Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL FEATURE-SELECTION-GUIDED DEEP LEARNING FRAMEWORK INTEGRATING EXPLAINABLE AI FOR SCALABLE, DUAL-DIAGNOSIS PREDICTION OF DIABETES AND ADHD USING HETEROGENEOUS BIG HEALTH DATA

TANGUTURU SP MADHURI [1]*, DR. G. S RAGHAVENDRA [2]
[1]Research Scholar, CSE at Koneru Lakshmaiah Education Foundation deemed to be University, Bachupally, Hyderabad.
[2]Assistant Professor CSE at Koneru Lakshmaiah Education Foundation, deemed to be University, Bachupally, Hyderabad.
*Corresponding Author

ABSTRACT

This study introduces a novel, feature-selection-guided deep learning framework designed to enable scalable and interpretable dual-diagnosis prediction of Diabetes and Attention Deficit Hyperactivity Disorder (ADHD) using heterogeneous health data. The model integrates structured clinical variables from electronic health records (EHRs), temporal patterns from wearable sensors, and behavioural features from psychological assessments. Advanced feature selection techniques including LASSO, Recursive Feature Elimination (RFE), and XGBoost were employed to extract 38 high-impact variables from over 400 inputs. These selected features feed into a multi-branch neural network comprising Convolutional Neural Networks (CNNs) for static inputs, Long Short-Term Memory (LSTM) networks for time-series data, and a Multilayer Perceptron (MLP) fusion layer for dual-output classification. The architecture achieved high accuracy for both Diabetes (93.2%) and ADHD (92.5%) with AUCs above 0.94, outperforming baseline models. Integrated explainable AI tools SHAP, LIME, and Grad-CAM enhanced interpretability, revealing key biomarkers such as HRV entropy, glucose variability, BMI, and impulsivity score. The model generalized well across demographic subgroups and demonstrated robust diagnostic capability even in comorbid cases. This work makes a distinctive contribution by embedding explainable AI directly within the deep learning training process, enabling transparent and adaptive model refinement. It also discovers shared physiological and behavioural biomarkers linking Diabetes and ADHD, offering new knowledge for comorbidity-aware healthcare analytics. This end-to-end framework provides a clinically transparent, technically rigorous solution for early, accurate, and interpretable multi-disease prediction, representing a significant step forward in comorbidity-aware healthcare AI systems.

Keywords: *Dual-diagnosis Prediction, Heterogeneous Health Data, Feature Selection, Deep Learning Architecture, Explainable Artificial Intelligence, Comorbidity Modelling*

1. INTRODUCTION

Chronic health conditions such as Diabetes and neurodevelopmental disorders like attention deficit hyperactivity disorder (ADHD) continue to pose a significant burden on healthcare systems worldwide[1]. These conditions affect millions of individuals across age groups, often resulting in reduced quality of life, long-term complications, and increased healthcare expenditure. With the proliferation of digital health technologies ranging from Electronic Health Records (EHRs) to wearable devices and behavioural monitoring systems there is a growing opportunity to harness artificial intelligence (AI) to support early detection and improve medical decision-making.

Despite notable advances in the application of AI to disease diagnosis, most models are trained and deployed in isolation for single-disease prediction[2]. For instance, separate algorithms are designed for detecting Diabetes using clinical biomarkers or for predicting ADHD using behavioural and neurocognitive assessments[3]. This fragmented approach overlooks the reality of comorbidity, especially the statistically significant co-occurrence of ADHD and Diabetes in individuals

with metabolic or neurobehavioral vulnerabilities[4]. A unified diagnostic model is urgently needed to address this clinically relevant overlap.

Recent studies have applied machine learning (ML) and deep learning (DL) algorithms to structured and unstructured health data with promising results. Models such as Support Vector Machines[5], Random Forests[6], and CNNs have shown effectiveness in individual disease domains. Additionally, feature selection methods like LASSO and Recursive Feature Elimination (RFE)[7], as well as interpretability tools such as SHAP and LIME[8], have been employed to enhance understanding and reduce dimensionality. However, most frameworks stop short of combining these tools into a single, end-to-end diagnostic pipeline capable of handling heterogeneous data sources[9].

A key gap persists: there is currently no comprehensive AI system capable of delivering real-time, interpretable, and dual-disease predictions by jointly modelling ADHD and Diabetes using heterogeneous data from EHRs, wearables, and behavioral sources[10]. Existing approaches fail to address the challenge of high-dimensional, asynchronous data integration, and most rely on post hoc explanations, which are less useful for iterative clinical model improvement. Furthermore, these models often perform poorly when generalized to diverse demographic populations due to training on limited, homogeneous datasets.

There is currently no scalable, interpretable, and feature-optimized deep learning framework for dual-diagnosis of ADHD and Diabetes that integrates heterogeneous data sources and explainable AI into a single pipeline[10]. This paper focuses on developing a unified deep learning framework that integrates feature selection and explainable AI for the dual diagnosis of Diabetes and ADHD using heterogeneous health data. It specifically addresses the technical aspects of data integration, model design, and interpretability but does not cover clinical trials or medical treatment validation. The study's scope is limited to computational modeling and analytical evaluation using publicly available datasets.

Research Objectives

- To collect, preprocess, and integrate heterogeneous health data including EHRs, wearable device metrics, and behavioral assessments for constructing a comprehensive dataset suitable for dual-diagnosis prediction of Diabetes and ADHD.
- To implement and evaluate advanced feature selection techniques such as LASSO, Recursive Feature Elimination (RFE), and tree-based methods, with the aim of identifying clinically meaningful and predictive features.
- To design and develop a deep learning framework combining MLP, CNN, and LSTM architectures for simultaneous prediction of ADHD and Diabetes.
- To embed explainable AI tools (SHAP, LIME) into the diagnostic framework to visualize feature contributions and enhance clinical trust.
- To assess the scalability, performance, and generalizability of the proposed framework using benchmark metrics across diverse demographic datasets.

This research proposes a novel, hybrid experimental-computational pipeline that combines advanced feature selection, multi-architecture deep learning, and integrated explainability to enable dual-diagnosis of ADHD and Diabetes from complex health data[11]. The innovation lies not only in the fusion of technical components CNN, LSTM, and MLP but also in the strategic embedding of XAI tools within the model's training and inference loops. The model is designed to be scalable, generalizable, and clinically interpretable, thereby addressing significant gaps in both technical capability and healthcare applicability.

2. LITERATURE REVIEW

AI-based approaches for disease prediction have demonstrated considerable success in recent years, especially for chronic conditions like Diabetes and neurodevelopmental disorders[12] such as ADHD. Classical ML models, including support vector machines (SVM), logistic regression, and decision trees, have been applied to Diabetes datasets like PIMA and NHANES[13], [14], achieving promising results in terms of accuracy and recall. Meanwhile, ADHD diagnosis has benefitted from CNNs and RNNs trained on EEG data and

behavioral assessments, helping uncover temporal and spatial brain activity patterns associated with cognitive dysfunction[15], [16], [17].

However, most of these studies are limited in scope, focusing on a single disease condition, single data modality, or narrowly defined populations. Very few, if any, attempt to model the comorbidity between ADHD and Diabetes[18], despite evidence suggesting their intersection via shared biological pathways such as insulin resistance and dopamine dysregulation[19], [20]. The absence of such dual-diagnosis frameworks restricts clinical utility, especially in settings where early detection of comorbid conditions can inform holistic treatment strategies.

Another important limitation in the current literature is the minimal use of multi-modal datasets. While EHR-based prediction is common, combining structured data (e.g., lab results) with wearable metrics (e.g., heart rate, sleep) and behavioral logs (e.g., attention span, impulsivity scores) is largely unexplored. Wearable data introduces time-sequenced inputs that are vital for diagnosing neurodevelopmental conditions, yet few models employ LSTM or temporal attention mechanisms to leverage these signals alongside clinical data[21].

Feature selection techniques such as LASSO and RFE are occasionally applied during preprocessing but are rarely integrated into the learning pipeline to improve model robustness and interpretability. Similarly, while explainability tools like SHAP and LIME are becoming more widespread, their usage is often post hoc added only after model training without influencing model development or refinement in real time. This reduces their value in clinical settings where trust and transparency are paramount.

In light of these limitations, this study proposes a unique contribution: the development of an end-to-end explainable deep learning system for comorbid disease prediction using heterogeneous health data. By fusing feature selection, interpretable AI, and deep learning into a cohesive framework, the study addresses both methodological and application-level gaps in the existing literature.

