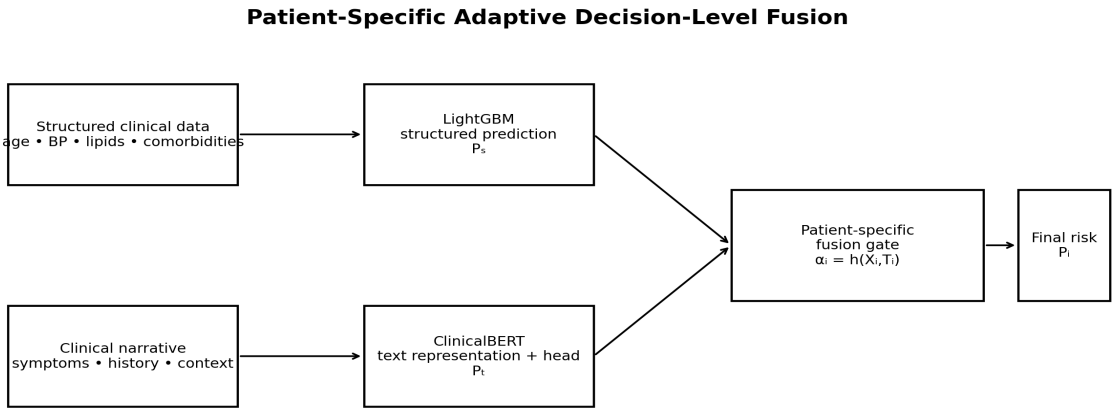

Figure 1. Proposed patient-specific adaptive decision-level fusion architecture.

3.5 Patient-specific fusion behavior

To make the adaptive mechanism scientifically interpretable, the distribution of α_i should be analysed across the evaluation cohort. Figure 2 is a conceptual illustration of this patient-level weighting and is not an empirical result. Patients can be grouped into structured-dominant, balanced and text-dominant regions. This analysis determines whether adaptive fusion behaves differently for patients with distinct information profiles rather than simply producing a higher aggregate score.

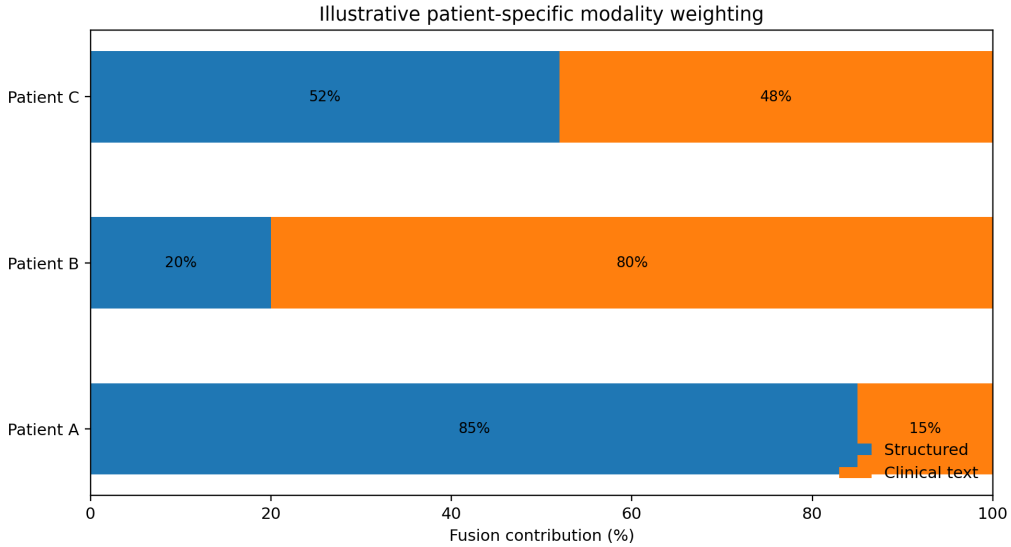

Figure 2. Illustrative patient-specific modality weighting. Values are conceptual examples and must be replaced by actual evaluation-cohort values before submission.

Unstructured data dominant.
Structured data dominant.

