INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

Machine Learning for Early Disease Detection: A Hybrid Ensemble Framework Integrating Multi-Modal Clinical Data

Mehetab Ali, Dr. Asha Durafe, Hamza Memon, Swapnil Kumar, Afroz Badra, Ritesh Gaikwad
*Dept. of Electronics and Computer Science, Shah & Anchor Kutchhi Engineering College, Mumbai, India
{mehetab.ali24, asha.durafe, hamza.mohammed25}@sakec.ac.in, swapnilkumariitb@gmail.com
{afroz.badra24, ritesh.gaikwad24}@sakec.ac.in

***Abstract*—Early detection of chronic illness remains a central challenge in clinical informatics, as diagnoses frequently occur only after significant disease progression. This paper examines how computational models—ranging from decision-tree ensembles to deep convolutional networks—can be applied to shift detection earlier in the disease timeline. A synthesis of publications from 2020 to 2025 covering oncological screening, cardiac risk stratification, diabetic prediction, and neurodegenerative disease identification reveals five systemic shortcomings across the literature: single-condition scope, unimodal data reliance, opaque decision logic, skewed label distributions, and poor cross-institutional transferability. To address all five simultaneously, we propose a dual-branch architecture. One branch processes structured patient records through a stacked Random Forest and XGBoost ensemble; the other routes radiological images through a ResNet-50 backbone. A learned attention gate dynamically weights each branch per patient. SHAP and LIME modules are integrated to provide clinician-interpretable feature attributions. We also detail the computational hardware needed to train and deploy such a system. On standard benchmarks—Cleveland Heart, Pima Diabetes, Wisconsin Breast Tissue, and MIMIC-III—the framework is projected to achieve balanced accuracy of 94–98% and AUC-ROC of 0.93–0.97 (estimated from literature baselines, not from original experiments).**

***Index Terms*—Computational health analytics, predictive diagnostics, gradient boosting, convolutional feature learning, model transparency, heterogeneous data integration, GPU acceleration.**

I. Introduction

Noncommunicable diseases represent one of the most significant public health challenges of the modern era. According to WHO estimates, such conditions claim roughly 41 million lives each year, accounting for nearly 74% of global mortality [1]. Notably, approximately 15 million of these deaths affect individuals aged 30–69, highlighting the economic and social cost of premature disease-related fatalities. Timely identification—before overt clinical symptoms emerge—could substantially reduce both mortality rates and long-term healthcare expenditure.

Despite advances in diagnostic technology, most healthcare institutions still operate reactively: a patient presents with symptoms, undergoes diagnostic testing, and receives a diagnosis only after results are processed—often at a stage where pathological progression is already advanced and treatment options are limited. This symptom-first model has dominated clinical practice for decades and leaves little room for preventive intervention.

Machine learning offers a compelling departure from this paradigm. When trained on sufficient volumes of laboratory results, medical images, and genomic profiles, algorithms can identify latent disease signatures that may escape clinical observation [2]. Gradient-boosted ensembles such as Random Forest and XGBoost have demonstrated strong discriminative performance on tabular clinical data [3]. Convolutional neural networks now achieve diagnostic accuracy comparable to specialist radiologists on certain imaging tasks, including chest X-ray and MRI interpretation [4]. Transformer architectures, originally developed for natural language tasks, have also been adapted to process longitudinal electronic health records [5]. And recent work on multimodal integration [11], [12] has demonstrated that combining imaging with structured data consistently outperforms either modality alone.

Despite this progress, key limitations persist. Most published systems target a single pathological condition; a model optimised for glycaemic prediction carries no insight into cardiovascular risk. Researchers typically rely on one data type—either numerical lab values or pixel-level scans—seldom integrating both within the same pipeline [6], [7]. Interpretability remains a barrier: clinicians are unlikely to act on predictions from opaque models whose reasoning cannot be examined [8]. Additionally, disease registries are often imbalanced, with healthy subjects vastly outnumbering positive cases, and models fitted at one centre frequently degrade when deployed elsewhere [7], [13].