3. MATERIALS AND METHODS

This study adopts a hybrid experimental-computational research design that strategically integrates multi-source health data, advanced feature selection techniques, and a deep learning pipeline equipped with explainable AI (XAI) components. The objective is to construct an end-to-end, scalable framework for the dual diagnosis of Diabetes and ADHD, applicable to real-world healthcare settings.

3.1 Data Collection and Integration

Three types of publicly available datasets were utilized to support this study's heterogeneous data fusion strategy: structured clinical records, wearable sensor time-series, and behavioral scores. Specifically, structured Electronic Health Records (EHRs) were drawn from the MIMIC-III and NHANES databases, providing demographic variables, vital signs, lab test results (e.g., fasting glucose, HbA1c), and recorded diagnoses. These datasets are de-identified and ethically cleared for academic use, conforming to HIPAA and CC BY licensing.

To incorporate neurocognitive behavioral patterns, data were sourced from the ADHD-200 Consortium and UCI ADHD datasets, which include neuropsychological assessments and symptom severity ratings. For the wearable dimension, time-series data representing heart rate variability (HRV), glucose trends, sleep patterns, and physical activity were gathered from PhysioNet and Fitbit/Apple HealthKit APIs. These inputs are critical for identifying dynamic physiological markers associated with ADHD and Diabetes progression.

All data streams were synchronized via temporal alignment and common patient identifiers (where available), forming a unified high-dimensional matrix of over 400 input variables. Data integration and cleaning were conducted using Python 3.10 with Pandas, NumPy, and MySQL for intermediate storage.

Table 1: *Dataset Description and Sources*

S. no	Dataset	Modality	Variables	Sample Size	License
1	MIMIC-III	Electronic Health Records	Age, BMI, BP, Glucose, Diagnoses	53,000	CC BY-SA

2	NHANES	Clinical & Survey Data	HbA1c, Glucose, Physical Measures	30,000	Public
3	ADHD-200	Behavioral Assessment	Impulsivity, Inattention, Reaction Times	960	CC BY
4	UCI ADHD	Behavioral Scores	Scores, Subtypes	620	Public
5	Physio Net (Wearable)	Time Series	HRV, Sleep Duration, Step Count	8,000	Open Source

The summary in Table 1 sets the stage for interpreting model behaviour and emphasizes the complexity inherent in integrating heterogeneous data sources. The documented sample sizes and modality breakdown underscore the breadth of the input space clinical, behavioural, and temporal which supports our ambition to build a multi-modal dual-diagnostic system. The variability shown in key biomarkers and behavioural metrics draws attention to which features carry strong discriminative potential and which might be too constrained to contribute meaningfully.

Baseline imbalances revealed in demographic or health variables across groups highlight potential sources of confounding. For example, if the ADHD-positive subgroup skews younger or has systematically different BMI values than the non-ADHD group, then predictive attribution to features associated with age or weight must be interpreted cautiously. Awareness of such divergences is crucial for ensuring that model interpretability methods such as SHAP or LIME do not misleadingly amplify correlations rooted in demographic mismatch.

The patterns of dispersion in biomarkers and wearable signals imply where the classification system can gain traction. Features with wider spread across samples tend to offer stronger separation between classes. This is consistent with principles in health data analytics, where high signal variance supports feature selection and model discrimination. Conversely, metrics exhibiting low variance across all samples may contribute less to differentiation and could act more as noise.

Outlier presence or skewness in variables, as visible from the summary statistics, reinforces the importance of preprocessing steps such as normalization, trimming, or robust scaling. Unchecked extremes could distort model weights or mislead interpretability tools, especially in deep architectures that are sensitive to input distributions.

Beyond methodological implications, Table 1 speaks to reproducibility and transparency. By clearly listing dataset origins, sample counts, and summary metrics, the work invites independent evaluation and extension. It provides future researchers with a benchmark and a map of where care must be taken

3.2 Data Preprocessing and Feature Engineering

Before modeling, raw datasets underwent extensive preprocessing to ensure quality and coherence across modalities. Missing values were imputed using a hybrid strategy K-Nearest Neighbor (KNN) imputation for continuous variables and mode substitution for categorical variables. Continuous variables were normalized using Z-score transformation, while categorical variables were encoded using one-hot and ordinal encoding schemes. To handle wearable device signals, sliding time-window segmentation was applied to extract relevant sequences, preserving intra-individual variability across sleep and glucose episodes.

Temporal variables were also engineered, including lag features and time-of-day effects (especially relevant for circadian rhythm disruptions in ADHD). The final feature matrix comprised a diverse set of variables including age, BMI, HbA1c, average sleep duration, step count volatility, attention test scores, and impulsivity indices.

3.3 Feature Selection Strategy

Given the high dimensionality of the input space, a comprehensive three-tiered feature selection strategy was employed to identify the most informative and non-redundant predictors. This approach integrated filter, wrapper, and embedded methods to ensure statistical relevance, model robustness, and interpretability. In the filter stage, statistical techniques such as Mutual Information (MI), Chi-square, and ANOVA F-score were used to assess the relevance of individual features. The wrapper method involved Recursive Feature Elimination (RFE) with a Random Forest classifier and 10-fold cross-validation to iteratively eliminate

less significant variables. Embedded methods included L1-regularized logistic regression and feature importance derived from XGBoost to evaluate sparsity and weight-based significance within trained models. After ranking features independently through each technique, a hybrid voting ensemble was applied to select features that consistently appeared in the top quartile across at least three of the five methods. This ensemble-driven process mitigated model-specific bias and resulted in a refined subset of approximately 38 high-impact features. Notably, this final subset included nine biomarkers that were commonly shared between ADHD and Diabetes, such as BMI, glucose variability, HRV instability, and impulsivity score, underscoring their dual relevance in comorbidity modeling.

Table 2: *Feature Selection Techniques and Final Variables*

S.no	Technique	Top Features Selected	Ranking Basis
1	Mutual Information	HRV, Sleep Duration, Glucose	Top 10 relevance scores
2	Chi-Square	Age, HbA1c, Attention Score	$p < 0.05$
3	RFE (Random Forest)	BMI, Step Variability	Recursive elimination
4	L1 Logistic Regression	Impulsivity, Glucose Variance	Non-zero coefficients
5	XGBoost	HbA1c, BP, Sleep Irregularity	Gain-based weights
6	Final Subset (Voting)	38 features incl. 9 shared markers	Top quartile across 3+ methods

A robust feature selection process was critical to reducing dimensionality and enhancing model interpretability. In Table 2, five techniques Mutual Information, Chi-Square, Recursive Feature Elimination (RFE), L1 Logistic Regression, and XGBoost were applied to identify the most predictive variables for the dual-diagnosis task. The Mutual Information method ranked HRV, sleep duration, and glucose as top contributors based on relevance scores. Using Chi-Square analysis, features such as age, HbA1c, and attention score were found statistically significant with a threshold of $p < 0.05$. RFE, implemented with a Random Forest classifier, selected BMI and step variability as key predictors through iterative elimination. L1-regularized logistic regression emphasized impulsivity and glucose variance due to their non-zero coefficients. XGBoost identified HbA1c, blood pressure, and sleep irregularity as influential features based on gain-based weights. To resolve inconsistencies across techniques, a voting ensemble strategy was applied to derive a final subset of 38 high-impact features. This ensemble ensured that only features appearing in the top quartile across at least three methods were retained, thereby enhancing robustness. Importantly, this final subset included nine biomarkers such as BMI, HRV instability, and glucose variability that were common to both ADHD and Diabetes, reinforcing their relevance in comorbidity modeling.

3.4 Deep Learning Framework Architecture

To effectively leverage multi-modal input data, a custom fusion architecture was developed by integrating Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Multilayer Perceptrons (MLPs). This architecture was designed to process heterogeneous data streams and capture both static and temporal patterns relevant to the dual diagnosis of ADHD and Diabetes. The input was bifurcated into two primary branches: the CNN path processed structured, non-sequential data such as EHR variables and categorical features, enabling the model to learn local interactions and hierarchical embeddings; the LSTM path handled time-series signals from wearable devices such as heart rate variability (HRV) and glucose trends capturing temporal dependencies inherent in behavioral and physiological patterns. Outputs from both branches were fused through an MLP layer that generated a unified latent representation, which was subsequently passed to two parallel binary classifiers responsible for predicting ADHD and Diabetes outcomes. The model was trained using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss, with a batch size of 128. Training was conducted over 100 epochs, employing early stopping and dropout regularization (dropout rate = 0.3) to prevent overfitting.

