



Review Article

Headache Pain Res 2025;26(1):48-65
pISSN: 3022-9057 · eISSN: 3022-4764
<https://doi.org/10.62087/hpr.2024.0024>

The Current Role of Artificial Intelligence in the Field of Headache Disorders, with a Focus on Migraine: A Systemic Review

Wonwoo Lee¹, Min Kyung Chu²

¹Department of Neurology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

²Department of Neurology, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

The application of artificial intelligence (AI) in the field of headache disorders, particularly migraine, is rapidly expanding, and AI has demonstrated significant potential for diagnosis, treatment, and research. This review examines the current role of AI in migraine management, categorizing AI applications into diagnosis and classification, assessment of treatment response, prediction of migraine attacks, and research. A systematic review of literature published between 2000 and 2024 was conducted, following PRISMA guidelines and utilizing the snowball technique. Of the 398 articles identified, along with five additional articles, 61 were finally reviewed. The results highlight promising AI applications, including the use of patient questionnaires, natural language processing, and imaging for migraine diagnosis, as well as predicting treatment responses and forecasting migraine attacks. Nonetheless, challenges remain in improving the accuracy, generalizability, validation, and clinical relevance of AI applications. Despite the substantial promise of AI for improving migraine management, it does not always guarantee better results than traditional methods. Careful consideration of the study design and method selection is crucial. Additionally, the interpretation of AI-generated results, particularly those from generative models, requires caution to avoid potential pitfalls.

Keywords: Computational, Deep learning, Large language models, Machine learning, Neural network

INTRODUCTION

The evolution of artificial intelligence (AI) has been transformative, significantly impacting various aspects of the medicine, including diagnosis, treatment, research, and the development of medical devices. However, the application of AI in the field of headache disorders, including migraine, has been relatively slow. A meta-analysis published

in 2020 revealed that only four (<1%) of the 985 selected articles published on Google Scholar between 2010 and 2019 that utilized deep learning (DL) techniques focused on migraine. In contrast, 303 (40%), 161 (21%), and 131 (18%) of these articles addressed Alzheimer's disease, autism, and epilepsy, respectively.¹ Nevertheless, research, tools, and applications related to migraine and headache disorders have expanded considerably since then, leading

Received: August 15, 2024; **Revised:** October 8, 2024; **Accepted:** November 23, 2024

Correspondence: Min Kyung Chu, M.D., Ph.D.

Department of Neurology, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
Tel: +82-2-2228-1600, Fax: +82-2-393-0705, E-mail: chumk@yonsei.ac.kr

© 2025 The Korean Headache Society

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

to a significant increase in published studies. According to a systematic review of computerized migraine diagnostic tools, the number of such tools has increased by 4.5 times since 2005, compared to the period before 2005.²

The current concept of AI and its application in the field of headache disorders is summarized in Figure 1. Briefly, AI can be categorized into symbolic and statistical methods. The symbolic method is based on logic and rule-based reasoning, using knowledge as inputs to produce knowledge that can be directly interpreted.³ Statistical methods generally rely on raw, continuous inputs and use statistical techniques to produce associations that need to be interpreted with background knowledge.

Examples of symbolic AI include Deep Blue for chess gameplay and MYCIN in the medical field, a computer-based consultation system designed to assist physicians

in the diagnosis and therapy selection for patients with bacterial infections.⁴

The evolution of computer systems has driven the rapid advancement of AI technologies, particularly in the area of statistical AI. Statistical methods can be divided into ‘supervised’ and ‘unsupervised’ learning, based on whether they have answers, known as ‘labels.’ Machine learning (ML) is a type of statistical AI that involves algorithms for data-driven pattern analysis, decision-making, and prediction. Among ML algorithms, neural networks are models inspired by the human neural network. Among artificial neural networks (ANN), convolutional neural networks (CNN) are better suited for image analysis, while recurrent neural networks and long short-term memory networks are more appropriate for linear and wavelet data. DL refers to neural network algorithms with multiple, deep layers.

