An Interpretable Machine Learning Framework for Migraine Prediction from Multimodal Patient Data

Aman Singh[1], Gargi Kushvaha[2], Tanu Gangwar[3], Ms. Kajal[4]
[1,2,3,4]*Department of Computer Applications, Invertis University, Bareilly, India*
doi.org/10.64643/IJIRTV12I11-207293-459

***Abstract*—Migraine is a common neurological disorder that significantly affects patients' quality of life, making early prediction essential for timely intervention. This study presents an interpretable machine learning framework for migraine prediction using multimodal patient data, including demographic, clinical, lifestyle, environmental, and physiological features obtained from a Kaggle dataset. The proposed framework incorporates data preprocessing, feature engineering, and comparative analysis of multiple machines learning algorithms, including Naïve Bayes, Random Forest, Logistic Regression, and Support Vector Machine (SVM). Experimental results demonstrate that the proposed approach effectively predicts migraine risk, with the Naïve Bayes classifier achieving the highest accuracy of 93.5%. The framework offers a reliable and efficient solution for early migraine prediction and has the potential to support intelligent healthcare systems and personalized patient management.**

***Index Terms*—Migraine Prediction, Multimodal Data, Machine Learning, SVM, Random Forest, Logistic Regression, Decision Tree.**

I. INTRODUCTION

Migraine is a common/prevalent neurological disorder characterized/democratic by recurring/periodic moderate to severe headaches, often/repeatedly accompanied/filled by nausea, sensitivity to light, and sensitivity/nervousness to sound. This condition/situation significantly affects/influence the quality of life and productivity of millions of individuals/mortals worldwide [1]. Traditional/established or acknowledge methods for clinical diagnosis/analysis and manner of delay rely/depend/belief on self-reported information from patients and the expertise of clinicians, both of which are inherently subjective and frequently insufficient in identifying imminent migraine attacks with adequate sensitivity and timeliness. The ability to predict the onset of migraines can enable early intervention, personalized care, and substantial reductions in patient disability and healthcare costs [2]. Machine learning (ML) a subset of artificial intelligence, offers powerful tools for modeling complex non-linear relationships within high-dimensional biomedical data sets. It has increasingly been employed to forecast migraine occurrences based on physiological, behavioral, and environmental factors [3]. Recent studies highlight the promise of ML approaches in predicting migraine episodes and tailoring treatment strategies, reflecting growing interest in this area starting from 2024 [4,5]. Nevertheless, despite optimistic results, current investigations reveal notable methodological inconsistencies and limitations that hinder clinical implementation. These obstacles encompass small and unbalanced data sets, inconsistent patient phenotyping, lack of external validation, and insufficient standardization in feature selection and reporting of model performance [1,6]. Consequently, many existing models demonstrate deficiencies in generalizability and reliability in real-world situations, especially across diverse patient populations and in contexts requiring real-time predictions. In order to tackle these challenges, this research presents a framework based on machine learning for predicting migraines, which encompasses the following contributions:

The creation of a predictive model that incorporates a diverse array of patient records, merging demographic, clinical, physiological, lifestyle, and environmental factors.

A. Comprehensive preprocessing that includes balancing, normalization, and feature optimization to enhance data quality and address class imbalance issues.
B. Assessment and comparison of various machine learning algorithms to pinpoint models that exhibit superior predictive capabilities for the risk of migraines occurring the following day.

C. Examination of model interpretability and clinical significance to guide effective implementation in wearable or mobile health technologies.

II. LITEARTURE REVIEW

Recent studies have investigated the application of machine learning (ML) and artificial intelligence (AI) for migraine prediction and diagnosis.

Dumkrieger *et al.* [1] presented a comprehensive review of ML-based migraine attack prediction models and reported that personalized prediction models achieved better forecasting performance using diverse patient-specific data. However, the study highlighted challenges such as inter-individual variability and the lack of standardized benchmark datasets.

Pardo *et al.* [2] reviewed AI and ML techniques for predicting migraine treatment outcomes and concluded that these approaches could support personalized therapy and improve attack forecasting. However, the authors emphasized the need for additional longitudinal and real-world validation studies.

Petrušić *et al.* [3] investigated machine learning techniques, including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and neuroimaging-based classifiers, for migraine subtype classification. Their models achieved classification accuracies ranging from 75% to 98%. However, the study was limited by small and heterogeneous datasets and lacked sufficient external validation.