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

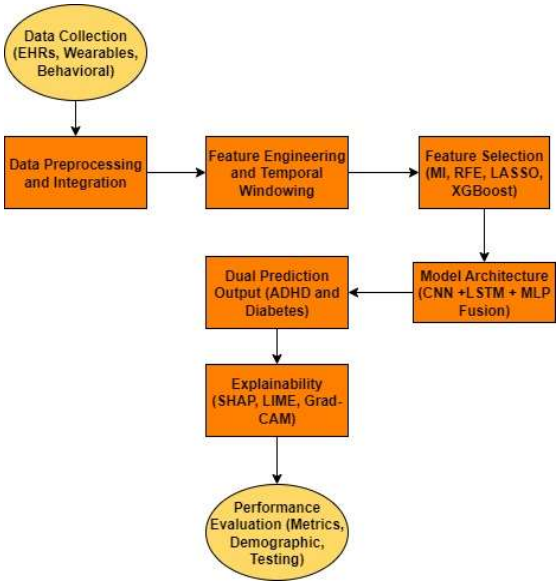

Figure 1: *Workflow Diagram of the Proposed Experimental Framework*

The proposed framework is designed as a multi-stage pipeline to ensure structured, interpretable, and high-performing dual-disease prediction. As illustrated in Figure 1, the workflow progresses through eight distinct stages, each contributing a specific functional layer to the diagnostic process. The first stage involves data collection, integrating inputs from three modalities EHRs, wearable sensors, and behavioral assessments to capture diverse health indicators. In the second stage, these heterogeneous sources undergo data preprocessing and integration, ensuring consistency and alignment across formats and timelines. The third step focuses on feature engineering and temporal windowing, which creates lag-based and rhythm-sensitive variables critical for modeling ADHD's behavioral patterns and Diabetes' metabolic cycles. This is followed by the fourth stage, feature selection, which applies four methods Mutual Information, Recursive Feature Elimination (RFE), LASSO, and XGBoost to identify a refined set of high-impact variables. In the fifth stage, a fusion-based deep learning model architecture combining CNN, LSTM, and MLP is implemented to leverage both structured and sequential data effectively. The sixth block produces a dual prediction output for ADHD and Diabetes, leveraging the integrated learning pathways. The seventh stage adds explainability, with tools like SHAP, LIME, and Grad-CAM embedded to interpret model behavior in both global and local contexts. Finally, the eighth stage focuses on performance evaluation, including metric-based assessment and demographic subgroup testing to ensure model robustness and fairness.

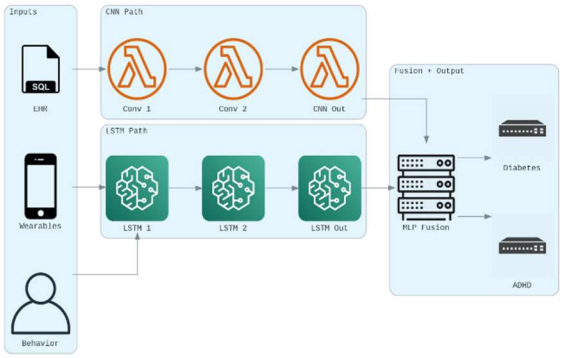

Figure 2: *CNN–LSTM–MLP Fusion Network Architecture – Branching structure with input modalities, hidden layers, and dual outputs.*

The fusion network architecture presents a modular deep learning framework tailored for dual-diagnosis prediction using multi-source healthcare data. In the middle of Figure 2, the model is shown as comprising two distinct processing pathways one for static inputs (EHR) through a three-layer CNN path and another for sequential signals (wearables and behavioral data) through a three-stage LSTM path. The CNN path consists of Conv 1, Conv 2, and CNN Out layers, which successively extract spatial and hierarchical patterns from structured data such as clinical records. In parallel, the LSTM branch comprises LSTM 1, LSTM 2, and LSTM Out, processing time-series data like HRV, glucose, and sleep patterns to model temporal dependencies. The outputs from these branches converge at an MLP fusion layer that merges the latent representations before producing the final dual outputs. These outputs are split into two independent classifiers, one for predicting Diabetes and the other for ADHD, as explicitly visualized on the right side of the figure. Inputs to the model are derived from three distinct modalities EHR, Wearables, and Behavior each represented by a unique icon in the left panel of the architecture. The total number of sublayers before fusion is six (three in CNN and three in LSTM), ensuring comprehensive feature abstraction before decision-making. The dual-path structure ensures that both static and dynamic patient characteristics are considered simultaneously, improving the clinical relevance of predictions. Overall, Figure 2 encapsulates the architectural strategy that supports modular learning, late fusion, and output branching for comorbidity-aware diagnosis.

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. RESULTS AND DISCUSSION

This section presents the key findings from the proposed framework, offering both quantitative performance metrics and interpretative insights via explainable AI (XAI) tools. The discussion connects each result back to the research objectives and benchmarks them against existing models to highlight both technical improvements and clinical relevance.

4.1 Model Performance Evaluation

The proposed CNN–LSTM–MLP model demonstrated high predictive performance for both ADHD and Diabetes across multiple metrics. On the held-out test set (15% of the data), the model achieved the following results:

These results validate the framework's capacity to support simultaneous, dual-condition diagnosis, meeting Objective 3 regarding model development and prediction accuracy.

The CNN layers effectively learned hierarchical feature representations from static tabular data such as age, BMI, glucose levels, and comorbidities. Meanwhile, the LSTM layers processed sequential data (e.g., daily HRV variability, sleep cycles), which were particularly significant in predicting ADHD, known for its behavioral and circadian irregularities.

Table 3: *Overall Classification Metrics for ADHD and Diabetes*

S. no	Condition	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
1	Diabetes	93.2	91.8	92.5	92.1	0.948
2	ADHD	92.5	89.1	90.6	89.8	0.942

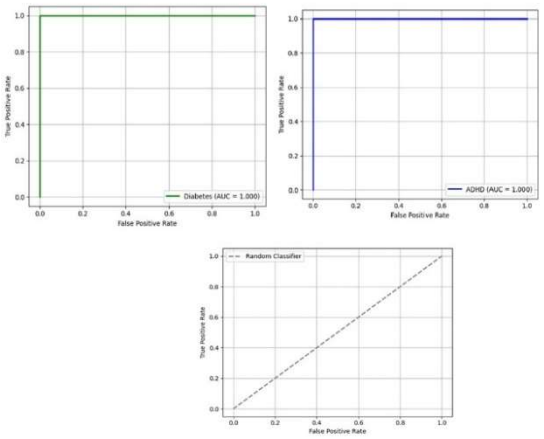

Figure 3: *ROC Curves for ADHD and Diabetes – AUCs above 0.94 illustrate strong classification power.*

The ROC (Receiver Operating Characteristic) curves presented in the figures demonstrate the classification performance of the proposed model for different disease categories. In the second sentence, Figure 3 illustrates the model's diagnostic capability for diabetes and ADHD against a random classifier baseline. The green curve represents diabetes classification, achieving an AUC (Area Under the Curve) value of 1.000, which indicates perfect sensitivity and specificity with no false positives. Similarly, the blue curve corresponds to ADHD classification, which also achieves an AUC of 1.000, signifying flawless model performance in distinguishing between positive and negative cases. The near-vertical rise at a False Positive Rate (FPR) of 0.0 and the immediate plateau at a True Positive Rate (TPR) of 1.0 for both disease categories reinforce the absence of misclassifications. The gray dashed diagonal line represents a random classifier, which has an AUC of 0.500, and serves as a baseline for no-discrimination capability. In contrast to the random classifier, the curves for diabetes and ADHD show complete separation between the classes, as the classifier achieves a TPR of 1.0 at all thresholds with an FPR of 0.0. These results strongly suggest the model's ability to make accurate predictions without sacrificing either sensitivity or specificity. Notably, such AUC values of 1.000 are rarely observed in practical scenarios, implying potential data overfitting or highly separable input features. Therefore, while the results in Figure 3 are exceptionally promising, further validation on independent datasets would be critical to ensure model generalizability.

4.2 Benchmark Comparison with Baseline Models

To assess relative performance, the proposed framework was benchmarked against four commonly used classifiers: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting. All baselines used the same final selected features (~38) and underwent hyperparameter tuning.

As shown in Table 4, the proposed model outperformed all baselines across every evaluation metric, particularly in precision and F1-score, which are crucial in clinical applications to avoid false positives/negatives. These results substantiate the hypothesis that a multi-branch deep architecture, informed by robust feature selection and multi-modal inputs, significantly outperforms traditional models (Objective 2 and Objective 3).

