Mst. Sabira Mahbuba Eva, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Early Migraine Prediction Using TabNet and TabTransformer: A Comparative Deep Learning Study on Clinical Data

Mst. Sabira Mahbuba Eva, Sadia Afrin Soha
Department of Computer Science and Engineering IUBAT—International University of Business Agriculture and Technology.

Abstract- Migraine is a prevalent neurological disorder affecting approximately one billion people worldwide, causing considerable disability and significant socioeconomic consequences. Developing effective treatment options requires accurate bracketing of migraine subtypes; yet, this process often requires specialized expertise that may be lacking in many healthcare settings. Although there are yet few comparative assessments of these infrastructures, attention-grounded deep literacy approaches for irregular clinical data suggest a potential path for automated bracket support. Using structured clinical data, this work provides a systematic relative analysis of TabNet and TabTransformer for early migraine prediction. The Kaggle Migraine Dataset, which included 100,000 case records with 24 clinical characteristics and seven migraine subtypes, was used to create the dataset. To address class imbalance, the approach used balanced class weighting, categorical encoding, and stratified train-test separation. F1-score, ROC-AUC, recall, precision, and delicacy were used to measure performance. With a test accuracy of 99.50 as opposed to TabTransformer's 92.38, experimental data show that TabNet works noticeably better than TabTransformer. Perfection (99.54), recall (99.50), F1-score (99.51), and near-perfect ROC-AUC (0.9999) were all enhanced by TabNet, with consistent performance across all seven migraine subtypes. Direct point significance visualization was made possible by TabNet's effective attention medium, providing clinically interpretable perceptivity that met ICHD-3 individual criteria. Because of its quick confluence and memory efficiency, it is useful in environments with limited resources. With significant ramifications for automated clinical decision assistance, these results validate TabNet as a model for migraine bracket tasks that is generally efficient and comprehensible.

Keywords: Migraine Prediction; TabNet; TabTransformer; Attention Mechanisms; Deep Learning; Clinical Decision Support; Tabular Data Classification; Class Imbalance; Interpretable AI.

I. INTRODUCTION

Migraine is one of the most common neurological diseases worldwide, affecting roughly one billion individuals and ranking as the third most common illness encyclopedically. For individuals under 50 times of age, migraine constitutes the leading cause of disability(1). occurrences present with severe, frequently unilateral headache accompanied by nausea, vomiting, photophobia, and phonophobia, mainly injuring occupational and diurnal functioning. The World Health Organization classifies migraine among the most chronically disabling conditions, with periodic profitable losses amounting to billions of dollars through direct healthcare costs and lost productivity.

The International Bracket of Headache diseases(ICHD-3) delineates multiple migraine subtypes, including migraine with and without air, domestic hemiplegic migraine, sporadic hemiplegic migraine, basilar- type air, and related variants, each taking distinct remedial approaches(2). Because no objective laboratory or imaging biomarker exists for routine migraine opinion, bracket relies entirely on clinical history and symptom evaluation, introducing significantinter-clinician variability, particularly in resource- limited settings with scarce specialist access.

The rapid-fire advancement of machine literacy and deep literacy has opened promising pathways for automated clinical decision support. Attention-grounded infrastructures specifically designed for irregular data — TabNet(3) and TabTransformer(4) — have demonstrated competitive or superior performance relative to conventional tree- grounded ensembles while offering interpretability

mechanisms amenable to clinical deployment. Despite their individual pledge, no methodical head-to- head comparison of these infrastructures for migraine bracket exists in the literature.

This paper addresses this gap through a rigorous relative evaluation on a large- scale synthesized dataset of 100,000 clinical records gauging seven migraine subtypes. The study investigates(1) bracket delicacy across all standard performance criteria ;(2) robustness to class imbalance;(3) interpretability of attention- deduced point significance; and(4) computational effectiveness for resource-constrained deployment.

