Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

CLINICAI - Hybrid Machine Learning and Transformer Architecture for Clinical Diagnosis and Drug Recommendations

[1]Kamal Kishor Sahu, [2]Talish Rai, [3] Ms. Rashmi Banchhor
[1,2]B. Tech CSE Student, [3]Assistant Professor
[1,2,3]Computer Science Engineering (Artificial Intelligence and Machine Learning)
[1,2,3]Shri Shankracharya Institute of Professional Management and Technology Raipur, India
[1] kamalk@ssipmt.com, [2]talish@ssipmt.com, [3]rashmi.banchhor@ssipmt.com

Abstract: We introduce CLINICAI, a unified hybrid architecture that synergistically couples gradient-boosted ensemble machine learning with pretrained transformer-based deep learning to advance clinical decision support at scale. CLINICAI addresses two intertwined challenges in AI-assisted medicine: (1) accurate multi-label disease diagnosis from heterogeneous electronic health record (EHR) data comprising both structured laboratory values and unstructured clinical narratives, and (2) personalised, safety-constrained drug recommendations informed by pharmacological knowledge graphs encoding known drug-drug interactions (DDIs). A cross-modal attention fusion layer dynamically weights evidence across the two modalities, while an auxiliary pharmacological safety module penalises recommendations that would introduce clinically significant interactions. Evaluated on MIMIC-IV, the PhysioNet Sepsis Challenge 2019, and the DrugBank Interaction Corpus, CLINICAI achieves a diagnostic AUROC of 0.967, top-1 accuracy of 94.7%, and drug recommendation Precision@20 of 91.3%, reducing the DDI rate by 38% relative to the strongest single-modality baseline. A clinical explainability module produces SHAP-based feature attributions and natural-language justifications for every prediction, supporting clinician trust and regulatory compliance. A 12-week prospective shadow-mode feasibility study across 4,218 patient encounters at a tertiary care centre confirms near real-world deployability.

Keywords: clinical decision support, transformer architecture, gradient boosting drug recommendation, EHR, explainable AI, pharmacovigilance, cross-modal attention

1. INTRODUCTION

The intersection of artificial intelligence and clinical medicine has entered a transformative era. Electronic health record systems now accumulate hundreds of clinical variables per patient per day — laboratory panels, vital-sign trajectories, medication administrations, diagnostic imaging reports, physician notes, and genomic markers — far exceeding the cognitive bandwidth of any individual clinician. Clinical decision support systems (CDSS) powered by machine learning offer a compelling path toward addressing diagnostic delays, medication errors, and unjustified treatment variability, three leading causes of preventable morbidity and mortality worldwide.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

Despite impressive advances in both classical machine learning and deep learning for healthcare, existing systems typically suffer from one of two fundamental limitations. Models trained solely on structured EHR data achieve interpretable, calibrated predictions but fail to leverage the rich semantic content of clinical narratives, imaging reports, and temporal physiological signals. Conversely, large pretrained language models capture nuanced clinical semantics but lack the probabilistic calibration and tabular feature-processing efficiency required for safe, high-stakes recommendations. Drug recommendation compounds these difficulties further, requiring navigation of complex polypharmacy interaction landscapes, patient-specific contraindications, and continuously evolving pharmacological evidence.

CLINICAI resolves this tension. By fusing gradient-boosted ensemble methods — which excel on tabular EHR features — with a fine-tuned clinical transformer encoder — which excels on sequential and textual data — through a learned cross-modal attention fusion mechanism, CLINICAI achieves state-of-the-art performance across both tasks while maintaining full clinical explainability. This report presents the architectural design, training methodology, experimental evaluation, ablation analysis, and prospective clinical feasibility study for the CLINICAI framework.

The principal contributions of this work are as follows:

- A hybrid ML-transformer architecture that jointly processes structured EHR features and unstructured clinical text through a novel bidirectional cross-modal attention fusion layer.
- A pharmacological safety module integrating drug-drug interaction (DDI) knowledge graphs and patient contraindication embeddings into the drug recommendation pipeline.
- A clinical explainability module generating SHAP-based feature attributions and natural-language justifications for every prediction.
- Comprehensive benchmarking on three publicly available clinical datasets with results surpassing all current state-of-the-art baselines by statistically significant margins.
- A 12-week prospective feasibility study across 4,218 patient encounters at a 650-bed academic medical centre.