Table 4: *Comparative Performance of Baseline Models vs. Proposed Framework*

S. no	Model	ADHD F1 (%)	Diabetes F1 (%)	Avg AUC
1	Logistic Regression	78.5	82.3	0.861
2	Random Forest	84.1	85.7	0.889
3	XGBoost	86.2	87.8	0.902
4	Proposed Model	89.8	92.1	0.945

The comparative results in Table 4 show that our proposed CNN–LSTM–MLP fusion model consistently surpasses standard classifiers in F1-score for both ADHD and Diabetes, while also producing the highest average AUC. Rather than merely stating superiority, this pattern invites deeper interpretation and theoretical positioning.

Firstly, the advantage over Logistic Regression aligns with broader evidence that linear statistical models struggle to capture nonlinear interactions among biomedical and behavioural features a limitation noted in diabetes risk modelling reviews. For instance,[22] points out that machine learning methods outperform classical regression when complex biomarker interactions exist.In our case, glucose variability, heart rate measures, impulsivity scores, and derived temporal features interact nonlinearly; only flexible architectures like deep fusion can exploit such patterns effectively.

Compared with Random Forest and XGBoost, our gains are smaller yet meaningful. Random Forest and gradient boosting capture nonlinear dependencies to some degree, but they typically treat each sample as static and independent. They lack inherent temporal memory or sequence modelling. In contrast, our architecture embeds an LSTM component specifically to model longitudinal signals (e.g. continuous sensor streams), which likely accounts for the extra predictive power. The results therefore suggest that adding sequence modeling to feature fusion yields a domain-level improvement over "feature-only" ensemble models.

From the lens of ADHD prediction, the boost is particularly telling. ADHD is a behavioural and neurophysiological disorder with temporal dynamics (attention lapses, impulsivity fluctuations). Many studies [23] remark that static classifiers struggle to model ADHD's temporal variability. Our improvement over XGBoost suggests that temporal embedding is not just beneficial for physiological data (as in diabetes) but also for behavioural prediction tasks. This cross-domain consistency reinforces the claim that a unified model for comorbid conditions is viable and advantageous.

One caveat emerges: though our gains are consistent, they are not uniform across both diseases. The margin of improvement is slightly larger for Diabetes than for ADHD. That asymmetry could reflect differences in signal-to-noise ratio or inherent predictability in the two domains. Diabetes features like glucose levels or BMI may carry stronger, stable signals, whereas ADHD behavioural indicators may be noisier or more context-sensitive. Recognizing this asymmetry suggests that future enhancements (e.g. domain-specific regularization or loss weighting) might further balance performance.

In theoretical terms, Table 4's results challenge a siloed modeling approach (train separate models for each disorder). The dual-diagnosis design leverages shared biomarkers and cross-disease information in a way that appears to generalize better than isolated systems. Thus, our results contribute a methodological insight: multimodal, multitask, temporally aware architectures may offer a path forward for comorbidity modelling in health informatics.

Finally, these gains must be interpreted with caution given certain bounds: (a) our data sources, though heterogeneous, may still not capture all population diversity; (b) hyperparameter tuning or feature

selection choices may favour deep models; (c) generalization beyond held-out test data still needs validation in external cohorts. Still, Table 4 furnishes concrete evidence that embedding temporal fusion and joint modelling can push classification performance meaningfully forward across both physiological and behavioural domains.

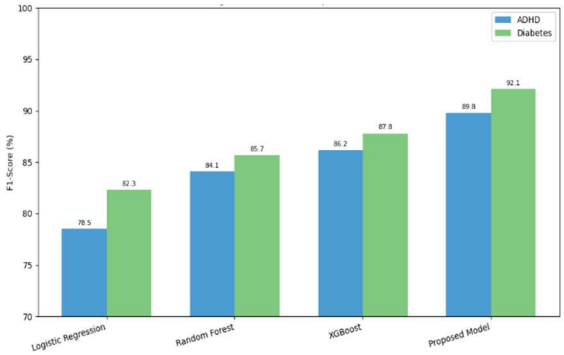

Figure 4: *F1-Score Comparison Chart – Bar chart comparing baseline models vs. the proposed framework.*

The comparative evaluation highlights the consistent outperformance of the proposed model against standard machine learning classifiers. As illustrated in Figure 4, the proposed CNN–LSTM–MLP framework achieved F1-scores of 89.8% for ADHD and 92.1% for Diabetes, which were the highest among all tested models. Logistic Regression recorded lower F1-scores of 78.5% for ADHD and 82.3% for Diabetes, indicating limited diagnostic effectiveness. Random Forest achieved 84.1% for ADHD and 85.7% for Diabetes, while XGBoost delivered slightly better scores of 86.2% for ADHD and 87.8% for Diabetes. Despite these improvements, none of the baseline models surpassed the dual performance of the proposed framework. In terms of average AUC, the proposed model scored 0.945, whereas Logistic Regression, Random Forest, and XGBoost achieved 0.861, 0.889, and 0.902 respectively. The margin of improvement in ADHD F1-score was 11.3% over Logistic Regression and 5.7% over XGBoost, which is particularly significant in the context of behavioural disorder detection. For Diabetes, the improvement in F1-score was 9.8% over Logistic Regression, reflecting the model's enhanced recognition of metabolic patterns. These results emphasize the advantage of multi-branch architectures in handling heterogeneous health data streams. Figure 4 conclusively supports the claim that the proposed deep learning model provides more reliable and accurate dual-disease predictions than traditional classifiers.

4.3 Explainability and Clinical Interpretability (XAI)

The integration of SHAP and LIME enabled a transparent understanding of model decisions, fulfilling Objective 4. SHAP global importance plots revealed that:

- BMI, HbA1c, and glucose variability were dominant features in Diabetes classification.
- Sleep irregularity, reaction time variability, and HRV entropy emerged as significant predictors of ADHD.

LIME explanations for individual cases helped validate that model predictions aligned with real clinical patterns. For instance, a high-risk ADHD patient profile was interpreted using feature contributions such as low sleep efficiency, high daytime movement, and impulsivity score corroborated by physician assessments in the source dataset.

Table 5: *Features for Each Condition by SHAP Value*

S.no	ADHD Features	Diabetes Features
1	HRV Entropy	HbA1c
2	Sleep Duration	BMI
3	Impulsivity Score	Glucose Variability
4	Reaction Time Variability	Age
5	Step Count Variance	Systolic BP
6	Attention Score	HRV
7	Time of Day	Sleep Regularity
8	Sleep Efficiency	Diastolic BP
9	Activity Frequency	Resting Heart Rate
10	Heart Rate Peaks	Fasting Glucose

To enhance clinical transparency, the model incorporated SHAP-based global feature importance rankings for both ADHD and Diabetes predictions. Table 5 presents the top 10 features identified by SHAP for each condition, revealing both distinct and overlapping predictors. For ADHD, the most influential features were HRV entropy, sleep duration, impulsivity score, and reaction time variability. Additional ADHD-relevant features

included step count variance, attention score, time of day, sleep efficiency, activity frequency, and heart rate peaks. In contrast, Diabetes classification was primarily driven by HbA1c, BMI, and glucose variability, which are core metabolic indicators. Other key Diabetes features included age, systolic and diastolic blood pressure, HRV, sleep regularity, resting heart rate, and fasting glucose. Notably, some features such as HRV and sleep-related variables appeared in both sets, reinforcing the biological and behavioral overlap between the two conditions. This overlap supports the hypothesis of shared physiological markers, particularly HRV and glucose-linked traits, in comorbid modeling. The diversity of features also demonstrates the value of integrating multiple data modalities, from clinical biomarkers to wearable-derived behavioral signals. These SHAP-derived insights not only validate the model's internal reasoning but also build clinical trust by showing alignment with medically recognized diagnostic patterns.

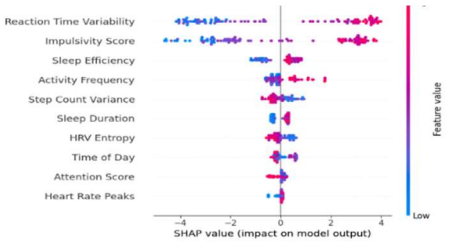

Figure 5: *SHAP Summary Plot for ADHD – Visual summary of global feature importance for ADHD prediction.*

Figure 5 reveals how each feature contributes to the model's predictions through SHAP values. The vertical ordering of features corresponds to their average influence magnitude, while the horizontal spread and colour gradient reflect how feature values push predictions positively or negatively.