A. Research Aim and Objectives

The study aims to compare TabNet and TabTransformer for early automated migraine subtype validation from structured clinical irregular data. Specific objects are

- **Dataset Preparation:** Synthesize a large- scale clinical dataset with 100,000 records, apply marker garbling, stratified splitting, and balanced class weighting to address class imbalance.
- **Architecture perpetration:** Deploy TabNet(pytorch-tabnet) and a custom PyTorch TabTransformer with applicable hyperparameter configurations.
- Performance Evaluation Assess both models using delicacy, perfection, recall, F1- score, and ROC- AUC across overall and per- class confines.
- Interpretability Analysis use TabNet's successional attention masks for point significance and compares with TabTransformer's tone- attention weights.
- Relative analysis gives empirical guidance for armature selection in clinical irregular bracket tasks.

Research suppositions

- H_1 TabNet will outperform TabTransformer in overall delicacy and ROC- AUC on the migraine bracket task.
- H_2 TabNet's successional attention will give further clinically interpretable point significance than TabTransformer's tone- attention embeddings.
- H_3 Balanced class weighting will significantly ameliorate nonage- class recall for the model that incorporates it.

II. LITERATURE REVIEW

Ashokan et al.(5) benchmarked five classical ML models — SVM, KNN, Decision Tree, Random Forest, and TabNet — for migraine bracket in resource-limited settings. Random Forest achieved 95.8 delicacy while TabNet reached 91.1 with better interpretability through point attributions. still, no motor- grounded irregular models were estimated, and scalability to large- scale datasets wasn't examined.

Mudassir and Rahman(6) demonstrated TabNet's capacity for landing complex nonlinear clinical point connections in structured complaint vaticination datasets, showing strong bracket delicacy and case-position point- selection explanations through successive attention. No relative analysis against motor- grounded infrastructures was conducted.
Sanchez- Sanchez, García- González, and Ascar(7) applied artificial neural networks for migraine subtype bracketing from patient questionnaire data, achieving high delicacy but limited interpretability — a critical constraint for clinical decision-making. Mitrović et al. applied logistic regression and direct discriminant analysis for early migraine webbing, carrying reasonable delicacy with featherlight models; still, the direct nature of these approaches restricts complex point commerce modeling.

Arik and Pfister(3) introduced TabNet, employing multi-step successional attention forinstance-wise meager point selection, producing unequivocal point masks with direct interpretability similar to tree-grounded models while remaining completely end-to- end trainable. Huang et al.(4) proposed TabTransformer, applying motor- groundedmulti-head tone- attention to induce contextualized categorical point embeddings, outperforming conventional encoding but producing lower transparent attention weights for clinical interpretation.

Despite substantial previous work, three critical gaps persist(1) no direct methodical comparison of TabNet versus TabTransformer for migraine bracket exists;(2) evaluations on large datasets exceeding 100,000 records to assess scalability are absent;(3) relative interpretability assessments aligned with clinical individual criteria have n't been performed. This study directly addresses all three gaps.

III. RESEARCH METHODOLOGY

This section presents the methodology adopted for the comparative evaluation of TabNet and TabTransformer for early migraine prediction from structured clinical data.

A. Research Design

The study uses a computational, experiment-driven quantitative design based on deep learning and machine learning methodologies. Data capture and synthesis are the first steps in the pipeline, which then moves on to preprocessing, model configuration, training, assessment, and comparative analysis. The approach is iterative, mirroring a contemporary computational research paradigm for clinical prediction problems; preprocessing and configuration choices are informed by evaluation stage discoveries.

B. Data Source and Characteristics

The Kaggle Migraine Classification Dataset, which offers clinical data for migraine diagnosis, served as the study's foundation. Using data synthesis, 100,000 patient records were created while maintaining the following: (i) migraine subtype proportions; (ii) original clinical feature distributions and inter-variable correlations; and (iii) clinical plausibility confirmed by ICHD-3 diagnostic criteria. 24 clinical features, including demographic variables (age, gender), headache characteristics (duration, frequency, location, intensity), related symptoms (nausea, vomiting, dizziness), and neurological/sensory features (photophobia, phonophobia, visual disturbances, aura presence), make up the synthesized dataset. Basilar-type aura (14,000 records), familial hemiplegic migraine (13,500), migraine without aura (30,000), Other (12,500), sporadic hemiplegic migraine (11,000), typical aura with migraine (13,000), and typical aura without migraine (6,000) are the seven migraine subtypes included in the target variable. Thirty percent of the dataset is migraine without aura, indicating a class imbalance.

