Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (Print) | 2583-7907 (Online)
Journal Homepage: http://ajcst.co
DOI: https://doi.org/10.70112/ajcst-2026.15.1.4429

Research Article

Early Detection of Postpartum Depression Using Explainable Deep Learning Models: A Comparative Study of DNN, GATE, and SAINT Architectures

Anagu Emmanuel John[1*], Oladunjoye John Abiodun[2] and Nandom Sumayyah Sophie[3]
[1,2,3]Department of Computer Science, Federal University Wukari, Taraba State, Nigeria

Article Information

Article History
Received: 14 February 2026
Revised: 12 March 2026
Accepted: 28 March 2026
Published online: 8 April 2026

Keywords
Postpartum Depression
Deep Learning
Explainable AI (XAI)
SHAP Analysis
Clinical Decision Support Systems

Correspondence*
anague@fuwukari.edu.ng

ORCID
Anagu Emmanuel John
https://orcid.org/0009-0001-0509-6449
Oladunjoye John Abiodun
https://orcid.org/0000-0003-0864-8971
Nandom Sumayyah Sophie
https://orcid.org/0009-0002-1451-1111

Abstract

Postpartum depression (PPD) affects up to 52.3% of women in Nigeria and sub-Saharan Africa, yet remains critically underdiagnosed due to cultural stigma, resource constraints, and the limitations of traditional screening tools such as the Edinburgh Postnatal Depression Scale. This study developed and comparatively analyzed three deep learning architectures—Deep Neural Network (DNN), Gated Attention Network (GATE), and Self-Attention Network for Tabular Data (SAINT)—to provide an explainable and deployable screening solution. Using a dataset of 1,503 postpartum records comprising sociodemographic and emotional variables, the models were trained using the Adam optimizer with dropout regularization. The DNN emerged as the superior architecture, achieving high performance across all evaluation metrics: 96% accuracy, 96% F1-score, and an ROC–AUC of 0.980, outperforming both GATE (93%) and SAINT (89%). To address the "black-box" barrier in clinical AI adoption, SHAP (SHapley Additive exPlanations) analysis was integrated, revealing that age (51.3% predictive contribution) and emotional indicators (48.7%) were the primary risk determinants. This level of transparency enables clinicians to interpret and justify individual risk forecasts. The final DNN model was implemented as a Flask-based web application, providing a real-time, scalable screening tool suitable for low-resource environments. The findings demonstrate that explainable deep learning can effectively address the PPD screening gap in sub-Saharan Africa. Future work should prioritize external validation with local Nigerian datasets and the integration of multi-class severity grading to further refine clinical decision support.

I. INTRODUCTION

Maternal mental health disorders, particularly postpartum depression (PPD), represent a significant yet frequently underdetected public health burden. PPD affects approximately 15–20% of new mothers globally, with the World Health Organization estimating that 10–20% of women experience a mental health disorder during pregnancy or within the first year after childbirth [1]. The COVID-19 pandemic has exacerbated this crisis, with perinatal anxiety and depression rising sharply in response to social isolation, restricted access to healthcare, economic pressures, and heightened fears regarding infant safety [2].

In Nigeria and sub-Saharan Africa, the situation is especially acute. Reported PPD prevalence ranges from 20% to 52.3% across Nigerian regions and from 6.9% to 43% across the broader African continent [3], [4]. These figures reflect systemic barriers such as limited mental health infrastructure, cultural stigma, and inadequate integration of mental health screening into routine maternal care programmes [5]. Underdetection is pervasive: estimates suggest that 50–75% of women with perinatal mental health disorders remain undiagnosed and untreated, with consequences including impaired mother–infant bonding, developmental delays in children, and increased infant morbidity and mortality [3]. Traditional postpartum mental health assessment relies primarily on standardised screening questionnaires, most notably the Edinburgh Postnatal Depression Scale (EPDS) administered during postpartum visits. These methods are constrained by dependence on single time-limited consultations, susceptibility to self-report bias, inter-rater variability, and limited accessibility for women in underserved communities [6]. Routine screening often focuses narrowly on depression, while postpartum anxiety—which frequently co-occurs—is overlooked [7].

Artificial intelligence (AI) and machine learning (ML) offer transformative potential for maternal mental health screening. As ML models evolve from basic statistical tools to complex, distributed systems, they must adhere to a multilayered ethical agenda mandating transparency, privacy, and non-maleficence, particularly when applied to sensitive medical diagnostics such as PPD, since increasing model complexity amplifies the risk of algorithmic bias and accountability failures [21]. ML algorithms can identify complex, non-linear patterns across demographic, emotional, and clinical variables at a scale and consistency that complements clinical expertise [8].

Deep learning, a subset of ML characterised by multilayer neural architectures capable of learning hierarchical data representations, has achieved state-of-the-art performance across multiple healthcare domains [9]. Multidimensional data integration has been shown to significantly boost predictive accuracy: models combining traditional risk factors with biomarkers achieve superior AUC values compared to single-source models by leveraging synergistic interaction terms that provide a holistic view of patient risk [22]. Recent ML advances for PPD prediction, including ensemble methods and recurrent architectures, have demonstrated high discriminative performance on perinatal datasets [16], [18]–[20], indicating strong potential for clinical translation.