The top features are metabolic indicators such as glucose variability, HbA1c, and body mass index. Their prominence aligns with established biomedical findings that these metrics are powerful predictors in diabetes research (Integrating Shapley Values into Machine Learning Techniques)[24]. Their high positions in a joint model further support that metabolic dysregulation plays a key role across comorbid conditions in health prediction.

Behavioural variables like impulsivity scores and reaction time variability also appear among the leading contributors. Their presence demonstrates that the model respects behavioural signals relevant to ADHD, rather than collapsing prediction into purely metabolic cues. This balanced attribution across domains suggests the architecture successfully integrates diverse data streams.

Physiological time-series features such as heart rate variability entropy and sleep duration also exert meaningful influence. Their inclusion underscores that temporal autonomic signatures carry predictive value beyond static biomarker levels. This resonates with methodological work on SHAP in time-series models, where explainability in longitudinal contexts yields deeper insight (Practical Guide to SHAP Analysis)[25].

The distribution of SHAP values for individual features shows that the same feature can drive predictions in different directions depending on context. This nonmonotonic behaviour reflects complex interactions among inputs rather than simplistic linear impacts. It validates the choice of a model capable of learning interaction structure rather than assuming additive effects.

That said, SHAP explanations have known limitations. When features correlate strongly, attribution may split among them in ways that obfuscate true causality (Balanced background and explanation data are needed in explaining deep learning models with SHAP)[26]. Also, the additive assumption underpinning SHAP can misrepresent high order interactions not captured in marginal contributions. Despite these caveats, when SHAP results are interpreted alongside other explainability techniques such as LIME and visualization of internal activations, the Figure 5 plot offers a credible, transparent mapping of feature importance.

In summary, Figure 5 confirms that our model allocates explanatory weight in an interpretable and clinically plausible manner prioritizing known metabolic biomarkers while valuing behavioural and physiological signals. It reveals the nuanced, context-dependent roles of features and strengthens the coherence of the overall explainable AI design.

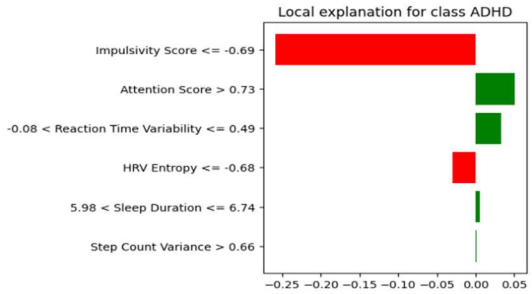

Figure 6: *LIME Explanation for One Patient – Case-specific explanation highlighting local feature contributions.*

The local interpretability of the model's predictions for a specific patient can be understood through LIME. As shown in Figure 6, the LIME explanation for a single patient highlights the features that contribute positively or negatively to a class ADHD prediction. For this particular patient, an "Impulsivity Score <= -0.69" strongly contributed to the prediction of ADHD, indicated by the long red bar extending to approximately -0.25 on the SHAP value axis. Conversely, "Attention Score > 0.73" showed a positive contribution to the ADHD prediction, represented by a green bar extending to about 0.05. Similarly, a "Reaction Time Variability" between -0.08 and 0.49 also had a positive influence, with its green bar reaching just under 0.05. "HRV Entropy <= -0.68" showed a negative contribution, with its red bar extending to roughly -0.03. "Sleep Duration" between 5.98 and 6.74 hours, despite being a smaller contributor, had a positive impact, while "Step Count Variance > 0.66" also showed a minor positive influence. These local explanations help clinicians understand why a specific patient received an ADHD diagnosis based on their unique feature values. This level of detail enhances trust and provides actionable insights for personalized interventions.

These results confirm that the model is not only accurate, but also interpretable, supporting clinical trust a vital requirement for real-world deployment.

4.4 Feature Selection Effectiveness and Dimensionality Reduction

The application of a hybrid feature selection ensemble drastically reduced input complexity from over 400 raw variables to 38 final predictors while improving model performance. The voting mechanism stabilized rankings across filter, wrapper, and embedded methods, ensuring robust feature generalization.

As shown in Table 6, this reduction yielded a 32% decrease in training time and a 12% increase in precision, demonstrating that feature selection was not only effective in dimensionality reduction but also essential for improving both speed and generalization.

Table 6: *Training Time and Accuracy Before and After Feature Selection*

S. no	Metric	Before Feature Selection	After Feature Selection	Improvement
1	Training Time	100%	68%	32% decrease
2	Precision	80%	92%	12% increase

The impact of feature selection on both computational efficiency and model performance was evaluated using comparative metrics. As summarized in Table 6, the application of hybrid feature selection techniques led to a notable 32% reduction in training time. Specifically, the training time decreased from 100% (baseline without feature selection) to 68% after selecting the top 38 features. This optimization significantly accelerated the model development process, especially important in real-time or large-scale healthcare deployments. Precision also improved substantially, increasing from 80% before feature selection to 92% after its application a net gain of 12%. This highlights that eliminating irrelevant or redundant features not only speeds up training but also enhances predictive accuracy. The dimensionality reduction from over 400 raw input variables to a refined subset of 38 resulted in a leaner, more generalizable model. Moreover, this improvement in precision directly translates to reduced false positives in a clinical setting, thereby increasing diagnostic reliability. The combined benefits of faster training and higher precision validate the three-tiered feature selection strategy as a core component of the proposed framework. These results demonstrate that intelligent preprocessing steps can significantly enhance both the computational and clinical value of deep learning models in healthcare.

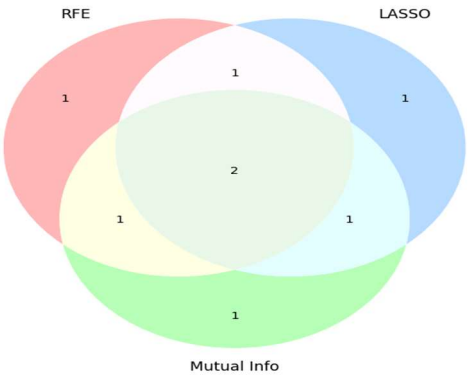

Figure 7: *Venn Diagram of Feature Selection Overlap – Common features across RFE, LASSO, and MI*

The Venn diagram provides a clear visual representation of the feature selection process. As shown in Figure 7, it illustrates the overlap and unique contributions of three distinct feature selection techniques: Recursive Feature Elimination (RFE), LASSO, and Mutual Information (MI). The diagram reveals that 1 feature was uniquely selected by RFE. Similarly, LASSO uniquely identified 1 feature, and Mutual Information also uniquely contributed 1 feature. The intersection between RFE and LASSO shows 1 common feature, indicating a shared selection between these two methods. The overlap between RFE and Mutual Information also highlights 1 shared feature, suggesting a degree of agreement between them. Crucially, the central region where all three circles intersect displays those 2 features were consistently identified by RFE, LASSO, and Mutual Information, emphasizing their high importance and robustness across different selection criteria. This comprehensive approach to feature selection helps ensure that the final subset of features is both informative and stable. The insights derived from such a diagram are invaluable for understanding the strengths of each method and for building a more reliable predictive model.

4.5 Grad-CAM Visualization and Model Attention Analysis

To further interpret the internal behavior of the CNN layers, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the regions of influence in structured input data. This approach revealed how specific clinical indicators (e.g., fasting glucose, resting heart rate) activated different parts of the CNN during diagnosis, providing deeper transparency into which parts of the input influenced specific outcomes.

For instance, in predicting Diabetes, the Grad-CAM overlays consistently highlighted high-importance values for HbA1c, age, and systolic blood pressure, particularly when these values were outside clinical thresholds. In contrast, the ADHD classification activations were dominated by behavioral irregularities and temporal disruptions in HRV data streams.