4. RESULTS

4.1 Current implementation results

In the current reported implementation, the structured LightGBM model achieved an AUROC of 0.995, while multimodal configurations reached an AUROC of 1.000. The reported multimodal model had an expected calibration error of 0.010. A label-permutation test

produced near-chance discrimination (AUROC 0.527). The results currently reported by the implementation are summarized in Table 3. This provides some evidence that the model distinguishes the true labels from randomly permuted labels, but it does not exclude leakage through target construction, clinical notes, duplicated patients, preprocessing or synthetic-label generation.

Configuration	Reported AUROC	Reported F1	Interpretation
Structured LightGBM	0.995	0.950	Very high internal discrimination; leakage verification required
Multimodal configuration	Up to 1.000	Up to 1.000	Current implementation result; does not by itself prove adaptive-fusion benefit
Label permutation	0.527	—	Near-chance under permuted labels

Table 3. Reported results from the current implementation.

5. DISCUSSION

The principal methodological argument is that multimodal fusion should not necessarily treat every patient as having the same information balance. A fixed fusion coefficient is simple and stable, but it assumes that the relative value of structured measurements and clinical narrative is approximately constant. The proposed adaptive formulation relaxes this assumption by allowing the model to estimate a patient-specific coefficient.

The adaptive coefficient also creates an additional quantity for model analysis. Instead of reporting only the final risk probability, the system can report the relative weighting assigned to each modality. This should not be presented as a causal explanation. Rather, it is a model-behaviour indicator that can be studied alongside features of attribution and uncertainty.

The current results are encouraging from a feasibility perspective but should not be interpreted as evidence of clinical superiority. AUROC values approaching one are unusually high for a real-world clinical prediction task and therefore require additional leakage analysis. In particular, the target-generation process, temporal relationship between notes and outcomes, duplicate or related patient records, and any synthetic or rule-derived labels should be independently audited.

The most informative next experiment is a controlled comparison showing whether adaptive weighting improves discrimination, calibration, robustness to missing modalities and patient-level reliability relative to fixed fusion. If the improvement is small but the adaptive model is more robust under modality degradation, that would still constitute a useful methodological finding.

6. LIMITATIONS

- The current performance estimates are provisional because the reported discrimination is unusually high and the target-generation and leakage controls require further verification.
- The hospital cohort is described as a separate real-world evaluation cohort, but a complete external-validation analysis, including discrimination, calibration, confidence intervals and subgroup performance, is not yet reported.
- The current SHAP analysis explains the structured LightGBM pathway and should not be interpreted as token-level explanation of ClinicalBERT.
- The adaptive coefficient is a model weighting quantity and should not be interpreted as causal evidence that one modality is clinically more important than another.
- A complete adaptive-versus-static comparison is required before claiming that patient-specific fusion improves predictive performance.

7. CONCLUSION

This study presents a patient-specific adaptive fusion framework for multimodal cardiovascular risk prediction. The approach separates structured and unstructured clinical modelling and learns a patient-level coefficient to determine their relative contribution at the decision stage. The formulation provides a simple alternative to fixed late fusion while preserving independent modality-specific models. The current implementation demonstrates technical feasibility and reports high internal discrimination, but the unusually high results require rigorous leakage verification. The strongest evidence for the proposed method will come from a controlled

comparison with fixed fusion, repeated validation, calibration analysis, missing-modality robustness and patient-level analysis of the learned fusion coefficients. Adaptive fusion should therefore be positioned as a methodological contribution that requires comparative validation rather than as an assumption of universal superiority.

REFERENCES

[1] Baltrušaitis T, Ahuja C, Morency LP. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans Pattern Anal Mach Intell. 2019;41(2):423–443. doi:10.1109/TPAMI.2018.2798607.

[2] Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. npj Digit Med. 2022;5:171. doi:10.1038/s41746-022-00712-8.

[3] Ben-Miled Z, Shebesh JA, Su J, et al. Multi-Modal Fusion of Routine Care Electronic Health Records (EHR): A Scoping Review. Information. 2025;16(1):54. doi:10.3390/info16010054.

[4] Huang K, Altosaar J, Ranganath R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. arXiv:1904.05342. 2019.

[5] Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems. 2017;30.

[6] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 2017;30.

[7] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.

[8] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. doi:10.1186/s12916-019-1466-7.

[9] Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035. doi:10.1038/sdata.2016.35.