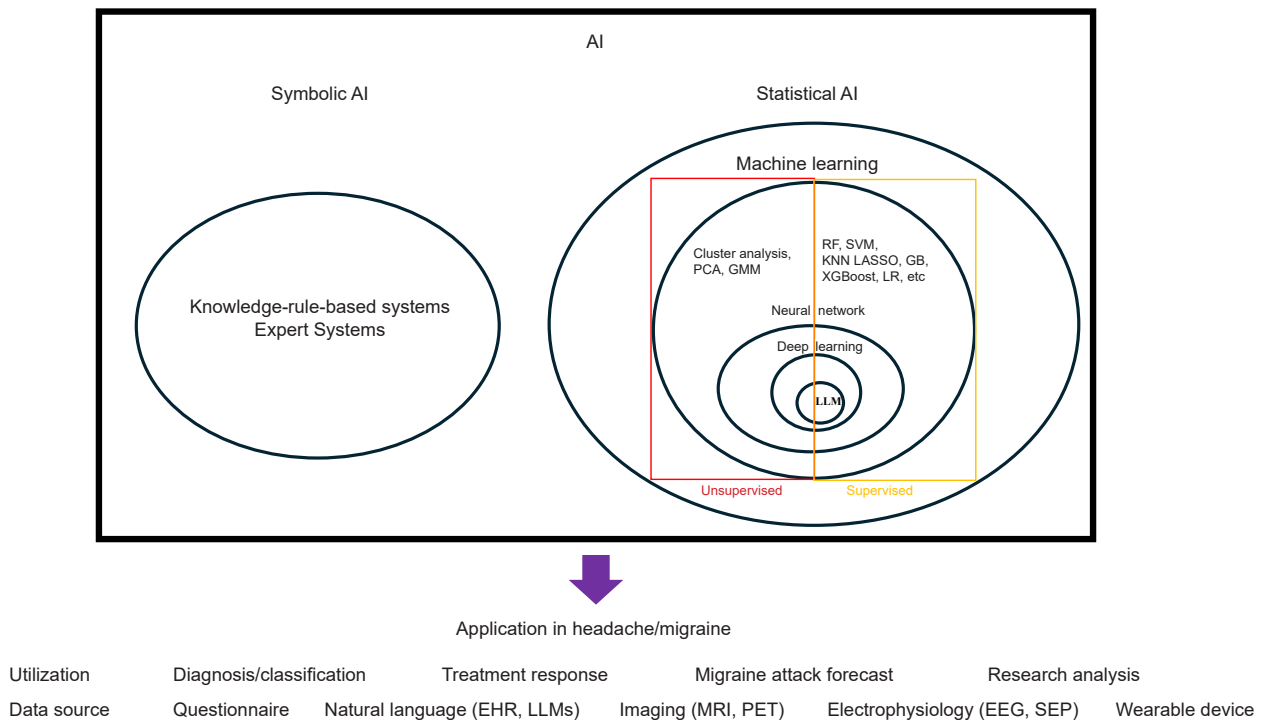


Figure 1. Schematic diagram of AI and its applications in the headache field AI can be divided into symbolic and statistical methods. Machine learning, neural networks, deep learning, and LLMs are examples of statistical methods. These methods can also be categorized as unsupervised or supervised based on their use of labeled data. The applications of AI in headache and migraine can be analyzed in terms of its utilization and the data source.

AI, artificial intelligence; PCA, principal component analysis; GMM, Gaussian mixture models; RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LASSO, least absolute shrinkage and selection operator; GB, gradient boosting; XGBoost, extreme gradient boosting; LR, logistic regression; LLM, large language model; EHR, electronic health records; MRI, magnetic resonance imaging; PET, positron emission tomography; EEG, electroencephalography; SEP, somatosensory evoked potentials.

Numerous DL architectures are available, each proven effective for specific type of data.

Utilization of AI in headache medicine can be categorized into several key areas: diagnosis or classification of headache disorders, assessment of treatment response, forecasting of migraine attacks, and as a tool for analysis during research. Regarding data sources and methods, AI applications utilize a range of inputs including questionnaires, language data (e.g., generative language models or electronic health records [EHR]), medical devices or tools such as magnetic resonance imaging (MRI), results from electrophysiology studies (e.g., electroencephalography [EEG], somatic evoked potential [SEP]), and wearable devices, either individually or in combination. This review aims to outline the current use and role of AI in the field of headache disorders, with a focus on migraine, and to discuss future perspectives.