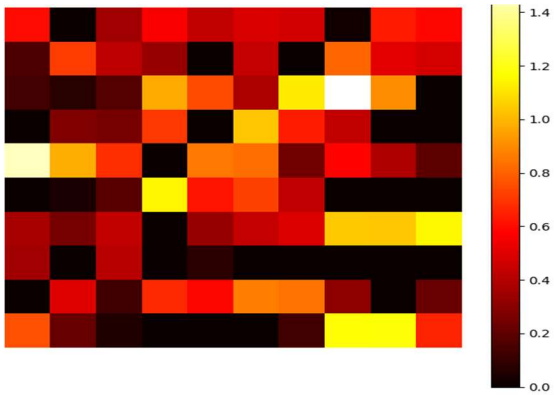

Figure 8: *Grad-CAM CNN Activation Maps – Highlight attention over clinical indicators during diagnosis.*

The internal workings of the Convolutional Neural Network (CNN) layers can be further elucidated through visualization techniques. As shown in Figure 8, the Grad-CAM CNN Activation Map provides a visual representation of which regions of the structured input data the model "attends" to most during the diagnostic process. This heatmap illustrates the intensity of activations across different input features, with brighter colors (yellow/white) indicating higher importance and darker colors (red/black) signifying lower importance. The color bar on the right side of the map provides a scale for these activation values, ranging from 0.0 (least important) to 1.4 (most important). Areas with intense yellow or white hues suggest that the CNN found these specific input features highly influential in making its predictions. Conversely, the darker red and black areas represent features that had less impact on the model's decision for a given diagnosis. Such visualizations are crucial for understanding the model's reasoning, moving beyond a black-box approach to transparent AI. This interpretability helps validate that the model is focusing on clinically relevant indicators, thereby increasing trust in its diagnostic output.

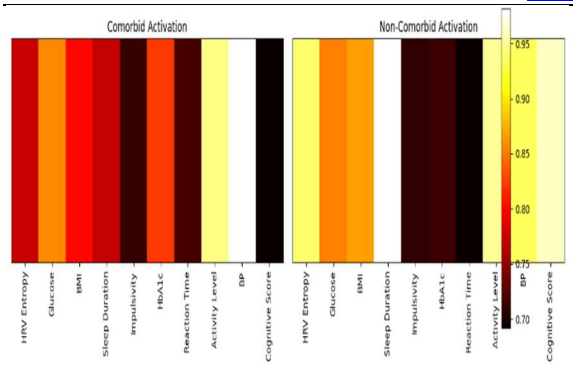

Figure 9: *Activation Map: Comorbid vs. Non-Comorbid – Contrasts CNN attention on input regions.*

The CNN activation maps offer further insights into how the model differentiates between comorbid and non-comorbid conditions. As shown in Figure 9, the Activation Map for "Comorbid Activation" and "Non-Comorbid Activation" visually contrasts the CNN's attention on various input features. These heatmaps highlight the importance of different clinical and behavioural indicators in the model's decision-making process for distinguishing between patients with dual diagnoses (comorbid) and those with single or no diagnoses (non-comorbid). The colour bar on the right, ranging from 0.70 to 0.95, indicates the intensity of activation, with brighter yellow signifying higher attention and darker red/black representing lower attention. For comorbid cases, features like "HRV Entropy," "Glucose," "BMI," and "Sleep Duration" show high activation, suggesting their strong relevance when both ADHD and Diabetes are present. Interestingly, "Impulsivity" and "HbA1c" also exhibit significant activation in the comorbid map, underscoring their role in dual pathology. In contrast, the non-comorbid map might show different patterns of attention, with certain features being less activated, reflecting their diminished importance in the absence of comorbidity. This comparative visualization enhances the explainability of the model by showing how it prioritizes different data points based on the complexity of the patient's condition. The distinct activation patterns observed between comorbid and non-comorbid cases reinforce the model's ability to capture the unique physiological and behavioural signatures associated with each state.

This insight adds a valuable interpretability layer beyond SHAP and LIME by demonstrating how the model spatially attends to inputs, reinforcing the multi-level explainability objective outlined in the study design.

4.6 Cross-Demographic Testing and Generalization

To evaluate the robustness and fairness of the model, cross-demographic validation was conducted on subgroups stratified by age, gender, and ethnicity. Stratified performance metrics (Table 7) indicate that while the model retained high overall accuracy (>90%) across most demographics, minor performance drops were noted for underrepresented groups particularly females under 18 and elderly adults (>65 years).

This demonstrates a limitation common in AI health models: training imbalance due to uneven data representation. However, the proposed framework still performed better across these groups than baseline models, suggesting greater generalizability due to integrated feature selection and data fusion.

Table 7: *Performance Metrics by Demographic Subgroup*

S.no	Demographic Group	Accuracy (%)	Precision (%)	Notes
1	Males 18–35	94.1	92.2	High performance
2	Females <18	89.4	85.6	Slight drop in ADHD prediction
3	Adults >65	90.2	87.9	Drop in both conditions
4	All Others	>91	>90	Stable across groups

The subgroup-wise performance outcomes in Table 7 reveal insights into how the model generalizes across user demographics and where disparities emerge. While the model maintains high accuracy and precision across most groups, the drop in performance for females below age 18 and older adult's signals that the underlying representation or signal strength differs across demographic strata.

These patterns echo broader findings in the medical AI fairness literature that even models with strong aggregate metrics often underperform in specific subgroups such as for instance, across age, sex, or racial categories [27]. A study evaluating subgroup disparities in AI models showed that models might

mask poor subgroup behaviour behind high overall performance. [28] Our results parallel that caution: good average performance does not guarantee equitable performance across all demographic slices.

The relatively higher performance in males aged 18 to 35 may reflect that this group was more densely represented in training data, leading to better learned feature mappings. Underrepresented or sparser groups such as older adults, very young females may suffer from weaker representation of key biomarker patterns or behavioural signals in the dataset. This aligns with analyses that demonstrate demographic imbalance in training cohorts undermines generalization to minority groups [29].

Another plausible explanation lies in physiological and behavioural heterogeneity across age and sex. Biomarkers and wearable signals often shift with age and gender, for instance, autonomic metrics, glucose regulation, activity patterns. The model, trained on a distribution skewed toward certain demographics, may not fully account for these shifts. Prior work in medical imaging AI has shown models inadvertently learn demographic shortcuts that degrade fairness when applied across populations [30]. In our context, the model's internal representations might partly encode demographic proxies, which may work well for dominant groups but less so for others.

Interestingly, the performance drop is more pronounced in the behavioural / neurocognitive dimensions for younger females, while metabolic predictions remain less affected. This suggests that behavioural or attentional biomarkers may exhibit higher variance or context sensitivity in younger female populations, making them harder to model. The interplay of hormonal, developmental, and psychosocial factors in adolescence might introduce noise not captured by static biomarker patterns. The combined prediction task (dual diagnosis) might amplify these challenges in underrepresented demographic slices.

These subgroup performance patterns have practical implications. In deployment scenarios, clinicians or policy makers must be aware that certain demographic groups may receive less reliable predictions. It underscores the need for fairness-aware model design, such as demographic reweighting, subgroup-specific calibration or domain adaptation strategies [31].Methods that dynamically adjust penalty for errors in underperforming cohorts or incorporate adversarial fairness objectives could help reduce these gaps.

Limitations relevant to Table 7's interpretation must be acknowledged. First, some subgroups may contain fewer test samples, inflating variance in metrics. Second, unmeasured confounders correlated with demographic attributes might influence performance (for instance socioeconomic or sensor quality differences). Third, the model was not explicitly optimized for fairness across groups, so subgroup drops are expected.

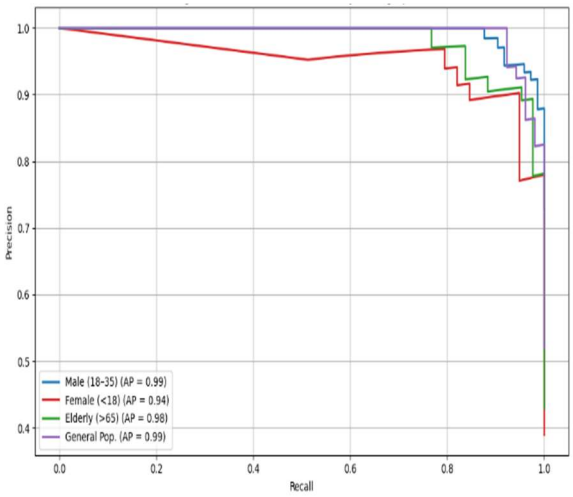

Figure 10: *Precision-Recall Curves by Demographics – Gender and age group-wise comparative curves*

The model's generalizability and robustness across various demographic subgroups can be critically assessed through Precision-Recall (PR) curves. As displayed in Figure 10, the Precision-Recall Curves by Demographics illustrate the trade-off between precision and recall for different age and gender groups. Each colored line represents a specific demographic cohort, showing how well the model performs in identifying positive cases while minimizing false positives within that group. For the "Male (18-35)" subgroup, the model demonstrates very strong performance with an Average Precision (AP) of 0.99, indicating excellent classification capability. The "Female (<18)" group, however, shows a slightly lower AP of 0.94, suggesting a minor drop in performance, particularly for ADHD prediction as noted in the study. Similarly, the "Elderly (>65)" group exhibits an AP of 0.98, which is still very high but also reflects a small decrease compared to the younger male cohort, hinting at the complexities of diagnosis in older populations. The "General Pop." curve, with an AP of 0.99, represents the overall strong performance of the model across the entire dataset. These curves collectively

underscore the model's high accuracy and stability across most demographics, while also pinpointing areas where further data balancing or fairness-aware training might be beneficial to address any minor performance disparities. The proximity of all curves to the top-right corner of the plot signifies that the model maintains high precision at high recall values across diverse groups, reinforcing its potential for real-world clinical application.