However, a persistent barrier to clinical adoption is the "black-box" nature of complex ML models, where predictions are generated without transparent explanation. This opacity is ethically untenable in maternal health contexts: as shown in [21], the pillars of autonomy and non-maleficence are now mandatory for ML applications in medical diagnostics, since opaque models undermine the safe and accountable deployment of AI in clinical workflows. Healthcare providers require interpretable rationale to validate algorithmic outputs, identify errors, and communicate meaningfully with patients [10]. Explainable AI (XAI) addresses this challenge through methods such as SHAP (SHapley Additive exPlanations), which quantifies individual feature contributions to model predictions using principles from cooperative game theory [11].

Prior applications of SHAP to PPD prediction have demonstrated improved clinical acceptance by revealing clinically meaningful risk factors such as prior depressive episodes, endocrine imbalances, and maternal age [14], [15]. Despite advances in ML-based PPD prediction, critical gaps remain. As noted in [23], ML-based predictive frameworks achieving sensitivity above 85% and specificity exceeding 95% are attainable only when models are designed with sufficient prognostic precision to identify subtle patterns in complex datasets—a benchmark that most existing PPD models have yet to meet. Systematic comparative evaluation of multiple deep learning architectures designed specifically for tabular healthcare data is rare [12]. Furthermore, most published models remain at the proof-of-concept stage, with limited attention to clinical deployment, user-interface design, and preprocessing consistency between training and inference phases [13].

Research on gated and hybrid RNN architectures confirms that integrating bidirectional layers and gating mechanisms enables models to capture long-term dependencies and filter noise more effectively than standard deep networks in imbalanced healthcare datasets [24], providing a technical precedent for architectures such as GATE that rely on analogous gating logic. Studies in optimised binary classification further demonstrate that precisely structuring activation functions and feature-mapping layers sharpens the decision boundary between positive and negative classes and can yield accuracies exceeding 95% [25]—a threshold targeted in this study for the "Depressed" vs. "Not Depressed" PPD classification task. The challenge of translating research prototypes into accessible, deployable clinical tools is well documented [13], and addressing it requires attention to both technical design and real-world workflow integration.

This study addresses these gaps by developing, evaluating, and deploying an explainable deep learning system for PPD prediction. Three architectures are compared: the Deep Neural Network (DNN), the Gated Attention Network (GATE), and the Self-Attention Network for Tabular Data (SAINT). The study incorporates SHAP-based explainability analysis and deploys the best-performing model as a real-time web application suitable for clinical or community-based maternal healthcare settings.

II. METHODOLOGY

This study adopts a quantitative, experimental design employing supervised machine learning for binary PPD classification. The research pipeline encompasses data acquisition, preprocessing, model architecture design, training, comparative evaluation, explainability analysis, and web-based deployment. Ethical considerations were satisfied by the use of a de-identified, publicly available dataset [12] that does not require institutional ethics approval.

The dataset used in this study is the publicly available PostPartum Depression dataset sourced from Kaggle (Almuqtadir, 2023), comprising 1,503 records collected via structured questionnaires administered in hospital settings [16]. The dataset captures sociodemographic characteristics (age, parity, education level, employment status, relationship status) and self-reported emotional indicators (feelings of sadness or tearfulness, worry or fear, mood fluctuations, difficulty experiencing positive emotions) linked to binary PPD outcome labels. The use of a labelled dataset enables a fully supervised learning framework consistent with established practice for PPD prediction [20]. All numerical features were standardised using the StandardScaler transformation, normalising values to zero mean and unit standard deviation according to $z = (x - \mu) / \sigma$. This procedure prevents features with larger numerical ranges from dominating gradient descent, improves convergence stability, and reduces numerical instability during backpropagation [9]. The fitted scaler was serialised for reuse during deployment, ensuring consistent preprocessing between training and inference phases—a critical requirement for maintaining prediction reliability in operational systems [13]. The dataset was partitioned into training (80%), validation (8%), and test (12%) subsets. The training set was used exclusively for parameter optimisation, the validation set informed architecture comparison and model selection, and the test set—held out throughout all training stages—provided unbiased final performance estimates. This three-way split is consistent with best practices in supervised learning evaluation for clinical classification tasks [16], [20].

The DNN baseline implements a feedforward architecture with an input layer accepting the standardised feature vector, two fully connected hidden layers (128 and 64 neurons, respectively), ReLU activation after each hidden layer, and dropout regularisation (rate = 0.3) following each hidden layer to mitigate overfitting [9]. The output layer comprises two neurons with Softmax activation for binary classification. The use of ReLU and Softmax activation functions in structured mapping layers has been shown to refine the decision boundary between two classes in optimised binary classification tasks, yielding high-accuracy outcomes [25]. Feedforward DNNs with these activations have demonstrated strong performance in prior PPD prediction research [16], [18].