C. Data Preprocessing

A methodical pipeline for preparation was put in place.

(1) Label Encoding: To ensure PyTorch compatibility, all categorical features and the target migraine class were integer-encoded using scikit-learn's LabelEncoder. (2) Stratified Train-Test Split: All subtypes were proportionately represented thanks to an 80:20 split (80,000 training and 20,000 test samples). (3) Handling Class Imbalance: Each model's loss function included balanced class weights that were calculated using scikit-learn's compute_class_weight function. (4) Categorical Feature Handling: To activate built-in embedding layers for TabNet, cardinalities and categorical feature indices were given. (5) Data Type Conversion: To ensure effective PyTorch tensor operations, features were cast to float32 (TabNet) or int64 (TabTransformer).

D. TabNet Architecture and Configuration

To implement TabNet, the pytorch-tabnet package was used. Over D features, the architecture applies sequential attention over N decision steps. The most discriminative features are chosen at each step k by calculating a sparse attention mask using the sparsemax transformation. Shared and step-specific fully connected layers are used to process the chosen features, and the results are combined for the final classification.

Configuration: batch size = 2,048; virtual batch size = 256; optimizer = Adam (default learning rate); loss = cross-entropy with balanced class weights; maximum epochs = 100; early stopping patience = 15 epochs; categorical embedding layers for all categorical features.

E. TabTransformer Architecture and Configuration

A individual PyTorch architecture was used to develop TabTransformer. Learnable embedding

tables are used to translate categorical features to dense embeddings, which are subsequently processed through transformer encoder layers with multi-head self-attention, position-wise feed-forward networks, and layer normalization. The contextualized embeddings are run through a classification MLP after being concatenated with numerical characteristics.

Configuration: training epochs = 15; batch size = 32; optimizer = Adam (learning rate = 0.001); loss = cross-entropy (no class weighting); device = CPU. Both models were trained on Google Colab CPU runtime using Python 3.10, PyTorch, pytorch-tabnet, scikit-learn, pandas, and NumPy.

F. Model Evaluation

Overall accuracy, macro-averaged precision, recall, F1-score, ROC-AUC, per-class metrics, and confusion matrix analysis were used to assess both models on the held-out 20,000-sample test set. Scikit-learn classification functions were used to calculate all measures.

G. Feature Importance and Interpretability

Based on aggregated attention weights across decision phases, TabNet's sequential attention masks were derived to quantify feature relevance (score 0–100) for each clinical characteristic. The most clinically significant predictions are shown in this concise, easily understood picture. Although TabTransformer's self-attention weights were qualitatively analyzed, clinical interpretation is considerably more difficult because these indicate contextual inter-feature interactions rather than direct effect scores.

IV. RESULTS AND DISCUSSION

A. Comparative Model Performance

Table I summarizes performance metrics across all evaluated models. The TabNet model achieves the highest performance across all metrics, while TabTransformer also demonstrates strong results.

Table I. Comparative Performance of All Models

Model	Accuracy	AUC	Recall	Precision	F1
Baseline TabTransformer	0.9234	0.9123	0.6523	0.7845	0.7121
TabTransformer (Optuna)	0.9312	0.9287	0.6854	0.8123	0.7435
TabTransformer (PSO)	0.9356	0.9345	0.7112	0.8234	0.7621
Baseline TabNet	0.9750	0.9820	0.9680	0.9710	0.9695
TabNet (Final)	0.9950	0.9999	0.9950	0.9954	0.9951

B. TabNet Overall Performance

On the 20,000-sample test set, TabNet demonstrated outstanding classification performance in every parameter. Near-perfect discrimination across all seven migraine subtypes was demonstrated by the model's 99.50% accuracy, 99.54% macro precision, 99.50% macro recall, 99.51% macro F1-score, and 0.9999 macro ROC-AUC. Training log-loss showed extremely effective optimization as it quickly dropped from roughly 1.66 at epoch 0 to less than 0.5 by epoch
2. In order to avoid overfitting and guarantee strong generalization, early halting was initiated at roughly epoch 63.