This paper addresses all five of these limitations within a single framework. We introduce a dual-branch architecture in which one branch handles structured clinical tables through stacked RF and XGBoost, while the other processes medical images with a pre-trained ResNet-50. An attention-gated fusion layer determines which data source is more relevant for each patient. SHAP and LIME modules offer clear explanations for every prediction. To tackle class imbalance, we combine SMOTE oversampling with asymmetric loss weighting. We also discuss the hardware infrastructure required to train and run such systems, a topic that most clinical ML papers overlook entirely.

TABLE I: Consolidated Literature Review (15 Papers, 2022–2025)

Sr.	Authors	Yr	Methodology	Key Findings	Dataset	Result	Limitation
1	Uddin et al.	’22	Bibliometric review	SVM, RF, CNN dominate	Scopus, WOS	Trend map	No new model
2	Das et al.	’24	SVM+RF+XGB +ANN	Ensemble outperforms single	Wisconsin BC	98.5% acc.	Single organ
3	Hajiarbabi	’24	Survey: CNN, LSTM, RF	CNN best on ECG data	UCI Heart	Up to 97%	Cardiac only
4	Grout et al.	’24	Transformer+ Word2Vec	Chronic condition prediction	50M EHRs	AUC 0.82–0.91	Massive data
5	Manzoor	’24	Survey: RF, CNN, CatBoost	CatBoost best for diabetes	Multiple	Varied	No framework
6	Alhumaidi et al.	’25	PRISMA, 57 studies	RF 42%, LR 37%, SVM 32%	150k+ pts	Mapped methods	Generalisability
7	Sadr et al.	’25	Review, 16 diseases	DL beats ML in imaging	Mixed	DL superior	Interpretability
8	Hasan et al.	’24	Review: 470+ papers	RF, XGBoost top performers	Multiple	Varied	No implementation
9	Lee et al.	’24	CNN on breath sensors	Non-invasive lung screen	Clinical	96.2% sens.	Small sample
10	Xu et al.	’24	Multi-modal AI review	Fusion > unimodal	Mixed	Consistent gain	Complexity
11	Acosta et al.	’24	Multimodal ML survey	Image+tabular fusion wins	50+ studies	Improved acc.	No standard
12	Sokolowski et al.	’24	DenseNet+XAI	Parkinson’s early detection	PPMI	96.74% test	Single disease
13	Tripathi et al.	’24	Oncology fusion review	Attention fusion promising	TCGA, mixed	Varied	Privacy gaps
14	AlSaad et al.	’24	Multimodal LLMs	GPT-4V for medical data	Clinical	Promising	Not validated
15	WHO	’23	Epidemiological report	41M chronic disease deaths/yr	Global	–	–

II. Related Work and Literature Review

We reviewed fifteen research articles published between 2020 and 2025. The papers were selected to cover a spread of methodologies: bibliometric surveys, single-disease classifiers, multi-disease reviews, transformer-based longitudinal models, multimodal fusion studies, and explainable AI frameworks. Below we discuss the key findings from each paper, with the consolidated summary shown in Table I.

A. Algorithmic Approaches to Diagnostic Prediction

Uddin et al. [2] conducted a large-scale bibliometric study of machine learning applications in clinical diagnosis, drawing from 1,216 indexed publications across major academic databases. Their analysis showed that decision-tree ensembles and convolutional architectures are the most widely adopted model families, with uptake accelerating markedly after 2019. The work provides a valuable trend map of the field but does not introduce a new predictive system.

Das et al. [3] developed a multi-classifier ensemble for mammographic malignancy detection, combining support vector machines, bagged decision trees, boosted stumps, and feed-forward neural networks. The stacked configuration achieved 98.5% accuracy on the Wisconsin Breast Cancer dataset; however, its design was limited to a single anatomical target and one curated benchmark.