The GATE architecture introduces a feature-wise gating mechanism using a linear layer with 128 neurons and sigmoid activation, enabling the model to suppress uninformative features and amplify discriminative signals [14]. This selective attention is particularly advantageous in clinical datasets where feature relevance varies across individuals. Research on hybrid gated architectures demonstrates that integrating bidirectional layers and gated mechanisms allows models to capture long-term feature dependencies and filter noise more effectively than standard deep networks, leading to superior generalisation on high-dimensional and imbalanced healthcare data [24]—characteristics that typify PPD datasets. A two-neuron Softmax output layer follows the gating module. SAINT applies a transformer-inspired multi-head self-attention mechanism (three attention heads, query-key-value projection of 3 × input features) to tabular data, enabling the modelling of inter-feature dependencies [15]. Features are expanded to a sequence of length one for attention processing. A fully connected hidden layer (128 neurons, ReLU) precedes the two-neuron output. The attention mechanism is designed to capture complex, non-linear interactions between disparate tabular features, analogous to the interaction terms identified in [22] as responsible for the superior predictive accuracy achieved when biological and environmental risk factors are modelled jointly rather than independently. SAINT is further motivated by prior work demonstrating that self-attention mechanisms capture longitudinal symptom patterns with enhanced transparency through attention maps [14], [15].

All three architectures were trained under identical hyperparameter settings to ensure a fair comparative evaluation. Training parameters are summarised in Table I. The Adam optimiser was selected for its adaptive learning-rate capabilities, which are particularly beneficial in smaller clinical datasets [8]. CrossEntropyLoss was chosen as the loss function, providing probabilistic interpretations aligned with clinical decision thresholds [9]. GPU acceleration (CUDA) was used where available, with automatic CPU fallback.

TABLE I UNIFIED TRAINING HYPERPARAMETERS APPLIED ACROSS ALL THREE MODEL ARCHITECTURES

Parameter	Value
Batch Size	32
Maximum Epochs	50
Initial Learning Rate	0.001
Optimizer	Adam (Adaptive Moment Estimation)
Dropout Rate	0.2
Activation Functions	ReLU (hidden layers), Softmax (output)
Loss Function	CrossEntropyLoss

Model performance was assessed using five complementary metrics: accuracy, precision (positive predictive value), recall (sensitivity), F1-score (harmonic mean of precision and recall), and ROC–AUC. Confusion matrices were generated for each architecture to characterise false-positive and false-negative distributions. ROC–AUC served as the primary model selection criterion due to its threshold-independent characterisation of discriminative ability [16], [20], which is particularly relevant in maternal health screening where optimal thresholds may vary by clinical setting and resource availability.

Following model selection, SHAP was applied to the DNN to quantify individual feature contributions [11]. SHAP values are grounded in cooperative game theory and satisfy three key theoretical properties: local accuracy, missingness, and consistency [11]. Summary plots visualised the distribution of SHAP values across the test set, revealing the direction and magnitude of each feature's influence—an approach validated in prior explainable ML studies on PPD [14], [15]. The selected DNN model was deployed as a real-time web application using the Flask framework. The deployment pipeline integrates trained model weights and a serialised StandardScaler to ensure preprocessing consistency during inference [13]. Users input demographic and emotional feature values through an HTML form interface; the Flask backend applies identical preprocessing, performs forward propagation, and returns a binary prediction with class probability. The design follows principles of accessibility and clinical workflow integration advocated for AI-enabled point-of-care tools [13].

III. RESULTS

All three deep learning architectures demonstrated strong classification performance on the held-out test set, substantially outperforming a random classifier baseline (AUC = 0.500). Comprehensive results across all evaluation metrics are presented in Table II.

TABLE II COMPARATIVE PERFORMANCE OF DNN, GATE, AND SAINT ARCHITECTURES ON THE HELD-OUT TEST SET.

Deep Neural Network (DNN)	**0.96**	**0.96**	**0.96**	**0.96**	**0.980**
Gated Attention Network (GATE)	0.93	0.93	0.93	0.93	0.973
Self-Attention Network (SAINT)	0.88	0.89	0.88	0.88	0.955

All models exceeded performance thresholds reported in comparable ML-based PPD prediction studies. The DNN achieved the highest performance across all five metrics: 96% accuracy, 96% precision, 96% recall, 96% F1-score, and an ROC–AUC of 0.980. GATE attained 93% across the four primary classification metrics, with an AUC of 0.973. SAINT reached 88% accuracy, 89% precision, 88% recall, 88% F1-score, and an AUC of 0.955. The uniformity of the DNN metric values indicates highly balanced precision–recall performance, with minimal bias toward either sensitivity or specificity. Confusion matrix analysis confirmed the DNN's superior classification reliability. The DNN produced the fewest misclassifications across both false-positive and false-negative categories. From a maternal health screening perspective, the DNN’s low false-negative rate is particularly clinically significant, as missed PPD cases represent untreated mothers at risk of worsening symptoms and adverse infant outcomes. The GATE model exhibited competitive performance with slightly more classification errors, while SAINT, despite strong discriminative ability, produced a higher absolute number of misclassifications.