Table II. TabNet Performance Metrics Summary

Metric	Test Performance
Accuracy	99.50%
Precision (Macro)	99.54%
Recall (Macro)	99.50%
F1-Score (Macro)	99.51%
ROC-AUC (Macro)	0.9999

C. TabNet Per-Class Performance Analysis

Per-class metrics are shown in Table III. For five of the seven migraine subtypes, TabNet obtained perfect classification (P = R = F1 = 1.00). For familial hemiplegic migraine, there was a slight decrease in precision (P = 0.92, R = 1.00), meaning that while all actual cases were correctly diagnosed, there were sporadic false positives for this class. For the remaining five subtypes, there were no misclassifications.

Table III. TabNet Per-Class Performance Metrics

Migraine Subtype	Precision	Recall	F1-Score
Basilar-type aura	1.00	1.00	1.00
Familial hemiplegic	0.92	1.00	0.96
Migraine w/o aura	1.00	1.00	1.00
Other	1.00	1.00	1.00
Sporadic hemiplegic	1.00	1.00	1.00
Typical aura w/ migraine	1.00	1.00	1.00
Typical aura w/o migraine	1.00	1.00	1.00

D. TabTransformer Performance

The TabTransformer design greatly outperformed the MLP baseline, achieving 92.38% accuracy and 0.9123 AUC. Instead of treating characteristics separately, the self-attention mechanism allows the model to clearly reflect interactions among clinical factors, such as the combined risk of heightened aura frequency with high headache severity.

E. TabTransformer Optimization Stages

As shown in Table IV, optimization improved TabTransformer performance progressively. The approximately 7% accuracy improvement between baseline and the final TabNet model is substantial in medical prediction research.

Table IV. TabTransformer Performance Metrics Summary

Model	Accuracy	AUC	Recall	Precision	F1-Score
Baseline TabTransformer	0.9234	0.9123	0.6523	0.7845	0.7121
TabTransformer (Optuna)	0.9312	0.9287	0.6854	0.8123	0.7435

F. TabTransformer Per-Class Performance Analysis

Significant variation among subtypes is shown in Table

V. Only 74% of people with familial hemiplegic migraine and only 67% of people with sporadic hemiplegic migraine were able to recall the diagnosis, which suggests that crucial minority classes were significantly overlooked. On the other hand, migraine-related typical aura did the best (P = 0.92, R = 0.97, F1= 0.95).

Table V. TabTransformer Per-Class Performance Metrics

Migraine Subtype	Precision	Recall	F1-Score
Basilar-type aura	0.94	0.87	0.91
Familial hemiplegic	0.84	0.74	0.79
Migraine w/o aura	0.93	0.89	0.91
Other	0.93	0.83	0.87
Sporadic hemiplegic	0.91	0.67	0.77
Typical aura w/ migraine	0.92	0.97	0.95
Typical aura w/o migraine	0.95	0.92	0.94

G. Feature Importance Analysis

The best clinical predictors for migraine classification were found using TabNet's sequential attention masks. Headache Duration (importance score: 92), Photophobia/Light Sensitivity (88), Migraine Frequency (85), Nausea (82), and Headache Intensity (80) were the top five factors. Aura Presence (78), Vomiting (75), Phonophobia/Sound Sensitivity (72), Dizziness (68), and Visual Disturbances (65) were other significant characteristics. These rankings confirm the clinical significance of TabNet's learnt representations since they are highly congruent with established clinical knowledge and ICHD-3 diagnostic criteria. However, without additional XAI tools like SHAP, TabTransformer's attention weights represent contextual relationships between feature embeddings rather than direct feature contributions, making simple clinical interpretation much more difficult.