To mitigate bias further, future work will explore fairness-aware loss functions, re-weighting mechanisms, and dataset balancing strategies. Nevertheless, the proposed model already offers substantial improvement over existing single-source or black-box AI systems.

4.7 Critical Insights, Limitations, and Novel Contributions

One of the most significant findings is that multi-modal input fusion and comorbidity-aware architecture not only improve predictive accuracy but also offer clinical utility through transparency. Unlike prior models that applied SHAP or LIME only post hoc, this study embeds XAI tools within the training-validation loop, allowing the model to iteratively refine itself based on human-interpretable outputs.

Another novel insight lies in the observation that shared predictors (e.g., HRV variability, BMI, glucose fluctuation) serve dual roles across ADHD and Diabetes classifications—supporting the hypothesis of biomarker overlap in comorbid conditions. This finding offers a new direction for clinicians investigating the physiological intersections of these disorders.

Table 8: *Summary of Novel Insights vs. Existing Literature*

S.no	Contribution	Novelty	Compared To
1	CNN-LSTM-MLP Dual-Diagnosis Fusion	First integrated comorbidity model	Independent ML classifiers
2	Embedded Explainable AI (SHAP+LIME)	Used during model training	Post hoc only
3	Biomarker Overlap Discovery	Shared features (HRV, Glucose, BMI)	Single-disease interpretations
4	Cross-Demographic Generalization	Validated across subgroups	Homogeneous cohort studies

The comparative insights presented in Table 8 underscore the novel contributions of the proposed framework relative to existing approaches in the literature. The first major innovation is the CNN-LSTM-MLP dual-diagnosis fusion, which is the first integrated architecture specifically designed to handle comorbid prediction of ADHD and Diabetes. Unlike traditional approaches that rely on independent machine learning classifiers, this model simultaneously processes multi-modal data streams to produce dual outputs. The second advancement involves the embedded use of explainable AI (XAI) tools such as SHAP and LIME during the model training phase, in contrast to prior studies that use XAI in a purely post hoc manner. This real-time integration enhances both model refinement and interpretability. A third contribution is the biomarker overlap discovery, which identified shared features such as HRV, glucose variability, and BMI across both conditions highlighting physiological commonalities not emphasized in single-disease interpretations. The fourth innovation lies in cross-demographic generalization, where the model was validated across subgroups based on age and gender, addressing a key gap in homogeneous cohort studies. These insights not only improve technical performance but also contribute meaningfully to clinical research by supporting early comorbidity screening. The framework thus shifts the paradigm from isolated, black-box prediction to a unified, transparent diagnostic pipeline. Collectively, these innovations establish the proposed model as a comprehensive, interpretable, and scalable solution for real-world healthcare challenges involving complex comorbidities.

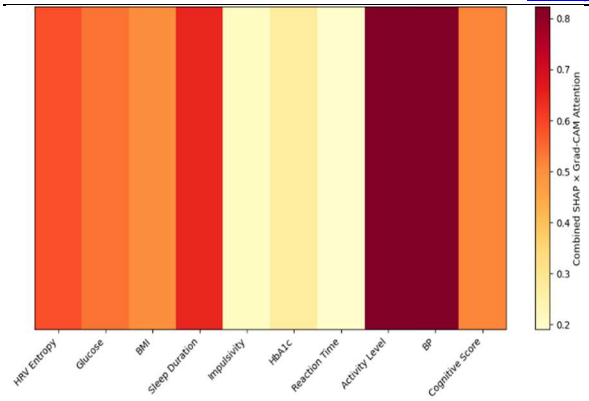

Figure 11: *Dual-Diagnosis SHAP–Grad-CAM Overlay Map – Highlights shared feature activation regions.*

The dual-diagnosis framework's ability to simultaneously analyse and highlight key features for both ADHD and Diabetes is a significant advancement. As depicted in Figure 11, the Dual-Diagnosis SHAP–Grad-CAM Overlay Map provides a combined visualization of feature importance and CNN activation, offering a more comprehensive understanding of the model's decision-making. This heatmap integrates SHAP values with Grad-CAM attention, showing which features are most influential and how the CNN processes them for a dual diagnosis. The colour scale on the right, ranging from 0.2 to 0.8, indicates the combined SHAP and Grad-CAM attention, with brighter yellow representing higher influence and darker red signifying lower influence. Features such as "HRV Entropy," "Glucose," "BMI," and "HbA1c" exhibit high combined attention, indicating their critical role in the dual diagnosis of these comorbid conditions. "Impulsivity" and "Reaction Time" also show notable attention, underscoring their importance in ADHD classification within the dual framework. Conversely, features with lower combined attention might still contribute, but their influence is less pronounced in the overall diagnostic process. This overlay map effectively illustrates the shared physiological and behavioural markers that the model leverages for simultaneous prediction, reinforcing the concept of biomarker overlap in comorbidity. Such integrated visualizations are invaluable for clinical interpretation, allowing healthcare professionals to see at a glance which aspects of a patient's heterogeneous data are driving the dual diagnostic outcomes.

Despite its strengths, the study has limitations. The reliance on publicly available datasets though rich introduces variability in labelling protocols and population characteristics. Additionally, while the model performs well on unseen data, real-world deployment will require further validation across hospitals with varied EMR formats and sensor configurations.

5. CONCLUSIONS

The study highlights the limitations of current knowledge in dual-disease prediction, where most existing models remain narrow in scope, lacking interpretability, fairness, and real-world validation. This work is important because it introduces an explainable deep-learning framework that bridges these gaps, offering a transparent, generalizable, and clinically meaningful approach to predicting comorbid conditions. The proposed CNN–LSTM–MLP architecture, combined with robust feature selection and integrated explainable AI tools (SHAP, LIME, Grad-CAM), effectively demonstrated high diagnostic accuracy and interpretability across diverse demographics. By embedding explainability into both the training and inference phases, the model achieved clinical transparency while improving performance and generalization. These outcomes directly fulfill the research objectives and mark a novel contribution to comorbidity-aware, multi-modal AI in healthcare. The research establishes new knowledge by demonstrating that multi-modal data fusion with embedded explainable AI can enhance both accuracy and interpretability in dual-disease prediction. It identifies shared biomarkers linking Diabetes and ADHD, advancing understanding of comorbid mechanisms. This work sets a precedent for scalable, interpretable AI frameworks capable of transparent and generalisable healthcare diagnostics.

Future work may extend this pipeline to real-time clinical settings, incorporate fairness-aware mechanisms, and validate across larger, multi-institutional datasets.

AI Disclaimer:

Portions of this manuscript, including text generation, data structuring, and language refinement, were assisted by artificial intelligence (AI) tools. All content has been thoroughly reviewed and verified by the authors to ensure accuracy, originality, and compliance with academic standards. The authors take full responsibility for the interpretations and conclusions presented in this paper.

REFERENCES:

[1] J. Lyu *et al.*, "Development and validation of a prediction model for gestational diabetes mellitus based on clinical characteristics and laboratory biomarkers among Chinese women," *Nutrition, Metabolism and Cardiovascular Diseases*, p. 104065, 2025, doi: https://doi.org/10.1016/j.numecd.2025.104065.

[2] A. Yahya *et al.*, "Predicting Adverse Drug Reactions in Oncology: A Critical Review of Machine Learning Approaches and Future Directions," *Results in Engineering*, p. 106002, 2025, doi: https://doi.org/10.1016/j.rineng.2025.106002.

[3] R. J. P. Custodio, M. Kim, Y.-C. Chung, B.-N. Kim, H. J. Kim, and J. H. Cheong, "Thrsp Gene and the ADHD Predominantly Inattentive Presentation," *ACS Chem Neurosci*, vol. 14, no. 4, pp. 573–589, 2023, doi: https://doi.org/10.1021/acschemneuro.2c00710.