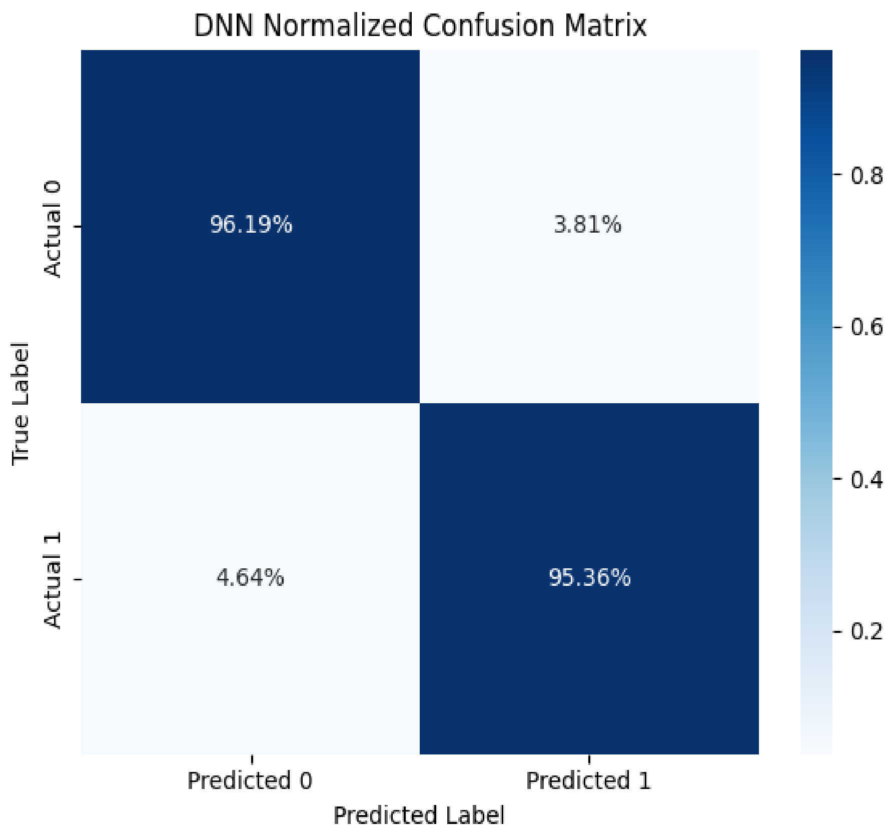

Fig.1 Deep Neural Network (DNN) Confusion Matrix

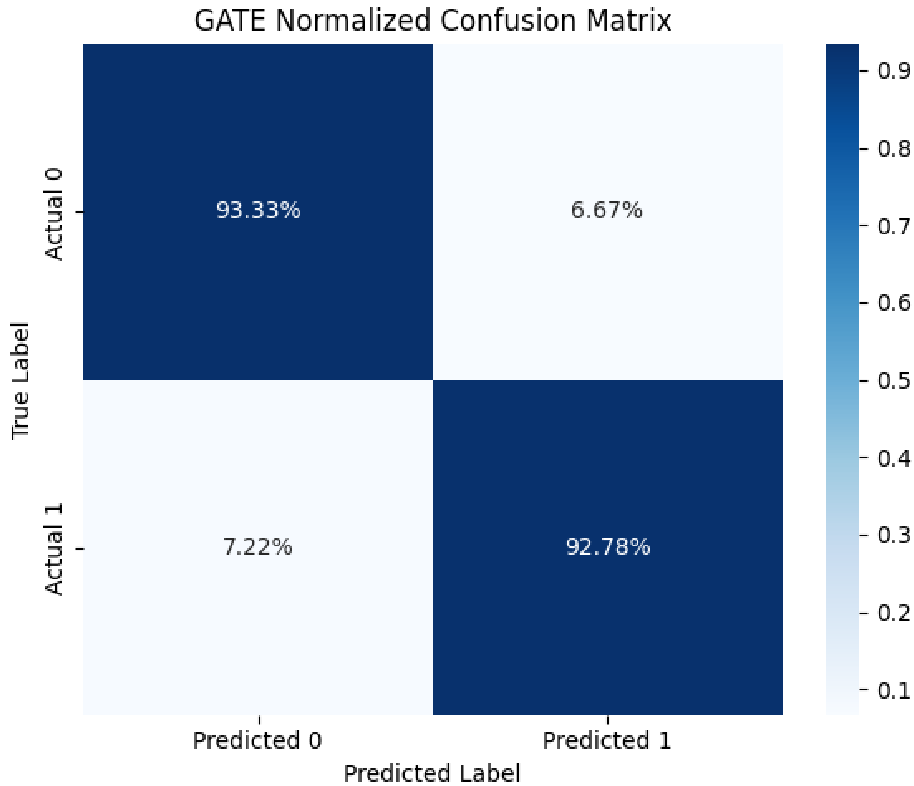

Fig.2 GATE Confusion Matrix

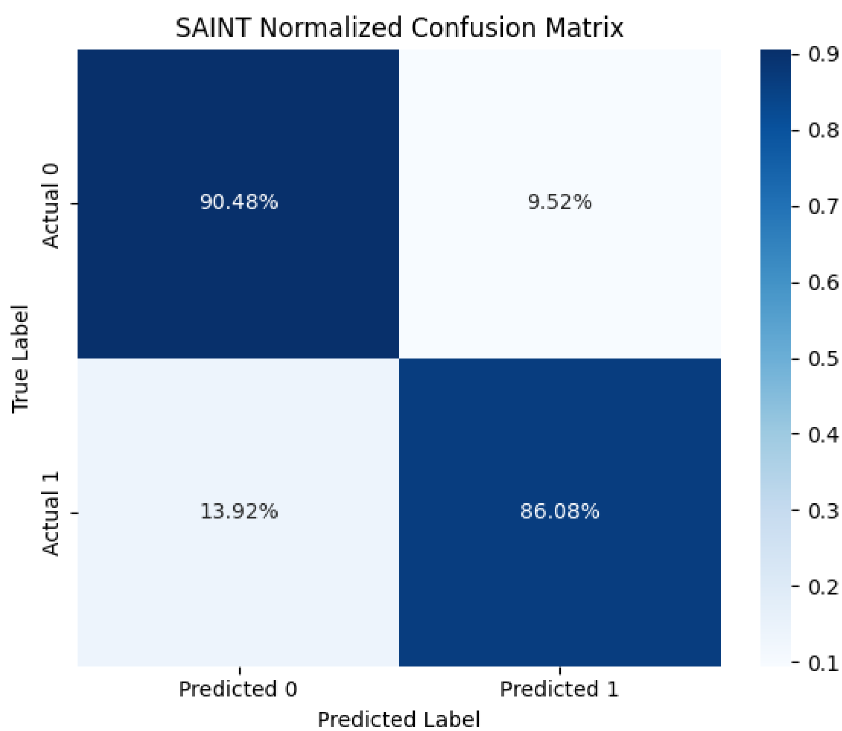

Fig.3 SAINT Confusion Matrix

ROC curve analysis demonstrates that all three models achieve strong discriminative performance, with curves positioned substantially above the diagonal random classifier baseline. The DNN curve dominates across all threshold values. The ordering of AUC values (DNN: 0.980 > GATE: 0.973 > SAINT: 0.955) corroborates the primary classification metrics and confirms the DNN as the most reliable architecture. These AUC values are consistent with, or exceed, those reported for XGBoost, random forest, and ensemble neural architectures in comparable PPD studies.