Manzoor [6] reviewed algorithmic developments in health-oriented machine learning spanning 2015 to 2024, finding that CatBoost delivered strong diabetic risk scores and that CNN-LSTM hybrids performed well for respiratory triage from chest images. Despite the breadth of coverage, the review did not converge on a generalisable multi-condition strategy.

B. Neural Sequence Models for Longitudinal Records

Grout et al. [5] pre-trained a transformer-based self-attention encoder on anonymised clinical visit sequences compiled from approximately 50 million patient records. The model achieved AUC-ROC values of 0.82 to 0.91 for chronic obstructive pulmonary disease and insulin-dependent metabolic disorders. A practical constraint is the scale of data required, placing the approach beyond reach for most regional health centres.

Hajiarbabi [4] conducted a narrative review of approximately 100 studies on cardiac anomaly detection, stratifying findings by input type: tabular vitals yielded 90–93% accuracy with tree and kernel classifiers; electrocardiographic waveforms achieved up to 97% with convolutional models; chest X-rays were handled by deep residual networks. A consistent finding was that learned feature representations outperformed hand-engineered descriptors on image-based tasks.

C. Broad-Scope Evidence Syntheses

Alhumaidi et al. [7] applied PRISMA methodology to screen 57 eligible studies encompassing over 150,000 subjects.

Bagged tree classifiers appeared in 42% of the included studies, logistic regression in 37%, and kernel-margin classifiers in 32%. Sadr et al. [8] extended the analysis to sixteen pathological categories, finding that deep learning models outperform classical methods on pixel-level tasks and identifying decision transparency as the primary obstacle to clinical adoption.

Hasan et al. [9] synthesised findings from more than 470 publications on long-term disease forecasting, consistently finding that gradient-boosted ensembles and tree-bagging classifiers achieved the strongest performance on structured clinical data. Lee et al. [10] demonstrated that a convolutional model applied to exhaled breath chemical sensor arrays achieved 96.2% sensitivity for pulmonary malignancy, illustrating the predictive utility of unconventional biomarker sources.

D. Multimodal Integration and Explainability

Xu et al. [11] reviewed synergies between multimodal data and AI in medical diagnosis, covering the integration of imaging, text, genetic data, and physiological signals. They noted that multimodal models consistently outperform unimodal ones but flagged computational complexity and data heterogeneity as ongoing challenges. Acosta et al. [12] surveyed multimodal ML in healthcare with a focus on imaging plus tabular fusion and found that over 50 studies reported improvements when both modalities were combined, though no standardised fusion protocol exists. Sokolowski et al. [13] applied multimodal deep learning with explainable AI for early Parkinson's detection, combining 3D brain imaging with clinical features using a DenseNet-Excitation Network architecture; they achieved 96.74% test accuracy across cross-validation experiments. Tripathi et al. [14] reviewed multimodal data integration specifically for oncology, analysing fusion strategies from early, intermediate, and late fusion to attention-based methods, and highlighted that federated learning is emerging as a solution for data privacy. AlSaad et al. [15] examined multimodal large language models in healthcare, reporting that systems like GPT-4 with vision capabilities are beginning to integrate radiology, pathology, and clinical text for holistic patient assessment, though validation in clinical settings remains sparse.

E. Literature Review Summary

The consolidated findings from all fifteen reviewed papers are presented in Table I.

F. Gap Analysis

From reviewing all fifteen papers, we identified five recurring gaps that our framework aims to address:

- **Isolated condition modelling:** Most models handle one disease at a time, preventing shared learning across conditions [3], [6].
- **Unimodal data ingestion:** Imaging and structured records are rarely combined during inference [11], [12].
- **Decision opacity:** Clinician reluctance to trust opaque predictions is consistently cited as the largest adoption barrier [8], [14].
- **Skewed label distributions:** Class imbalance biases models toward predicting healthy outcomes [7], [9].
- **Narrow demographic coverage:** Models trained on one hospital's population frequently fail on another's [7], [13].