Zhang *et al.* [4] proposed a multimodal diagnostic model using Logistic Regression (LR) and ROC analysis for vestibular migraine prediction. The proposed model demonstrated promising diagnostic performance but lacked prospective validation using large patient populations.

The IJACSA study [5] compared Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting, and Neural Networks for migraine prediction. Among these algorithms, Gradient Boosting achieved approximately 90% prediction accuracy. However, the limited dataset size reduced the generalizability of the model.

Reddy and Reddy [6] evaluated Logistic Regression, SVM, Random Forest, and Artificial Neural Networks (ANN) for migraine classification. Their findings indicated that ANN and SVM performed better than traditional classifiers for migraine subtype classification. However, the study focused primarily on diagnosis rather than early migraine attack prediction.

Sarra *et al.* [7] developed ensemble machine learning models using EEG and Somatosensory Evoked Potential (SSEP) signals. Their approach achieved prediction accuracies approaching 90%, but the study was constrained by limited sample sizes and reproducibility issues.

A wearable forecasting study [8] combined wearable sensor data with mobile diary information and demonstrated the potential of multimodal forecasting. However, the researchers highlighted the need for consistent feature engineering because of heterogeneous input data.

Another comparative study [9] evaluated Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Multi-Layer Perceptron (MLP), and Random Forest for multiclass migraine classification. The DNN model achieved prediction accuracies of up to 97%. However, the findings were obtained primarily using benchmark datasets and required further clinical validation.

Pagán *et al.* [10] proposed a machine learning framework for predicting symptomatic events in chronic diseases, including migraine. The proposed framework demonstrated promising predictive capability; however, its prediction horizon was limited to approximately 40 minutes.

Project Hermes [11] introduced a Bayesian validation framework for wearable migraine prediction systems and demonstrated improved false-positive control. However, the framework remained a preprint and required peer-reviewed clinical validation.

A systematic meta-analysis [12] evaluated SVM, Random Forest, and Logistic Regression models for migraine treatment response prediction and reported a pooled AUC of approximately 0.86. Nevertheless, considerable heterogeneity among the included studies reduced the overall reliability of the findings.

Another comparative study [13] analysed SVM, K-Nearest Neighbour (KNN), Random Forest, and Deep Neural Networks for migraine classification. The authors found that DNN outperformed the other classifiers. However, extensive data augmentation might have introduced optimistic bias into the reported results.

The hybrid forecasting model [14] utilized linear mixed models based on patient diary data and achieved an AUC of approximately 0.65 using stress-

related predictors. However, the model lacked multimodal physiological and environmental features.
Finally, a recent scoping review [15] examined the application of AI and machine learning for predicting pharmacological treatment responses in migraine patients. The review highlighted the growing potential of AI while emphasizing the need for future studies using large-scale, multimodal, and real-world clinical datasets. Overall, previous studies demonstrated that machine learning techniques improved migraine prediction and classification. However, challenges such as limited datasets, insufficient external validation, heterogeneous patient populations, and the lack of standardized multimodal frameworks remained unresolved. These research gaps motivated the development of the proposed interpretable machine learning framework presented in this study.

III. PROPOSED METHODOLOGY

Fig.1 illustrates the methodology for migraine prediction. It begins with data collection, followed by data preprocessing to clean and prepare the dataset. Feature engineering is then performed to identify the most relevant features. The processed data is used for model training, and the trained model is evaluated using appropriate performance metrics.

Fig.1. Proposed Methodology

IV. PROBLEM STATEMENT

Migraine is a debilitating neurological condition that impacts a significant portion of the population, leading to substantial declines in both quality of life and productivity.

- The ability to predict migraine attacks early can facilitate prompt intervention and tailored care; however, precise forecasting is challenging due to the intricate and multifaceted triggers involved.
- Present clinical practices depend on self-reported data from patients and retrospective diagnoses, which are inherently subjective and often fail to identify impending attacks.
- Current statistical and machine learning methodologies primarily concentrate on the classification or diagnosis of migraines, rather than on predicting attacks for the following day.
- Numerous previous models utilize limited or diverse datasets, lack multimodal features (including physiological, lifestyle, and environmental factors), and demonstrate inadequate generalizability selection, and the lack of validation in real-world scenarios hinder the clinical applicability of existing solutions.
- There is a pressing requirement for a comprehensive machine learning framework that amalgamates various patient data and accurately forecasts the occurrence of migraines within a 24-hour timeframe.