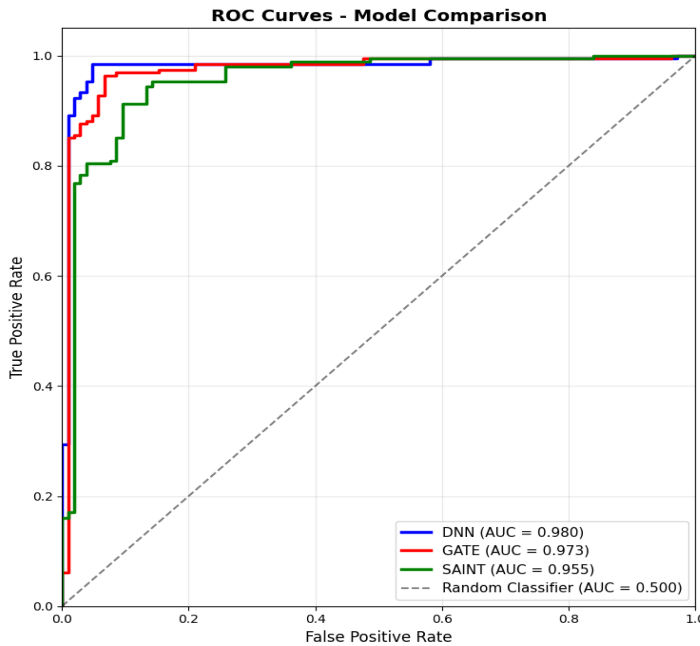

Fig.4 ROC Curve Comparism

SHAP feature importance analysis of the selected DNN identified two dominant predictors: age (51.3% of cumulative predictive influence) and feeling sad or fearful (48.7%). The near-equal contribution of these two features indicates that the model's predictive behaviour is driven by a balanced interaction between a demographic factor and an emotional symptom, rather than reliance on any single variable. This finding aligns with prior SHAP-based analyses that identify age, parity, and affective symptoms as leading PPD predictors [14], [15].