METHODS

1. Search strategies

Although this is not a systematic review, the search process was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.⁵ A literature search was performed in PubMed using the following terms: ((migraine*) AND ((artificial*) OR (artificial intelligence*) OR (AI*) OR (deep learning*) OR (machine learning*) OR (artificial intelligence [MeSH Terms]) OR (AI [MeSH Terms]) OR (deep learning [MeSH Terms]) OR (machine learning [MeSH Terms])))

The search was restricted to literature published between 1 January 2000 and 31 July 2024. Only abstracts in the English language were included for review.

2. Inclusion and exclusion criteria

In reviewing abstracts, only studies that explicitly included AI, ML, or DL methods in their analytical processes were considered for inclusion. Studies where the authors used ML or DL methods but did not specify this in the abstract were excluded. Semi-automated approaches that involved computational methods alongside expert-suggested algorithms were included if they were specified as AI-based methods or if they were well-organized for comparative

review. Medical tools, including imaging techniques such as MRI and positron emission tomography (PET), electrophysiology methods such as EEG and SEP, magnetoencephalography (MEG), and other devices such as wearable technologies, were included if the analytical methods utilized AI techniques. Review articles, editorials, opinions, and viewpoints were considered for snowballing purposes but were generally excluded from the systematic review. Additionally, studies for which the full text was unavailable or not published in English were excluded. Manuscripts were further excluded if they employed inappropriate methodologies, such as not applying the International Classification of Headache Disorders, 3rd edition (ICHD-3) criteria, or if they involved improper headache diagnosis or did not specify headache participants.

RESULTS

1. Search results and article inclusion/exclusion

Of the 398 articles identified, one was a duplicate and 317 were excluded based on abstract review. Of the remaining 80 articles, six were review articles, four were editorials or opinion pieces, two did not utilize AI methodology, one was not related to the headache field, seven did not adhere to ICHD-3 criteria or did not specify headache diagnosis methods, and four had full texts that were unavailable. An additional five articles were identified through the snowball technique. In total, 61 articles published between 2002 and 2024 were included in the review, with the majority published since 2020. The PRISMA flow chart outlining the selection process is shown in [Figure 2](#). The summaries of the included studies are demonstrated in [Table 1](#).

ARTIFICIAL INTELLIGENCE APPLICATIONS IN THE DIAGNOSIS OF HEADACHE DISORDERS

1. Questionnaire/survey

Traditionally, questionnaires have been valuable tools in aiding the diagnosis of headache disorders, given that such diagnoses are typically based on clinical profiles. Furthermore, previously collected data from these questionnaires facilitates the swift and effective application of AI technology.