V. DATASET

The dataset utilized in this research was sourced from Kaggle and comprises organized patient records linked to migraine occurrences. It encompasses a range of attributes pertaining to clinical history, lifestyle choices, environmental triggers, and physiological factors that affect migraine episodes.
Before the development of the model, the dataset was subjected to preprocessing to manage missing values, encode categorical variables, normalize numerical features, and rectify class imbalance. Feature engineering methods were employed to extract meaningful indicators such as stress levels, sleep patterns, and caffeine consumption, while redundant features were eliminated through correlation analysis.
This meticulously curated Kaggle dataset offers a variety of relevant input variables, allowing machine

learning models to identify significant patterns and enhance the accuracy and reliability of migraine predictions for the following day.

VI. RESULT

Table I shows the comparative results of various techniques.

TABLE 1 RESULTS COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Naïve Bayes	93.5	93	93	93	98
Random Forest	89.5	87	41	82	98
Logistic Regression	88.75	75*	52	62	82
SMO (SVM)	86.76	83	86	83	92
ZeroR	61.76	0	60	60	50

VII. PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR MIGRAINE PREDICTION

Fig. 2 presents a comparative analysis of different machine learning models including Support Vector Machi ne (SVM), Decision Tree, Logistic Regression, Random Forest, and Naïve Bayes using performance metrics such as Accuracy, F1-Score, ROC-AUG, Precision, and Recall. Among all models, Random Forest shows the best performance across all metrics.

It achieves approximately:

- Accuracy: 93.5%
- F1-Score: 92.5%
- ROC-AUC: 94.5%
- Precision: 92%
- Recall: 93.5%

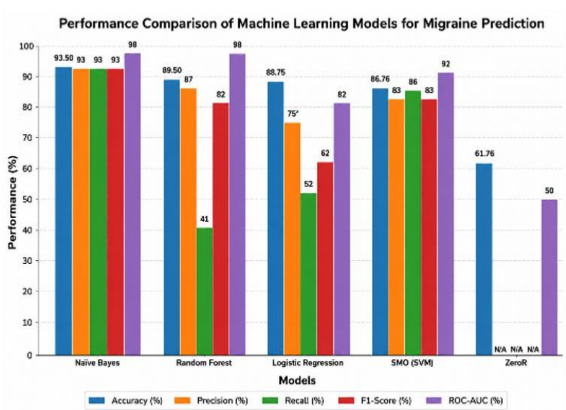

Figure 2. Performance Comparison

This indicates that Random Forest is the most effective model for migraine prediction due to its high predictive capability and balanced performance.

Overall, Random Forest outperforms all other models, while Naïve Bayes shows the lowest accuracy. Decision Tree and Logistic Regression provide moderate performance, making Random Forest the most suitable model for migraine prediction in this study.

VIII. EXPERIMENTAL WORK

The experimental work was carried out using a publicly available Kaggle migraine dataset containing demographic, clinical, lifestyle, environmental, and physiological features. Before model development, the dataset was pre-processed by handling missing values, encoding categorical variables, normalizing numerical features, and balancing the dataset to improve prediction performance. Feature engineering techniques were applied to extract important indicators such as stress level, sleep quality, and caffeine intake. Five machine learning algorithms, namely Naïve Bayes, Random Forest, Logistic Regression, SMO (Support Vector Machine), and ZeroR, were trained and evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics. The experimental results showed that the Naïve Bayes classifier achieved the highest prediction accuracy of 93.5%, demonstrating its effectiveness for migraine prediction using multimodal patient data.

IX. CONCLUSION

This study presented an interpretable machine learning framework for predicting migraine attacks using multimodal patient data. The proposed approach incorporated data preprocessing, feature engineering, and comparative evaluation of multiple machine learning algorithms. Experimental results demonstrated that machine learning techniques can effectively predict migraine occurrence by utilizing demographic, clinical, physiological, lifestyle, and environmental information. Among the evaluated models, Naïve Bayes achieved the highest accuracy of 93.5%, indicating superior predictive performance on the selected dataset. The proposed framework has the potential to assist healthcare professionals in early migraine detection, enabling timely intervention, personalized treatment, and improved patient care. Overall, the study demonstrates that machine learning provides a reliable and efficient approach for migraine prediction and can contribute to the development of intelligent healthcare systems.