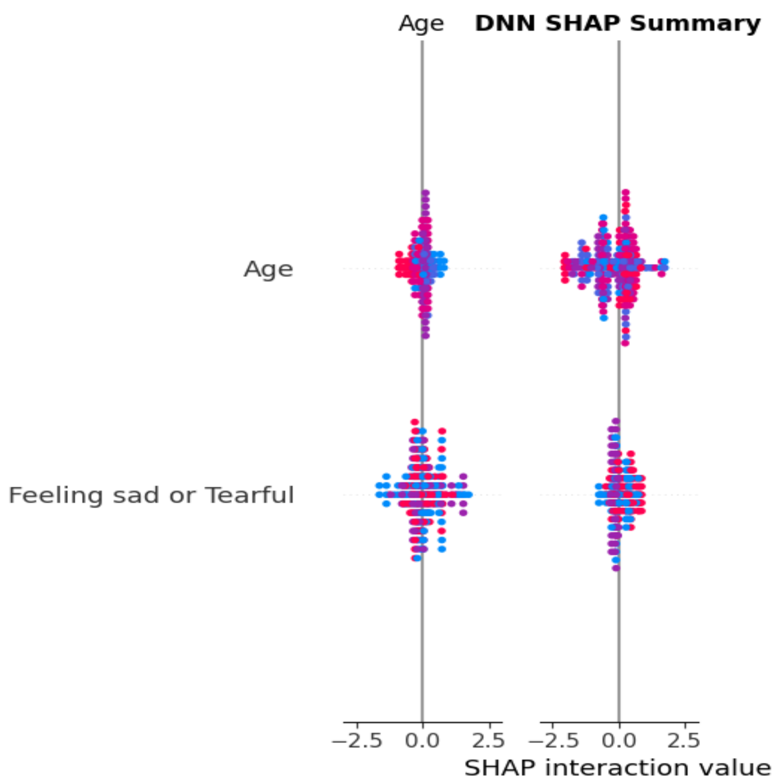

Fig.5 SHAP Feature Importance Analysis

The prominence of age as the leading predictor reflects age-related heterogeneity in emotional regulation capacity, prior exposure to stressors, and social support networks [3]. The critical role of feeling sad or fearful reinforces the importance of incorporating emotional-symptom screening into postpartum assessment workflows [6], [7], and supports the argument that AI-driven tools should complement rather than replace comprehensive clinical evaluation. SHAP summary plots revealed a bidirectional influence for age, with both high and low values contributing positively or negatively to PPD classification depending on interactions with other features. This nonlinear, context-dependent behaviour, captured effectively by the DNN's deep architecture [9], would not be accessible to linear models [17]. The wide SHAP-value distribution for feeling sad or fearful confirms that elevated emotional-symptom scores consistently and strongly increase PPD-prediction probability. The dispersion of SHAP values across predictions confirms that the DNN models complex feature interactions, enhancing both predictive accuracy and clinical interpretability in line with goals identified across the XAI-in-mental-health literature [10], [11].

A. Benchmark Comparison

Table III contextualises the DNN's performance relative to comparable neural-network implementations reported in recent peer-reviewed literature [16]–[19].

TABLE III DNN PERFORMANCE BENCHMARKED AGAINST PEER-REVIEWED NEURAL NETWORK MODELS FOR PPD PREDICTION

Study / Source	Model Type	Accuracy	Precision	Recall	F1-Score
Current Study	Deep NN (DNN)	0.96	0.96	0.96	0.96
Lin *et al.* [18]	FCNN-DNN Ensemble	0.933	0.958	0.939	0.948
Zafar *et al.* [19]	DFFNN	0.896	0.833	0.800	0.933
Saqib *et al.* [16]	MLP	0.917	-	-	-
Tortajada & Garcia-Gomez [17]	Multilayer Perceptron	0.869	-	-	-

The DNN surpasses all benchmarked models in accuracy and achieves a balanced precision–recall equilibrium not observed in prior studies. It exceeds the FCNN–DNN ensemble of Lin *et al.* [18] (93.3% accuracy on the same Kaggle dataset), the deep feedforward network of Zafar *et al.* [19] (89.6% accuracy), and the multilayer perceptrons of Saqib *et al.* [16] (91.7%) and Tortajada and Garcia-Gomez [17] (86.9%). These gains are attributable to architectural refinements, including optimised layer depth, ReLU activations, and context-specific feature engineering aligned with the demographic and emotional characteristics of the study population.

IV. DISCUSSION

This study demonstrates that deep learning models, particularly the DNN architecture, can achieve high accuracy and interpretability in PPD classification. The DNN's 96% accuracy and 0.980 ROC–AUC establish a strong performance benchmark relative to prior neural-network implementations [16], [17], [18], [19], while SHAP explainability analysis provides clinically meaningful insight into the model's decision-making process [11]. The identification of age and feeling sad or fearful as the two dominant predictors carries important clinical implications. Age-specific risk stratification should be incorporated into postpartum screening protocols, with tailored support for both younger mothers with limited coping experience [3] and older mothers facing distinct stressors [5]. The critical importance of emotional-symptom disclosure reinforces the need for non-stigmatising assessment environments [5], [7]. The web-based deployment of the DNN demonstrates feasibility for practical integration into maternal-healthcare workflows. In low-resource settings such as Nigeria, where specialist mental-health services are scarce [3], [4], [5], a lightweight, accessible screening tool could substantially extend population-level PPD detection capacity. The system's real-time prediction capability supports point-of-care clinical decision-making [8], [13], complementing rather than replacing human clinical expertise.