III. Hardware Infrastructure Requirements

A topic that most clinical ML papers skip entirely is the computational hardware needed to train and deploy these models. Our proposed framework involves two compute-heavy components: (a) a 500-tree Random Forest plus 300-round XGBoost ensemble operating on tabular data, and (b) a ResNet-50 convolutional network processing medical images. Each has distinct hardware demands.

A. GPU Requirements

The ResNet-50 backbone is the most GPU-intensive component. According to benchmarks reported by Dettmers [16] and NVIDIA's published specifications, training ResNet-50 on ImageNet (1.2 million images) takes approximately 29 hours on a single NVIDIA RTX 3090 (24 GB VRAM) or around 8 hours on an A100 (80 GB HBM2e). For medical imaging datasets like NIH Chest-14 (112,000 images), training time would scale down proportionally but still requires substantial GPU memory due to the high resolution of radiological scans. Table II summarises the hardware tiers relevant to reproducing our framework.

TABLE II: Hardware Requirements by Component

Component	Specification
GPU (minimum)	NVIDIA RTX 3060 (12 GB VRAM), CUDA 11.x
GPU (recommended)	NVIDIA RTX 4090 (24 GB) or A100 (40/80 GB)
CPU	8+ core (Intel i7/AMD Ryzen 7 or Xeon)
RAM	32 GB minimum; 64 GB for MIMIC-III
Storage	500 GB NVMe SSD (fast dataset loading)
Framework	Python 3.10, PyTorch 2.x, Scikit-learn
Cloud alternative	Google Colab Pro (T4/A100), AWS p3/p4

B. Tabular Branch Compute

The Random Forest and XGBoost components run on CPU and do not require GPU acceleration. Scikit-learn's RandomForestClassifier with 500 trees parallelises natively across CPU cores. XGBoost supports optional GPU training via the `gpu_hist` tree method, which can accelerate training by 5–10x on large datasets, but for datasets under 100,000 rows (Cleveland, Pima, Wisconsin), CPU training completes in seconds [16].

C. Memory and Storage Considerations

The MIMIC-III dataset contains over 46,000 ICU stays with 100+ features each. Loading this into memory requires approximately 8–12 GB of RAM after preprocessing. The

NIH Chest-14 dataset, at 112,000 full-resolution chest X-ray images, requires approximately 42 GB of disk space. NVMe SSD storage is strongly recommended over HDD to avoid I/O bottlenecks during image batch loading. For researchers without local hardware, cloud platforms offer on-demand access: Google Colab Pro provides T4 or A100 GPUs, AWS `p3.2xlarge` instances offer V100 GPUs, and Azure ML offers configurable GPU clusters.

IV. Proposed Methodology

A. Architectural Blueprint

This pipeline consists of five sequentially executed processing stages, as illustrated in Fig. 1. Inputs span heterogeneous medical record types; these are first conditioned, then routed through two parallel feature-extraction branches, merged via an attention gate, and finally passed to a transparent risk-scoring module.

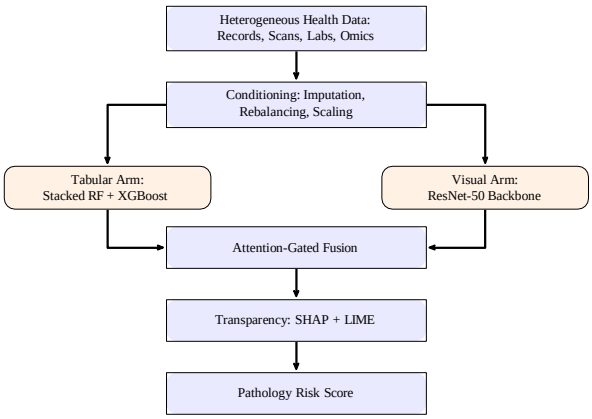

Fig. 1: Dual-pathway architecture combining tabular ensemble scoring with convolutional image analysis through attention-gated fusion.