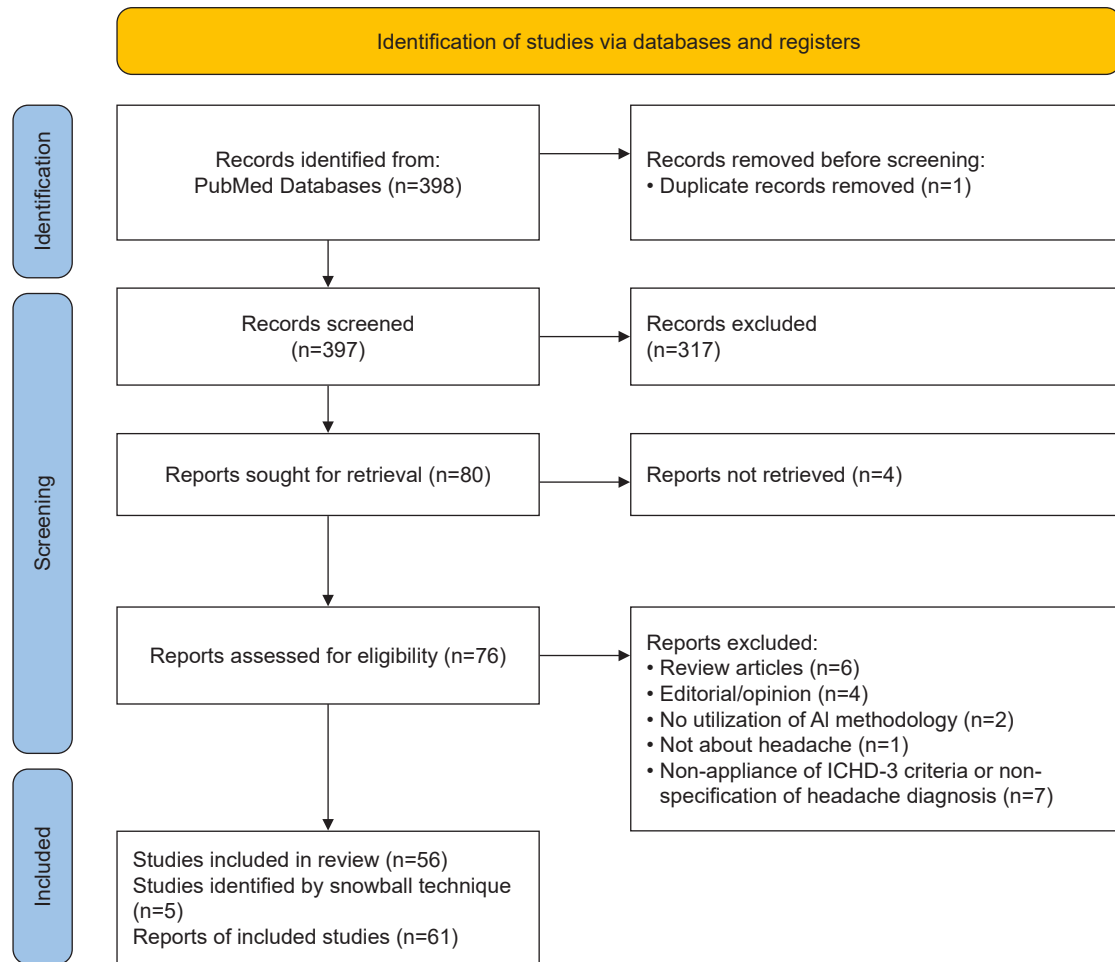


Figure 2. PRISMA 2020 flow diagram.⁵

AI, artificial intelligence; ICHD-3, International Classification of Headache Disorders, 3rd edition.

The number of items in the questionnaires varied from 17–22^{6–12} to 75.¹³ While the details differed, all questionnaires included demographic data (age, sex), headache characteristics, duration, frequency, and accompanying symptoms.

The number of participants and the number of classifying groups varied across studies. Liu et al.⁶ distinguished between 84 migraine and 89 tension-type headache (TTH) participants using a 19-item questionnaire. Simić et al.⁷ utilized a 20-item questionnaire to classify 1,022 subjects, identifying 169 with migraine, 224 with TTH, and 186 with other headache types. Kwon et al.¹³ employed a 75-item questionnaire from a headache center to classify 2,162 individuals with headache disorders, including migraine, TTH, trigeminal autonomic cephalalgias (TAC), epicranial headaches, and thunderclap headaches.

Most studies utilized supervised ML methods, including decision trees (DTs), random forests (RFs), gradient boosting (GB), logistic regression (LR), and support vector machines (SVMs). The performances were presented with sensitivity, specificity, accuracy, area under the receiver operating characteristic curve (AUROC), and F1 score. The F1 score is the harmonic mean of precision and recall ($2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$), where precision is calculated as ($\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$) and recall is calculated as ($\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$). The F1 score is particularly useful for evaluating predictive performance, especially when the dataset is imbalanced.

Kwon et al.¹³ used a stacked classifier model with four layers of eXtreme Gradient Boosting (XGBoost) classifiers, each layer classifying migraine, TTH, TAC, epicranial headaches, and thunderclap headaches. Different features