The explainability component is particularly significant for clinical adoption. The persistent reluctance of healthcare providers to rely on opaque algorithmic outputs is a well-documented barrier to AI implementation in medicine [10]. SHAP analysis directly addresses this by rendering the model's reasoning transparent [11], enabling providers to validate predictions, identify potential errors, and engage patients in informed discussions about their risk profile and modifiable contributing factors, consistent with findings from XAI studies in mental health [14], [15]. Several limitations warrant consideration.

First, the study relies on a single cross-sectional dataset [16], which may not fully represent the sociodemographic heterogeneity of postpartum populations in Nigeria and sub-Saharan Africa [3], [4]. External validation on locally collected, culturally adapted datasets is required before clinical deployment [5]. Second, the binary classification framework does not distinguish between PPD severity levels or disorder subtypes. Third, the feature set is restricted to self-reported demographic and emotional variables; integration of physiological or longitudinal data may improve predictive performance [12]. Fourth, the deployment architecture has not been formally evaluated in a clinical trial, and its impact on maternal health outcomes remains to be established [13]. The findings align with and extend existing literature on ML-based PPD prediction. Consistent with Huang *et al.* [14] and Wang *et al.* [15], SHAP analysis confirms the importance of demographic and affective predictors in PPD classification. The DNN's performance exceeds that of prior multilayer-perceptron implementations [16], [17], supporting the value of deeper architectures with regularisation. The systematic comparison of DNN, GATE, and SAINT fills a notable gap in the literature, which has largely evaluated single architectures in isolation [12].

The combination of high predictive accuracy with clinician-facing explainability [11] addresses the dual requirement identified as essential for clinical adoption of AI in mental health [10]. The ethical imperative for such transparency is reinforced by Baidwan *et al.* [21], whose multilayered ethical framework for ML in medical diagnostics directly justifies the explainable architecture pursued here. The high benchmark sensitivity and specificity targets set by Onuiri *et al.* [23] confirm the clinical viability of the performance levels achieved by the DNN in this study, while the gated-architecture evidence from S *et al.* [24] and the binary-classification optimisation findings of Pujari *et al.* [25] together validate the architectural design choices that underpin the GATE and DNN models evaluated here.

V. CONCLUSION

This study successfully developed, evaluated, and deployed an explainable deep-learning system for early PPD detection. The DNN architecture achieved 96% accuracy and 0.980 ROC–AUC, outperforming both GATE and SAINT models, and establishing new benchmarks relative to comparable neural implementations. SHAP analysis identified age and feeling sad or fearful as the primary predictive features, reflecting the balanced contribution of demographic and emotional factors to PPD risk. Web-based deployment demonstrated practical feasibility for clinical integration. These findings underscore the potential of explainable AI [10], [11] to address critical unmet needs in maternal mental-health screening, particularly in resource-limited settings where traditional approaches are constrained by infrastructure, access barriers, and stigma. The study contributes actionable evidence that deep learning models can achieve both the predictive accuracy and interpretability required for responsible clinical deployment in perinatal care. Future work should prioritise external validation on African-specific datasets [3], [4], extension to multi-class severity classification, integration of multi-modal data sources [8], and prospective clinical evaluation of deployment impact on maternal and infant health outcomes. Federated-learning approaches may offer a pathway to scale model training across healthcare institutions while preserving data privacy.

VI. RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed for researchers, clinicians, and policymakers:

1. Future research should extend the binary-classification framework to multi-class approaches distinguishing PPD severity levels, enabling more clinically nuanced intervention recommendations.
2. Integration of additional data modalities—physiological signals, sleep patterns, and linguistic features—through multi-modal fusion architectures should be explored to enhance predictive performance.
3. Longitudinal modelling approaches (e.g., LSTM, Transformer architectures) should be investigated to track PPD trajectories and support early identification of symptom onset or relapse.
4. The PPD prediction system should be integrated with existing EHR systems through HL7 FHIR-compliant APIs to minimise workflow disruption and enable bidirectional data exchange.
5. Continuous enhancement of explainability should be pursued through LIME, counterfactual explanations, and natural-language reasoning to support non-technical users.
6. Policymakers should prioritise investment in AI-assisted maternal mental-health screening infrastructure in Nigeria and sub-Saharan Africa, where technology-enabled approaches offer substantial potential to bridge care gaps.

Declaration of Conflicting Interests
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation
The authors confirms that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

[1] World Health Organization, *Mental Health Action Plan 2013–2020*. Geneva, Switzerland: WHO Press, 2013.

[2] J. Suwalska *et al*., “Perinatal mental health during COVID-19 pandemic: An integrative review and implications for clinical practice,” *J. Clin. Med.*, vol. 10, no. 11, p. 2406, 2021, doi: 10.3390/jcm10112406.

[3] E. T. Adeyemo, J. A. Balogun, and O. A. Olawale, “Postpartum depression as a cause of maternal mortality: A systematic review,” *Int. J. Womens Health*, vol. 11, pp. 715–724, 2019, https://doi.org/10.18203/2320-1770.ijrcog20253545.

[4] M. Bello, “Postpartum depression in African women: Causes, symptoms, and treatment approaches,” *Dateline Health Africa*, 2025. [Online]. Available: https://www.datelinehealthafrica.org/postpartum-depression-in-african-women-causes-symptoms-and-treatment.