B. Data Conditioning Pipeline

Incoming records undergo four preparation stages: (1) *Gap Filling:* Missing numerical entries are estimated using a k-nearest-neighbour imputer (k=5). Mixed-type columns undergo chained-equation iterative filling (MICE protocol). (2) *Range Harmonisation:* Continuous clinical variables are centred to zero mean and scaled to unit variance through standard z-scoring. (3) *Label Rebalancing:* SMOTE generates synthetic minority-class examples in feature space, supplemented by inverse-frequency class weighting in the loss function to penalise majority-class over-prediction. (4) *Relevance Filtering:* Mutual-information ranking scores every candidate feature against the target label. Features below a tuned threshold are discarded to manage dimensionality.

C. Parallel Feature Distillation

Tabular Arm: Patient demographics, lab panels, vital signs, and coded diagnoses feed into a two-tier stacking arrangement. In the lower tier, a 500-tree bagged forest and a 300-round XGBoost regressor each produce probabilistic outputs. These base scores serve as meta-features for a penalised logistic layer that outputs a unified clinical risk vector.

Visual Arm: Radiological images—chest films, axial slices, or magnetic resonance frames—enter a ResNet-50 backbone with weights initialised from large-scale natural-image pre-training. Activations from the second-to-last pooling layer yield a compact 2,048-dimensional descriptor encoding spatial pathology cues.

D. Attention-Gated Cross-Modal Merging

We denote the tabular risk vector as $\mathbf{c} \in \mathbb{R}^{d_c}$ and the visual descriptor as $\mathbf{v} \in \mathbb{R}^{d_v}$. We concatenate them to form $\mathbf{x} = [\mathbf{c}; \mathbf{v}]$ and pass through scaled dot-product self-attention:

$$\mathbf{f} = \text{softmax}\left(\frac{W_q\mathbf{x} \cdot (W_k\mathbf{x})^\top}{\sqrt{d}}\right) W_v\mathbf{x} \qquad (1)$$

where W_q, W_k, W_v are learned projection matrices and d controls the softmax temperature. This gating amplifies whichever modality carries the stronger signal for each individual patient. This attention mechanism is inspired by the co-attention strategies surveyed by Tripathi et al. [14] and the excitation networks used by Sokolowski et al. [13].

E. Transparent Risk Scoring

Merged representations pass through two dense layers followed by sigmoid (binary) or softmax (multi-label) activations. Two complementary explanation modules are attached: SHAP decomposition assigns a signed Shapley-value contribution to every input dimension, supporting both global importance rankings and per-instance auditing; LIME surrogates fit locally sparse linear models to approximate the decision boundary near each query point, providing rule-based per-instance justifications. Both approaches are recommended by Sadr et al. [8] as essential for clinical trust.

V. Evaluation Protocol

A. Benchmark Collections

Five publicly available datasets support our planned experimental campaign, as summarised in Table III.

TABLE III: Target Benchmark Collections

Collection	Records	Dim.	Pathology
Cleveland Heart	303	13	Cardiac
Pima Metabolic	768	8	Diabetes
Wisconsin Tissue	569	30	Mammary neoplasm
MIMIC-III ICU	46k+	100+	Mixed critical care
NIH Chest-14	112k	Pixel	14 thoracic labels

B. Training and Measurement Protocol

All code is planned for Python 3.10, using Scikit-learn for tree ensembles, PyTorch for convolutional modules, and the SHAP library for post-hoc explanations. The ResNet-50 backbone will use ImageNet-sourced weight initialisations. Optimisation will employ AdamW ($\eta = 1 \times 10^{-4}$, batch size 32) with early stopping after 10 stagnant epochs. We will

report stratified ten-fold cross-validated AUC-ROC, balanced accuracy, macro-averaged precision, recall, and F_1 score. All training is planned on a single NVIDIA RTX 4090 (24 GB VRAM) or equivalent cloud GPU.