2. RELATED WORK

2.1 Machine Learning for Clinical Diagnosis

Early CDSS systems relied on logistic regression, naïve Bayes classifiers, and rule-based expert systems. The advent of ensemble learning, particularly random forests (Breiman, 2001) and gradient boosting (Friedman, 2001), substantially improved predictive performance on structured EHR data. Rajkomar et al. (2018) applied deep learning to longitudinal EHR sequences, achieving significant improvements on in-hospital mortality and 30-day readmission prediction. Gradient-boosted decision

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

tree models — notably XGBoost (Chen & Guestrin, 2016) and LightGBM (Ke et al., 2017) — subsequently became the de facto standard for structured clinical data owing to their robustness to missing values, computational efficiency, and competitive performance. Their inability to natively process raw text, imaging features, or fine-grained temporal sequences, however, motivated the development of hybrid approaches.

2.2 Transformer Models in Healthcare

The transformer architecture (Vaswani et al., 2017) and its derivatives — BERT (Devlin et al., 2018), its clinical variant ClinicalBERT (Alsentzer et al., 2019), and BioBERT (Lee et al., 2020) — demonstrated that domain-specific pretraining substantially improved performance on downstream clinical NLP tasks. Med-PaLM 2 (Singhal et al., 2023) subsequently illustrated that large language models could approach expert-level performance on medical licensing examinations. In the clinical time-series domain, temporal transformers have been applied to ICU prediction, sepsis detection, and length-of-stay forecasting, confirming the architecture's capacity to model long-range temporal dependencies that challenge recurrent networks.

2.3 Drug Recommendation Systems

Drug recommendation in clinical AI evolved from collaborative-filtering approaches adapted from recommender systems to graph neural networks that explicitly model pharmacological knowledge bases. SafeDrug (Yang et al., 2021) integrated molecular fingerprints and DDI constraints into a graph convolutional framework, while GAMENet (Shang et al., 2019) introduced memory-augmented networks retaining longitudinal medication history. Despite these advances, no prior work has simultaneously integrated structured EHR features, unstructured clinical notes, and pharmacological graph constraints within a single trainable architecture — the gap CLINICAI fills.

2.4 Explainability in Clinical AI

Regulatory frameworks including the EU AI Act (2024) and the FDA's guidance on Software as a Medical Device increasingly require interpretable justifications for clinical AI predictions. SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) have been widely adopted for post-hoc local explanations. Concerns about the faithfulness of raw attention weights as explanations (Jain & Wallace, 2019; Wiegreffe & Pinter, 2019) motivated the development of Integrated Gradients (Sundararajan et al., 2017), which CLINICAI adopts for transformer attribution alongside TreeSHAP for the gradient-boosted component.

3. THE CLINICAI ARCHITECTURE

CLINICAI is a four-module hybrid architecture (Figure 1): the Structured EHR Encoder, the Clinical Transformer Encoder, the Cross-Modal Attention Fusion Layer, and the Prediction & Safety Head.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

Table 1 — CLINICAI module overview.

Module	Technology	Output
Structured EHR Encoder	LightGBM (1,000 trees) + Leaf Embedding	h_struct ∈ $\mathbb{R}^{256}$
Clinical Transformer Encoder	ClinicalBERT-Large (12L × 16H × 1024d)	h_text ∈ $\mathbb{R}^{256}$ (projected)
Cross-Modal Attention Fusion	Bidirectional cross-attention + FFN + LayerNorm	h_patient ∈ $\mathbb{R}^{256}$
Prediction & Safety Head	Diagnostic softmax + Drug R-GCN + DDI penalty	Diagnosis dist. (286 classes) + Drug ranking

3.1 Structured EHR Encoder

The Structured EHR Encoder processes tabular patient data including demographics, laboratory values (CBC, metabolic panel, coagulation studies), vital-sign time-series statistics (mean, variance, trend over 24-hour windows), and coded diagnoses/procedures (ICD-10 and CPT codes mapped via entity lookup tables). It is implemented as a LightGBM ensemble with 1,000 trees, maximum 63 leaf nodes, and learning rate 0.05, using native missing-value propagation. Leaf embeddings from the penultimate layer are extracted to produce a 256-dimensional feature vector h_struct ∈ $\mathbb{R}^{256}$, converting the ensemble from a pure classifier into a differentiable feature extractor amenable to gradient-based fusion training.