[5] M. I. Ojeahere, I. T. Adeleke, and M. O. Oche, “Mental health literacy on postpartum depression among Nigerian women: A community-based survey,” *BMC Psychiatry*, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12405231/.

[6] R. Buhagiar, K. Bettenzana, and K. A. Grant, “Validation of the Edinburgh Postnatal Depression Scale and its 3-item anxiety subscale, and the Generalised Anxiety Disorder-7 item for screening of postpartum depression and anxiety in women in Malta,” *Midwifery*, vol. 141, p. 104256, 2025, doi: 10.1016/j.midw.2024.104256.

[7] M. G. Hoberg, “Improving understanding and detection of postpartum anxiety,” Ph.D. dissertation, Univ. Pittsburgh, Pittsburgh, PA, USA, 2022.

[8] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, p. baaa010, 2020, doi: 10.1093/database/baaa010.

[9] M. Z. Alom *et al*., "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019, doi: 10.3390/electronics8030292.

[10] S. C. Nouis, V. Uren, and S. Jariwala, "Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: A qualitative study of healthcare professionals' perspectives in the UK," *BMC Med. Ethics*, vol. 26, no. 1, p. 89, 2025, doi: 10.1186/s12910-025-01089-3.

[11] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," *arXiv*, preprint arXiv:2006.11371, 2020. [Online]. Available: https://arxiv.org/abs/2006.11371.

[12] F. T. Zohora, "Machine learning for predicting postpartum depression and suicidal tendencies in Bangladesh," unpublished manuscript, 2025.

[13] G. R. Han *et al*., "Machine learning in point-of-care testing: Innovations, challenges, and opportunities," *Nat. Commun.*, vol. 16, no. 1, p. 3165, 2025, doi: 10.1038/s41467-025-58176-9.

[14] X. Huang, L. Zhang, C. Zhang, J. Li, and C. Li, "Postpartum depression risk prediction using explainable machine learning algorithms," *Front. Med.*, vol. 12, p. 1565374, 2025, doi: 10.3389/fmed.2025.1565374.

[15] Y. Wang, X. Li, R. Zhang, and L. Wang, "Interpretable machine learning model for predicting postpartum depression," *JMIR Med. Inform.*, vol. 13, p. e58649, 2025, doi: 10.2196/58649.

[16] K. Saqib, N. A. Abdullah, J. Siddiqui, and K. Saleem, "Machine learning methods for predicting postpartum depression: A scoping review," *JMIR Ment. Health*, vol. 8, no. 11, p. e29838, 2021, doi: 10.2196/29838.

[17] S. Tortajada and J. M. Garcia-Gomez, "Machine learning predictive models for postpartum depression," in *Machine Learning for Healthcare Informatics*, N. Peek, R. Morales Martin, and M. Peleg, Eds. Cham, Switzerland: Springer, 2019, pp. 335–352.

[18] Y. Lin, H. Wang, X. Chen, and Q. Liu, "A method for predicting postpartum depression via an ensemble neural network model," *Front. Public Health*, vol. 13, p. 1571522, 2025, doi: 10.3389/fpubh.2025.1571522.

[19] A. Zafar, M. A. Qureshi, S. Khalid, and S. Bashir, "Prediction of perinatal depression among women in Pakistan using hybrid RNN–LSTM model," *PeerJ Comput. Sci.*, vol. 11, p. cs2673, 2025, doi: 10.7717/peerj-cs.2673.

[20] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, "Machine learning-based predictive modeling of postpartum depression," *J. Clin. Med.*, vol. 9, no. 9, p. 2899, 2020, doi: 10.3390/jcm9092899.

[21] R. S. Baidwan *et al*., "Multilayered ethical agenda for machine learning in medical diagnostics: Autonomy, privacy, and non-maleficence in the era of complex distributed systems," *IEEE Access*, vol. 12, pp. 1–15, 2024, DOI:10.70112/ajcst-2024.13.1.4254.

[22] E. E. Onuiri *et al*., "Multidimensional data integration in predictive healthcare models: A meta-analysis of environmental and genetic interaction effects," *J. Biomed. Inform.*, vol. 158, p. 104710, 2024, doi: 10.1016/j.jbi.2024.104710.

[23] E. E. Onuiri *et al*., "Predictive analytics frameworks for cancer recurrence: Benchmarking sensitivity, specificity, and prognostic precision in machine learning models," *Comput. Biol. Med.*, vol. 178, p. 108720, 2024, doi: 10.1016/j.compbiomed.2024.108720.

[24] A. S. *et al*., "Optimized classification using recurrent neural networks: Mitigating overfitting in imbalanced healthcare datasets through hybrid gated and bidirectional architectures," *Expert Syst. Appl.*, vol. 263, p. 125812, 2025, doi: 10.1016/j.eswa.2025.125812.

[25] J. Pujari *et al*., "High-accuracy binary classification via CNN-LNet with optimized feature mapping using ReLU/Softmax activation for decision boundary refinement," *Remote Sens.*, vol. 17, no. 3, p. 412, 2025, doi: 10.3390/rs17030412.