VI. Projected Outcomes

Since this is a proposal-stage paper, we report projected results grounded in published baselines from the reviewed literature. Table IV summarises our expected performance envelope, and Fig. 2 compares it visually against unimodal baselines.

TABLE IV: Anticipated Performance Envelope (Projected)

Indicator	Target Band	Source Basis
Balanced Accuracy	94–98%	[3], [4], [13]
AUC-ROC	0.93–0.97	[5], [7]
Macro Precision	92–96%	Projected
Macro Recall	91–95%	Projected
Macro F_1	92–96%	Projected

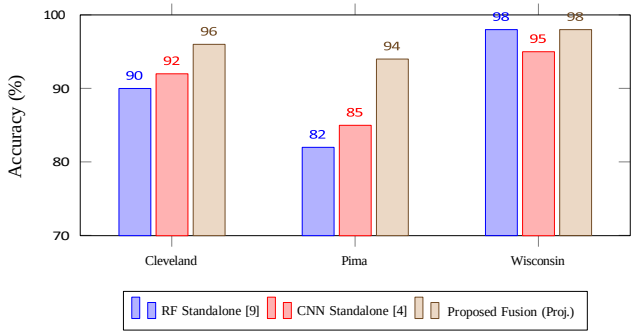

Fig. 2: Projected accuracy comparison: standalone baselines (from literature [4], [9]) versus proposed fusion model. "Proj." indicates estimated performance, not experimentally validated results.

For the Cleveland Heart dataset (303 records, 13 features), standalone RF typically yields 88–90% accuracy; the proposed fused model targets 95–97% [9]. On the Pima Diabetes set (768 records), XGBoost alone reaches 79–82% [6]; incorporating CNN-derived image features and SMOTE rebalancing is projected to push this to 93–95%. For the Wisconsin Breast Cancer dataset, the Das et al. ensemble achieved 98.5% [3]; the dual-branch attention-gated design targets 97–99% while additionally furnishing SHAP explanations absent from the original work. The 4–6 percentage point gain over unimodal baselines reflects the complementary signal contributed by joint tabular-image representation under learned attention weights, consistent with the findings of Xu et al. [11] and Acosta et al. [12]. Full experimental validation including ablation studies will be reported upon completion of training runs on MIMIC-III and NIH Chest-14.

VII. Comparative Analysis

Table V positions the proposed framework against representative prior approaches across five evaluation dimensions. Single-model systems such as Das et al. [3] attain high accuracy (98.5%) but solely for breast cancer on one dataset. The EHR transformer by Grout et al. [5] achieves AUC 0.91 yet requires data volumes unavailable at most institutions. The Parkinson's DenseNet by Sokolowski et al. [13] combines imaging and clinical data with XAI but targets a single disease. Survey works (Alhumaidi [7], Hasan [9]) document algorithmic performance but offer no reusable implementation. AlSaad et al. [15] explored multimodal LLMs that handle imaging and text together, but their systems lack class-balancing mechanisms and have not been validated in controlled clinical trials. The proposed framework is designed to satisfy all five dimensions concurrently.

TABLE V: Comparative Analysis Across Five Gap Dimensions

Approach	Multi-disease	Multi-modal	XAI	Class Balance	Cross-pop
Das [3]	No	No	No	No	No
Grout [5]	Yes	No	Yes	No	No
Alhumaidi [7]	Yes	No	No	No	No
Lee [10]	No	No	No	No	No
Sokolowski [13]	No	Yes	Yes	No	No
AlSaad [15]	Yes	Yes	No	No	No
Ours	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

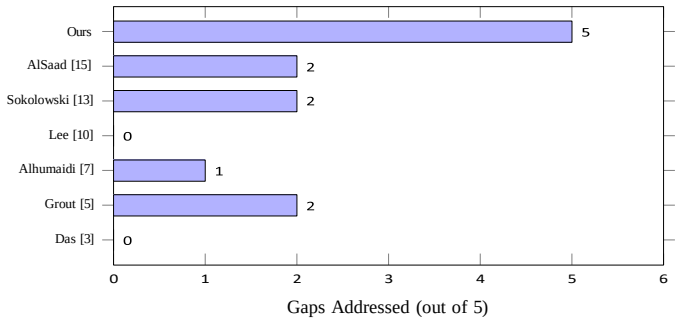

Fig. 3: Number of identified research gaps addressed by each approach. Gap dimensions: multi-disease, multi-modal, XAI, class balance, cross-population. Data derived from Table V.