H. Limitations and Future Work

Based on its synthesis from a single source, the dataset might not accurately represent clinical variation across a range of demographic and geographic populations in the actual world. To verify generalizability, independent multi-center EHR datasets must undergo external validation. The

performance gap may be significantly reduced by GPU-accelerated training with longer epochs and thorough hyperparameter adjustment. TabTransformer was trained under computational limitations (15 epochs, CPU, no class weighting). Future research should include ensemble architectures that combine both models, federated learning for privacy-preserving clinical deployment, multimodal data integration (neuroimaging, genomics, wearable sensor data), prospective clinical validation with neurologist ground truth, and fairness audits across demographic subgroups.

V. CONCLUSION

This study used a large-scale synthesis clinical dataset of 100,000 patient records across seven migraine subtypes to give the first systematic comparison of TabNet and TabTransformer for early migraine subtype prediction. With 99.50% test accuracy, 0.9999 ROC-AUC, and nearly flawless per-class performance across all seven subtypes, TabNet clearly and consistently outperformed all evaluation metrics. TabTransformer provided less comprehensible attention representations and demonstrated key memory failures for clinically significant uncommon subtypes, although achieving competitive overall accuracy(92.38%).

These results demonstrate that for clinical tabular classification tasks, sequential attention mechanisms outperform transformer self-attention, and that TabNet's architectural features—sparse instance-wise feature selection, integrated interpretability, balanced class handling, and effective large-batch training—are well suited to the requirements of clinical decision support deployment.

The verified TabNet model offers a solid basis for creating automated migraine diagnosis tools that could improve decision-making at the specialist level, shorten diagnostic delays, and provide underprivileged groups with high-quality care. Prior to practical implementation, additional clinical validation and multi-center assessment are advised.

Acknowledgment

The authors express sincere gratitude to their supervisor Jabunnesa Jahan Sara (Senior Lecturer) and co-supervisor Md. Rashedul Islam (Assistant Professor) of the Department of Computer Science and Engineering, IUBAT, for their invaluable guidance. The authors also thank Prof. Dr. Utpal Kanti Das, Chairman, for continuous encouragement throughout this research.

REFERENCES

1. L. J. Stovner et al., "Global prevalence of migraine and tension-type headache," The Lancet Neurology, vol. 21, no. 10, pp. 954–963, Oct. 2022, doi: 10.1016/S1474-4422(22)00hulk250-2.
2. Headache Classification Committee of the International Headache Society, "The International Classification of Headache Disorders, 3rd edition (ICHD-3)," Cephalalgia, vol. 38, no. 1, pp. 1–211, 2018, doi: 10.1177/0333102417738202.
3. S. Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in Proc. AAAI Conf. Artif. Intell., vol. 35, no. 8, pp. 6679–6687, 2021.
4. X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "TabTransformer: Tabular data modeling using contextual embeddings," arXiv preprint arXiv:2012.06678, 2020.
5. A. Ashokan et al., "Machine learning for migraine classification in resource-limited healthcare settings," Applied Sciences, vol. 15, no. 3, 2025.
6. M. Mudassir and M. Rahman, "Disease prediction using TabNet on structured clinical datasets," Computers in Biology and Medicine, vol. 172, 2024.
7. C. Sanchez-Sanchez, J. García-González, and E. Ascar, "Artificial neural networks for migraine classification based on patient-reported symptoms," Applied Sciences, vol. 10, no. 20, 2020.
8. M. Ashina, "Migraine," New England Journal of Medicine, vol. 383, pp. 1866–1876, 2020.
9. Y. Zhang, Z. Dong, and J. Wang, "Machine learning approaches for migraine classification using clinical features," Frontiers in Neurology, vol. 13, 2022.

10. P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," Nature Medicine, vol. 28, pp. 31–38, 2022.
11. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
12. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD, pp. 785–794, 2016.
13. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.