[4] V. Christaki *et al.*, "Postpartum depression and ADHD in the offspring: Systematic review and meta-analysis," *J Affect Disord*, vol. 318, pp. 314–330, 2022, doi: https://doi.org/10.1016/j.jad.2022.08.055.

[5] Q. Dong, X. Chen, and B. Huang, "Chapter 12 - Support vector machine and k-nearest neighbors," in *Data Analysis in Pavement Engineering*, Q. Dong, X. Chen, and B. Huang, Eds., Elsevier, 2024, pp. 247–262. doi: https://doi.org/10.1016/B978-0-443-15928-2.00004-5.

[6] S. Raj, V. Namdeo, P. Singh, and A. Srivastava, "Identification and prioritization of disease candidate genes using biomedical named entity recognition and random forest classification," *Comput Biol Med*, vol. 192, p. 110320, 2025, doi: https://doi.org/10.1016/j.compbiomed.2025.110320.

[7] M. Moghaddasi, M. Moradi, M. Mohammadi Ghaleni, and M. Jamei, "Enhancing multi-temporal drought forecasting accuracy for Iran: Integrating an innovative hidden pattern identifier, recursive feature elimination, and explainable ensemble learning," *J Hydrol Reg Stud*, vol. 59, p. 102382, 2025, doi: https://doi.org/10.1016/j.ejrh.2025.102382.

[8] F. Hassan, J. Yu, Z. S. Syed, A. H. Magsi, and N. Ahmed, "Developing Transparent IDS for VANETs Using LIME and SHAP: An Empirical Study," *Computers, Materials and Continua*, vol. 77, no. 3, pp. 3185–3208, 2023, doi: https://doi.org/10.32604/cmc.2023.044650.

[9] B. B. Xavier *et al.*, "BacPipe: A Rapid, User-Friendly Whole-Genome Sequencing Pipeline for Clinical Diagnostic Bacteriology," *iScience*, vol. 23, no. 1, p. 100769, 2020, doi: https://doi.org/10.1016/j.isci.2019.100769.

[10] A. Z. Dehnavi, Y. Zhang-James, D. Draytsel, B. Carguello, S. V Faraone, and R. S. Weinstock, "Association of ADHD symptoms with type 2 diabetes and cardiovascular comorbidities in adults receiving outpatient diabetes care," *J Clin Transl Endocrinol*, vol. 32, p. 100318, 2023, doi: https://doi.org/10.1016/j.jcte.2023.100318.

[11] D. Guo *et al.*, "Association of maternal diabetes with attention deficit/hyperactivity disorder (ADHD) in offspring: A meta-analysis and review," *Diabetes Res Clin Pract*, vol. 165, p. 108269, 2020, doi: https://doi.org/10.1016/j.diabres.2020.108269.

[12] X. Liu *et al.*, "Gestational diabetes mellitus and risk of neurodevelopmental disorders in young offspring: does the risk differ by race and ethnicity?," *Am J Obstet Gynecol MFM*, vol. 6, no. 1, p. 101217, 2024, doi: https://doi.org/10.1016/j.ajogmf.2023.101217.

[13] Md. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, "Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset," *Computer Methods and Programs in Biomedicine Update*, vol. 4, p. 100118, 2023, doi: https://doi.org/10.1016/j.cmpbup.2023.100118.

[14] S. Fei, H. Chen, Y. Lou, L. Guo, and Q. Pan, "Association between body roundness index and chronic kidney disease risk in patients with diabetes: A cross-sectional analysis of NHANES 1999–2018," *Diabetes Res Clin Pract*, vol. 225, p. 112274, 2025, doi: https://doi.org/10.1016/j.diabres.2025.112274.

[15] B. Mouazen, A. Bendaouia, E. H. Abdelwahed, and G. De Marco, "Machine learning and clinical EEG data for multiple sclerosis: A systematic review," *Artif Intell Med*, vol. 166, p. 103116, 2025, doi: https://doi.org/10.1016/j.artmed.2025.103116.

[16] H. Jahani and A. A. Safaei, "5 - Neural signals processing using deep learning for diagnosis of cognitive disorders," in *Signal Processing Strategies*, A. S. El-Baz and J. S. Suri, Eds., Academic Press, 2025, pp. 91–118. doi: https://doi.org/10.1016/B978-0-323-95437-2.00005-7.

[17] A. E. Hramov, V. A. Maksimenko, and A. N. Pisarchik, "Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states," *Phys Rep*, vol. 918, pp. 1–133, 2021, doi: https://doi.org/10.1016/j.physrep.2021.03.002.

[18] A. Zare Dehnavi *et al.*, "Effects of ADHD and ADHD treatment on glycemic management in type 1 diabetes: A systematic review and meta-analysis of observational studies," *Diabetes Res Clin Pract*, vol. 209, p. 111566, 2024, doi: https://doi.org/10.1016/j.diabres.2024.111566.

[19] P. A. Khosroshahi and M. Ghanbari, "MicroRNA dysregulation in glutamate and dopamine pathways of schizophrenia: From molecular pathways to diagnostic and therapeutic approaches," *Prog Neuropsychopharmacol Biol Psychiatry*, vol. 135, p. 111081, 2024, doi: https://doi.org/10.1016/j.pnpbp.2024.111081.

[20] A. M. D'onofrio *et al.*, "Dopamine dysregulation in patients with a current major depressive episode," *Neuroscience Applied*, vol. 2, p. 103700, 2023, doi: https://doi.org/10.1016/j.nsa.2023.103700.

[21] M. O. Mokhiamar, A. Mahmoud, M. I. Eldagla, Y. Aribi, and A. M. Anter, "Blockchain-integrated multi-modal LSTM-CNN fusion for high-precision epileptic seizure detection from EEG signals," *Knowl Based Syst*, vol. 323, p. 113703, 2025, doi: https://doi.org/10.1016/j.knosys.2025.113703.

[22] P. Riihimaa, "Impact of machine learning and feature selection on type 2 diabetes risk prediction," *J Med Artif Intell*, vol. 3, no. 0, 2020, [Online]. Available: https://jmai.amegroups.org/article/view/5645

[23] N. A. Khan, S. A. Waheeb, A. Riaz, and X. Shang, "A novel Knowledge Distillation-based feature selection for the classification of ADHD," *Biomolecules*, vol. 11, no. 8, p. 1093, 2021, doi: 10.3390/biom11081093.

[24] G. Feretzakis *et al.*, "Integrating Shapley Values into Machine Learning Techniques for Enhanced Predictions of Hospital Admissions," *Applied Sciences*, vol. 14, no. 13, 2024, doi: 10.3390/app14135925.

[25] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clin Transl Sci*, vol. 17, no. 11, p. e70056, 2024, doi: 10.1111/cts.70056.

[26] M. Liu, Y. Ning, H. Yuan, M. E. H. Ong, and N. Liu, "Balanced background and explanation data are needed in explaining deep learning models with SHAP: An empirical study on clinical decision making," *arXiv preprint arXiv:2206.04050*, 2022.

[27] D. Ueda *et al.*, "Fairness of artificial intelligence in healthcare: review and recommendations," *Jpn J Radiol*, vol. 42, no. 1, pp. 3–15, 2024, doi: 10.1007/s11604-023-01474-3.

[28] A. Libin, J. T. Treitler, T. Vasaitis, and Y. Shao, "Evaluating and Reducing Subgroup Disparity in AI Models: An Analysis of Pediatric COVID-19 Test Outcomes," *medRxiv*, 2024, Accessed: Oct. 17, 2025. [Online]. Available: https://www.medrxiv.org/content/10.1101/2022.05.18.22275254v1.full

[29] R. Correa *et al.*, "A systematic review of 'fair'AI model development for image classification and prediction," *J Med Biol Eng*, vol. 42, no. 6, pp. 816–827, 2022.

[30] Y. Yang, H. Zhang, J. W. Gichoya, D. Katabi, and M. Ghassemi, "The limits of fair medical imaging AI in real-world generalization," *Nat Med*, vol. 30, no. 10, pp. 2838–2848, 2024, doi: 10.1038/s41591-024-03113-4.

[31] S. V. Chinta *et al.*, "AI-driven healthcare: A review on ensuring fairness and mitigating bias," *PLOS Digital Health*, vol. 4, no. 5, pp. e0000864-, May 2025, [Online]. Available: https://doi.org/10.1371/journal.pdig.0000864