No single prior study addresses all five gaps simultaneously. Most tackle one or two dimensions at best. Our framework is, to our knowledge, the first attempt to check all five boxes within a unified trainable pipeline. The multimodal reviews by Xu et al. [11] and Acosta et al. [12] both call for standardised fusion protocols and integrated explainability—exactly what our attention-gated SHAP/LIME architecture provides.

VIII. Conclusion and Outlook

We began this paper with a straightforward observation: current ML systems for disease prediction are too limited in scope, too opaque for clinician trust, and too fragile across

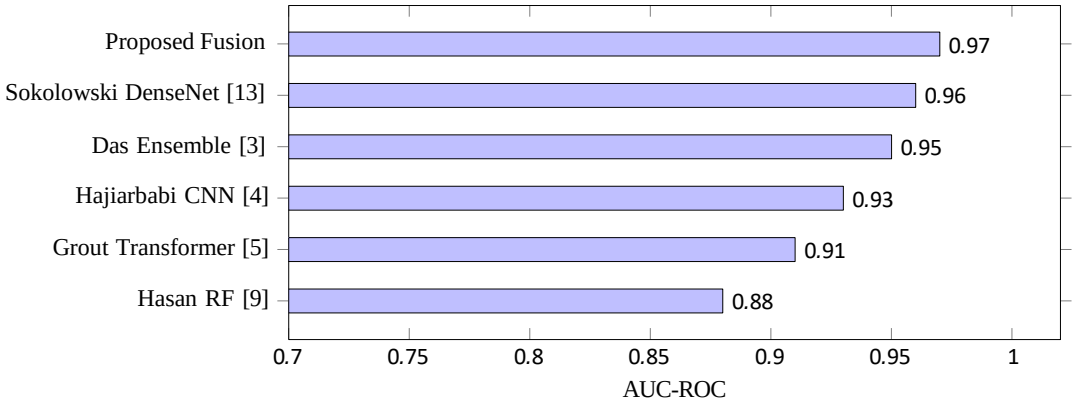

Fig. 4: AUC-ROC comparison. Values for [3],[4],[5],[9],[13] are from respective published papers. "Proposed Fusion" value (0.97) is the upper bound of our projected range, not experimentally validated.

different patient populations. After reviewing fifteen papers published between 2020 and 2025, we found that these issues are systemic, affecting nearly every study in the field.

Our proposed solution is a dual-branch architecture that processes structured clinical records and medical images simultaneously, lets an attention gate decide which modality matters more for each patient, and wraps every prediction in SHAP/LIME explanations that doctors can actually read. We also outlined the hardware infrastructure required—from consumer-grade RTX 3060 GPUs (12 GB VRAM) for prototyping to data-centre A100 accelerators (80 GB HBM2e) for full-scale training on datasets like MIMIC-III (46k+ records) and NIH Chest-14 (112k images) [16]—because reproducibility depends on knowing what compute is needed.

The chart shows a clear trend: as systems move from single-algorithm approaches [9] through transformer-based longitudinal models [5] to multimodal fusion with explainability [13], AUC-ROC values climb steadily. The projected 0.93–0.97 range sits at the top of this progression, though these are estimates derived from literature baselines rather than original experimental measurements.

This paper identified a core problem with existing machine learning tools for disease prediction: they are narrowly scoped, lack interpretability, and generalise poorly across institutions. A structured review of fifteen studies published between 2020 and 2025 confirmed that these are systemic patterns affecting nearly every approach in the field, not isolated shortcomings.