3.2 Clinical Transformer Encoder

The Clinical Transformer Encoder processes discharge summaries, clinical notes, imaging reports, and medication history token sequences. Initialised from ClinicalBERT-Large (12 layers, 16 attention heads, hidden size 1,024), it is fine-tuned with layer-wise learning rate decay (top: 2×10^{-5}; bottom: 5×10^{-6}) and gradient clipping at norm 1.0. A WordPiece vocabulary of 30,522 base tokens is augmented with 2,847 domain-specific medical tokens. Sequences are truncated at 512 tokens with stride-based sliding-window aggregation for longer documents. The [CLS] representation is projected from $\mathbb{R}^{1024}$ to $\mathbb{R}^{256}$ via a learned linear layer, yielding h_text ∈ $\mathbb{R}^{256}$.

3.3 Cross-Modal Attention Fusion

The fusion module integrates h_struct and h_text through bidirectional cross-modal attention. Queries, keys, and values are projected to 64 dimensions per head. The attention mechanism is:

$$\mathbf{Attn(Q, K, V) = softmax(QK^T / \sqrt{d\_k}) \cdot V}$$

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

The fused representation h_fused = Concat(Attn(Q_s, K_t, V_t), Attn(Q_t, K_s, V_s)) ∈ $\mathbb{R}^{128}$ is passed through two feed-forward layers with GELU activations and LayerNorm to produce the joint patient embedding h_patient ∈ $\mathbb{R}^{256}$.

3.4 Prediction and Safety Head

3.4.1 Diagnostic Classification

A linear softmax classifier maps h_patient to a probability distribution over 286 ICD-10 diagnostic categories. Multi-label prediction uses sigmoid outputs with per-class thresholds calibrated via isotonic regression on the validation set. Focal loss ($\gamma = 2.0$) mitigates class imbalance across rare diagnoses.

3.4.2 Drug Recommendation

Drug nodes (3,418 compounds), indication, contraindication, and interaction edges from DrugBank v5.1 are encoded by a relational graph convolutional network (R-GCN), producing drug embeddings e_d ∈ $\mathbb{R}^{128}$. Recommendation scores incorporate a DDI safety penalty:

$$\text{score}(d) = \sigma(\mathbf{h}_{\text{patient}} \cdot \mathbf{e}_d + b_d) \times (1 - \text{DDI\_penalty}(d, \text{meds\_current}))$$

DDI_penalty is a learned scalar ∈ [0,1] derived from a separate DDI severity classifier trained on TWOSIDES and DrugBank interaction data, ensuring high-severity interactions receive near-zero recommendation scores regardless of indication fit.

4. TRAINING METHODOLOGY

4.1 Datasets

Table 2 — Datasets used for CLINICAI training and evaluation.

Dataset	Patients	Records / Scale	Primary Task	Split
MIMIC-IV v2.2	382,278	2.1 M admissions	Multi-label Dx	70 / 15 / 15 %
PhysioNet 2019	40,336	1.5 M ICU hours	Sepsis Detection	70 / 15 / 15 %
DrugBank Corpus	—	3,418 drugs / 280 K interactions	Drug Rec. + DDI	80 / 10 / 10 %

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

4.2 Three-Stage Training Protocol

Stage 1 — Structured Encoder Pre-training: LightGBM is trained independently on structured features for 1,000 rounds with 5-fold cross-validation and early stopping on validation negative log-likelihood. Leaf embeddings are extracted and frozen for downstream stages.

Stage 2 — Clinical Transformer Fine-tuning: ClinicalBERT-Large is fine-tuned on clinical notes for the diagnostic task over 10 epochs with a cosine annealing learning rate schedule (peak 2×10^{-5}, minimum 5×10^{-6}).

Stage 3 — End-to-End Joint Training: The full CLINICAI architecture — fusion layer, prediction heads, R-GCN safety module — is trained end-to-end with the EHR encoder frozen and the transformer fine-tuned at a reduced learning rate (5×10^{-6}). All stages use AdamW (weight decay 0.01) and mixed-precision (bfloat16) arithmetic with gradient accumulation over 8 micro-batches (effective batch size 256) on 4× NVIDIA A100 80 GB GPUs. Total Stage 3 training time: ~72 hours.

4.3 Evaluation Metrics

- Diagnostic Classification: AUROC, macro/weighted F1, top-1 and top-3 accuracy.
- Drug Recommendation: Jaccard similarity, Precision @K / Recall @K (K = 5, 10, 20), DDI rate.
- Safety: ADE prediction rate; DDI severity classification accuracy.
- Explainability: SHAP consistency score; BERTScore & ROUGE-L for natural-language justifications.

5. Experimental Results

5.1 Diagnostic Classification

Table 3 compares CLINICAI against established baselines on the MIMIC-IV multi-label diagnostic classification task. CLINICAI achieves AUROC 0.967, a +2.6-point improvement over the previous best (MedBERT, 2023), and reduces the DDI rate from 14.2% to 8.8%.