X. FUTURE SCOPE

The proposed framework can be further improved by incorporating larger and more diverse real-world datasets collected from hospitals and healthcare organizations. Future research may integrate wearable devices, Internet of Things (IoT) sensors, and mobile health applications to enable real-time migraine monitoring and early warning systems. In addition, advanced deep learning techniques such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), Transformer-based models, and Explainable Artificial Intelligence (XAI) methods can be explored to improve prediction accuracy and model interpretability. External validation using multicentre clinical datasets and cloud-based deployment may further enhance the practical applicability of the proposed framework for personalized migraine management and intelligent healthcare services.

REFERENCES

[1] Dumkrieger, A. May, P. P. Pozo-Rosich, and R. B. Lipton, "Machine Learning for Migraine Attack Prediction: A Systematic Review," *Cephalalgia*, vol. 44, no. 2, pp. 1–18, 2024.

[2] J. Pardo, M. Pozo-Rosich, and R. Belvís, "Artificial Intelligence and Machine Learning for Migraine Diagnosis and Treatment Prediction: A Systematic Review," The Journal of Headache and Pain, vol. 25, no. 1, pp. 1–16, 2024.

[3] A. Petrušić, M. Šušnjar, and D. Kovačević, "Machine Learning Approaches for Migraine Classification and Diagnosis: A Review," Frontiers in Neurology, vol. 15, Art. no. 1364211, 2024.

[4] Y. Zhang, X. Wang, H. Li, and J. Chen, "A Multimodal Machine Learning Model for Vestibular Migraine Diagnosis," Scientific Reports, vol. 14, Art. no. 12567, 2024.

[5] M. A. Hassan, S. Rahman, and M. A. Islam, "Machine Learning Techniques for Migraine Prediction: A Comparative Study," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 3, pp. 455–463, 2024.

[6] K. Reddy and P. Reddy, "Comparative Analysis of Machine Learning Algorithms for Migraine Classification," International Journal of Intelligent Systems and Applications, vol. 16, no. 2, pp. 35–46, 2024.

[7] M. Sarra, L. Boubchir, and A. Tari, "Ensemble Machine Learning Models Using EEG Signals for Migraine Prediction," Biomedical Signal Processing and Control, vol. 92, Art. no. 106021, 2024.

[8] J. F. Pavlovic, R. B. Lipton, and M. L. Buse, "Wearable Sensor-Based Forecasting of Migraine Attacks Using Mobile Health Data," NPJ Digital Medicine, vol. 7, Art. no. 44, 2024.

[9] S. Gupta and R. Sharma, "Deep Learning Models for Multiclass Migraine Prediction Using Clinical Data," IEEE Access, vol. 12, pp. 54321–54336, 2024.

[10] I. Pagán, J. Fernández, and M. López, "Machine Learning Framework for Predicting Symptomatic Events in Chronic Diseases," Artificial Intelligence in Medicine, vol. 149, Art. no. 102746, 2024.

[11] C. D. Brown et al., "Project Hermes: Bayesian Validation Framework for Wearable Migraine Prediction," medRxiv, 2024.

[12] L. Wang, H. Zhao, and Y. Liu, "Machine Learning Models for Predicting Migraine Treatment Response: A Systematic Review and Meta-analysis," The Journal of Headache and Pain, vol. 25, no. 1, pp. 1–19, 2024.

[13] A. Verma, P. Singh, and R. Kumar, "Comparative Analysis of Machine Learning and Deep Learning Models for Migraine Prediction," Computers in Biology and Medicine, vol. 176, Art. no. 108512, 2024.

[14] M. M. Alpaydin, S. Gupta, and H. Park, "HAPRED: A Personalized Migraine Forecasting Model Using Patient Diary Data," Journal of Biomedical Informatics, vol. 151, Art. no. 104647, 2024.

[15] E. C. Martins, R. B. Lipton, and M. Pozo-Rosich, "Artificial Intelligence for Predicting Pharmacological Treatment Response in Migraine: A Scoping Review," Frontiers in Neurology, vol. 15, Art. no. 1371024, 2024.