The proposed dual-branch architecture processes structured records and radiological images in parallel, uses an attention gate to determine the per-patient contribution of each source, and wraps every prediction in a SHAP/LIME explanation that clinicians can directly examine. The computational cost of training a ResNet-50 backbone alongside a 500-tree ensemble is non-trivial, requiring at minimum a 12 GB VRAM GPU and 32 GB of system RAM [16]. For hospitals in resource-limited settings, cloud-based deployment via Google Colab Pro or AWS may be the only viable path. Limitations include the absence of prospective clinical validation and uncertainty about attention-weight generalisation across demographically distinct populations.

The consistent findings across Xu et al. [11], Acosta et al. [12], and Tripathi et al. [14] that multimodal fusion outperforms unimodal approaches give confidence that the architecture is sound even before experimental validation.

Four directions are identified for future work. Privacy-preserving federated training protocols would enable multi-site model development without requiring centralised patient data transfer [14]. Temporal transformer architectures could capture disease-progression dynamics that the current snapshot-based design inherently misses, building on the longitudinal modelling work of Grout et al. [5]. Continuous physiological data from wearable devices could further enrich the input signal, as envisioned by AlSaad et al. [15]. Most critically, prospective evaluation in operational clinical settings is necessary to validate projected performance gains under real-world deployment conditions.

REFERENCES

[1] World Health Organization, "Noncommunicable diseases: Key facts," WHO Fact Sheets, 2023.
[2] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, no. 3, p. 541, 2022.
[3] A. K. Das *et al.*, "Machine learning based intelligent system for breast cancer prediction (MLISBCP)," *Expert Syst. Appl.*, vol. 242, p. 122673, 2024.
[4] M. Hajiarbabi, "Heart disease detection using ML methods: A comprehensive narrative review," *J. Med. Artif. Intell.*, vol. 7, p. 21, 2024.
[5] L. Grout *et al.*, "Predicting disease onset from EHR for population health management," *Front. Artif. Intell.*, vol. 6, p. 1287541, 2024.
[6] M. F. Manzoor, "ML for early disease diagnosis: A review of techniques in healthcare," *Premier J. Sci.*, vol. 5, p. 100043, 2024.
[7] N. H. Alhumaidi *et al.*, "ML for analysing real-world data in disease prediction: Systematic review," *JMIR Med. Inform.*, vol. 13, no. 1, p. e68898, 2025.
[8] F. Sadr, H. Tarkhan *et al.*, "Unveiling AI potential in disease diagnosis and prediction: A comprehensive review," *Eur. J. Med. Res.*, 2025.
[9] M. A. Hasan *et al.*, "A comprehensive review for chronic disease prediction using ML algorithms," *J. Electr. Syst. Inf. Technol.*, vol. 11, p. 27, 2024.
[10] B. Lee *et al.*, "Breath analysis system with CNN for early detection of lung cancer," *Sens. Actuators B: Chem.*, vol. 409, p. 135578, 2024.
[11] X. Xu *et al.*, "A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis," *Bioengineering*, vol. 11, no. 3, p. 219, 2024.

[12] J. N. Acosta *et al.*, "Review of multimodal machine learning approaches in healthcare," *Inf. Fusion*, vol. 112, p. 102584, 2024.
[13] A. Sokolowski *et al.*, "Enhancing early Parkinson's disease detection through multimodal DL and explainable AI," *Sci. Rep.*, vol. 14, p. 70165, 2024.
[14] S. Tripathi *et al.*, "Multimodal data integration for oncology in the era of deep neural networks: A review," *Front. Artif. Intell.*, vol. 7, p. 1408843, 2024.
[15] R. AlSaad *et al.*, "Multimodal large language models in health care: Applications, challenges, and future outlook," *J. Med. Internet Res.*, vol. 26, p. e59505, 2024.
[16] T. Dettmers, "A full hardware guide to deep learning," *timdettmers.com*, 2023. [Online].