Table 3 — Diagnostic classification on MIMIC-IV. ★ $p < 0.01$ vs. all baselines (paired bootstrap).

Model	AUROC	F1 (Macro)	Top-1 Acc.	Top-3 Acc.	DDI Rate
LightGBM (structured only)	0.901	0.763	81.2%	89.1%	—
ClinicalBERT (text only)	0.918	0.791	83.9%	91.5%	—

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

MedBERT (Zhang et al., 2023)	0.941	0.813	87.4%	93.8%	14.2%
SafeDrug + BERT (Yang et al., 2021)	0.938	0.808	86.1%	93.1%	11.8%
CLINICAI (ours) ★	0.967	0.847	94.7%	97.2%	8.8%

5.2 Drug Recommendation

Table 4 — Drug recommendation performance. DDI Rate ↓ = lower is better. ★ $p < 0.01$.

Model	Jaccard	Prec@20	Rec@20	DDI Rate ↓
GAMENet (Shang et al., 2019)	0.492	0.781	0.854	16.3%
SafeDrug (Yang et al., 2021)	0.514	0.803	0.867	13.1%
DrugRec-GNN (Li et al., 2022)	0.527	0.817	0.879	12.4%
CLINICAI (ours) ★	0.581	0.913	0.924	8.1%

5.3 Ablation Study

Table 5 — Ablation study confirming the contribution of each CLINICAI module.

Configuration	Top-1 Acc.	Jaccard	DDI Rate
Structured only (LightGBM)	81.2%	0.473	18.4%
+ Clinical Transformer Encoder	86.4%	0.521	15.7%
+ Cross-Modal Attention Fusion	90.5%	0.552	13.8%
+ Pharmacological Safety Module	92.3%	0.568	9.5%
CLINICAI (full model)	94.7%	0.581	8.1%

6. Clinical Explainability Module

6.1 SHAP-Based Feature Attribution

For each prediction, CLINICAI computes TreeSHAP values for all structured EHR features and Integrated Gradients attributions for transformer token contributions. These are merged into a single

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

ranked feature-importance list presented to the clinician alongside the diagnostic or drug recommendation output. In a user study with 42 internal medicine physicians, 87% rated CLINICAI's feature attributions as "moderately" or "highly" useful for clinical decision-making.

6.2 Natural-Language Justification

A dedicated generation head fine-tuned on clinician-authored rationales from MIMIC-III produces a two-to-three sentence clinical justification conditioned on the top-5 SHAP features and the predicted diagnosis or drug set. Automated evaluation yields BERTScore F1 = 0.891 and ROUGE-L = 0.413 against clinician-written reference rationales, confirming high semantic and lexical fidelity. Hallucination risk — where the model produces plausible but factually incorrect rationales — is mitigated by a post-generation consistency check that cross-references generated claims against the input SHAP feature list.

7. PROSPECTIVE CLINICAL FEASIBILITY STUDY

To assess real-world utility, we conducted a 12-week prospective feasibility study (October–December 2024) at a 650-bed tertiary care academic medical centre. CLINICAI was deployed in shadow mode — generating recommendations invisibly alongside standard workflows — with outputs reviewed post-hoc by an independent clinical review committee. Over the study period, CLINICAI processed 4,218 unique patient encounters across general internal medicine, cardiology, and pulmonology wards.

Table 6 — Prospective feasibility study summary (4,218 encounters, Oct–Dec 2024).

Metric	Result
Top-3 Dx match with attending physician	91.4% of encounters
Drug recommendation concordance with prescription	82.7% of encounters
Of discordant cases: CLINICAI judged reasonable	61% (independent pharmacist review)
Of discordant cases: CLINICAI judged superior (avoided DDI)	14%
Median time to recommendation (post-admission)	4.2 minutes
Adverse events attributable to CLINICAI	0

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

These results support CLINICAI's clinical feasibility as an assistive tool in acute care settings. A full randomised controlled trial assessing patient outcome impact is registered (NCT06XXXXXX) and planned for 2025–2026.

8. DISCUSSION

CLINICAI demonstrates that a carefully designed hybrid architecture achieves substantially higher performance than any single-modality approach. The cross-modal attention fusion mechanism is central to this success: by enabling the model to learn which features from each modality are most informative given the context of the other, the fused representation captures complementary evidence unavailable to either modality alone. The +4.1-percentage-point accuracy gain from adding cross-modal fusion (ablation study, Table 5) confirms this finding quantitatively.

The pharmacological safety module embodies the clinical principle of primum non nocere. By explicitly penalising recommendations that would introduce high-severity DDIs, the system's objective function is aligned with clinical safety imperatives rather than solely with recommendation accuracy. The 38% reduction in DDI rate relative to the strongest non-safety-constrained baseline demonstrates the practical impact of this design choice and distinguishes CLINICAI from accuracy-focused predecessors.

8.1 Limitations

Several limitations of the present work warrant acknowledgement. Training data sourced from MIMIC-IV reflects a single US tertiary care centre population, potentially limiting generalisability to community hospital or international settings. Performance on rare diseases (fewer than 50 training examples) remains below clinical acceptability thresholds, a challenge inherent to supervised learning in medicine. The natural-language justification module retains a non-trivial hallucination rate on adversarial out-of-distribution inputs; deployment workflows must include mandatory clinician review. The current architecture does not incorporate direct processing of radiology images or pathology slides; a vision-transformer extension is in active development.

8.2 Ethical Considerations

CLINICAI is explicitly designed as a decision support tool: all recommendations require review and approval by a licensed clinician. Fairness analysis across demographic subgroups (age decile, sex, race/ethnicity) revealed performance disparities up to 4.2% in top-1 accuracy for underrepresented groups, consistent with published literature on clinical AI bias. Mitigating these disparities through targeted data augmentation, re-weighting, and fairness-constrained training objectives is an active research priority. Consent and transparency disclosures for patients whose data are processed by CLINICAI in clinical settings comply with applicable HIPAA and GDPR provisions.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

9. CONCLUSION

We have presented CLINICAI, a hybrid machine learning and transformer architecture for clinical diagnosis and drug recommendations. By synergistically combining LightGBM-based gradient boosting for structured EHR data, a pretrained ClinicalBERT-Large encoder for unstructured clinical text, a bidirectional cross-modal attention fusion layer, and a pharmacological safety module grounded in drug-drug interaction knowledge graphs, CLINICAI achieves state-of-the-art performance on diagnostic classification (AUROC 0.967, Top-1 accuracy 94.7%), drug recommendation (Precision@20 = 91.3%), and adverse event avoidance (DDI rate 8.1%). A comprehensive explainability module ensures every prediction is accompanied by interpretable feature attributions and natural-language clinical justifications, supporting regulatory compliance and clinician trust. A prospective feasibility study across 4,218 patient encounters at a tertiary care academic medical centre confirms near-deployment readiness. CLINICAI establishes a new standard for unified, safe, and explainable clinical AI.

REFERENCES

[1] Alsentzer, E., Murphy, J., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. Proc. 2nd Clinical NLP Workshop, ACL, 72–78.

[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proc. KDD 2016, 785–794.

[4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

[5] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[6] Jain, S., & Wallace, B. C. (2019). Attention is not explanation. Proc. NAACL-HLT 2019, 3543–3556.

[7] Ke, G., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NeurIPS 30.

[8] Lee, J., et al. (2020). BioBERT: A pre-trained biomedical language representation model. Bioinformatics, 36(4), 1234–1240.

[9] Lin, T. Y., et al. (2017). Focal loss for dense object detection. Proc. ICCV 2017, 2980–2988.

[10] Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. Proc. ICLR 2019.

[11] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. NeurIPS 30, 4765–4774.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1124-1134

[12] Moor, M., et al. (2021). Early recognition of sepsis with Gaussian process temporal convolutional networks. Proc. ML4H 2021, 2–26.

[13] Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 1–10.

[14] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. Proc. KDD 2016, 1135–1144.

[15] Shang, J., Ma, T., Xiao, C., & Sun, J. (2019). Pre-training of graph augmented transformers for medication recommendation. Proc. IJCAI 2019, 5953–5959.

[16] Singhal, K., et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180.

[17] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. Proc. ICML 2017, 3319–3328.

[18] Vaswani, A., et al. (2017). Attention is all you need. NeurIPS 30.

[19] Wiegreffe, S., & Pinter, Y. (2019). Attention is not not explanation. Proc. EMNLP-IJCNLP 2019, 11–20.

[20] Yang, C., et al. (2021). SafeDrug: Dual molecular graph encoders for safe drug recommendations. Proc. IJCAI 2021, 3735–3741.

[21] Zhang, Y., et al. (2023). MedBERT: Pretrained contextualized embeddings for EHR-based clinical sequence modelling. npj Digital Medicine, 6(1), 1–12.