Table 1. Summary of studies involving AI in the headache field

Purpose	Data source	Study	Year	AI method	AI method specification
Diagnosis	Questionnaire	Kwon et al. ¹³	2020	ML	Stacked classifier model with four layers of XGBoost classifiers, LASSO
	Questionnaire	Liu et al. ⁶	2022	ML	RF, GB, LR, SVM
	Questionnaire/NL	Katsuki et al. ¹⁴	2020	DL	NLP, ANN
	Questionnaire	Simić et al. ⁷	2021	Hybrid system	Calinski-Harabasz index, Analytical Hierarchy Process, and Weighted Fuzzy C-means Clustering algorithm (ML)
	Questionnaire	Katsuki et al. ¹⁰	2023	ML	GB, LR, Ridge Classifier, RF, Extra Trees Classifier, K Neighbors Classifier, Dummy Classifier, DT, SVM, AdaBoost Classifier, LDA, Naïve Bayes, QDA, best performance: GB
	Questionnaire	Katsuki et al. ⁸	2023	ML	Light GB machine, RF, LDA, Ridge Classifier, Extra Trees, GB Classifier, LR, AdaBoost Classifier, DT, K Neighbors, Naïve Bayes, Dummy Classifier, SVM, QDA, best performance: light GB machine classifier
	Questionnaire	Sasaki et al. ¹²	2023	ML	Light GB machine, RF, LDA, Ridge Classifier, Extra Trees, GB Classifier, LR, Ada Boost Classifier, DT K Neighbors, Naïve Bayes, Dummy Classifier, SVM, QDA, best performance: extremely randomized trees
	Questionnaire	Okada et al. ¹¹	2024	ML	Light GB machine classifier
	NL	Vandenbussche et al. ²⁰	2022	NLP/ML	NLP, LR, SVM
	NL (EHR)	Riskin et al. ¹⁹	2023	NLP/ML	Not specified
	Questionnaire/MRI	Chong et al. ²³	2021	ML	PCA, logistic classifier
	Clinical data/MRI	Dumkrieger et al. ²⁴	2023	ML	Ridge LR on principal component
	MRI	Rahman Siddiquee et al. ²⁵	2022	DL	ResNet-18
	MRI	Mitrović et al. ²¹	2023	ML	LDA
	MRI	Mitrović et al. ²²	2023	ML	SVM
	Resting-state fMRI	Chong et al. ²⁹	2017	ML	Diagonal QDA
	Resting-state fMRI	Yang et al. ³¹	2018	ML, DL	SVM, CNN
	Resting-state fMRI	Tu et al. ²⁶	2020	ML	Recursive feature elimination, SVM, LOOCV
	Resting-state fMRI	Nie et al. ^{27,28}	2021; 2023	ML	K-means clustering, hierarchical clustering, SVM
	Resting-state fMRI	Fernandes et al. ³⁰	2024	ML	Gaussian Process Classifier
	MEG	Hsiao et al. ³²	2022	ML	SVM
	MEG	Hsiao et al. ³³	2023	ML	DT, discriminant analysis, naïve Bayes classifiers, SVM, KNN
	EEG	Akben et al. ³⁹	2012	ML	MLP
	EEG (wearable)	Cao et al. ⁴⁰	2018	ML	LDA, KNN, MLP, Bayesian classifier, SVM
	EEG	Frid et al. ³⁷	2020	ML	Relif Family algorithm, SVM
	EEG	Aslan ³⁸	2021	ML	Rotation Forest, BFTree, RF, Bagging, AdaBoost, SPAARC, MultiBoost, Random Tree, NBTree ensemble classifiers
	EEG	Hsiao et al. ³⁵	2023	ML	DT, discriminant analysis, naïve Bayes classifiers, SVM, KNN
	EEG	Orhanbulucu et al. ³⁶	2023	DL	AlexNet, ResNet50, SqueezeNet
	SEP	Zhu et al. ⁴²	2019	ML, DL	RF, XGBoost trees, SVM, KNN, MLP, LDA, LR, CNN
	ECG	Chiang et al. ⁴¹	2022	DL	CNN
	Headache diary application/wearable device	De Brouwer et al. ⁴³	2022	ML	Knowledge-based classification, ML-based detection of activity, stress, sleep events

(Continued to the next page)

Table 1. Continued

Purpose	Data source	Study	Year	AI method	AI method specification	
Treatment efficacy/response	Functional near-infrared spectroscopy	Chen et al. ⁴⁴	2022	ML	LDA, QDA	
	Web-based survey	Ashina et al. ⁴⁷	2024	ML	RF, LASSO	
	NL (EHR)	Hindiyeh et al. ⁴⁸	2022	NLP	Not specified	
	NL (social media)	Guo et al. ⁴⁹	2023	NLP	Transformer-based models	
	NL (EHR)	Chiang et al. ⁵⁰	2024	NLP framework	ClinicalBERT regression model, GPT-2 Question Answering model zero-shot, GPT-2 QA model few-shot training fine-tuned on clinical notes, GPT-2 generative model few-shot training fine-tuned on clinical notes	
	NL (generative LLM)	Moskatel and Zhang ⁵²	2023	LLMs	ChatGPT-3.5	
	NL (generative LLM)	Li et al. ⁵¹	2024	LLMs	ChatGPT-3.5, ChatGPT-4.0, Google Bard, Meta Llama2, and Anthropic Claude2	
	Clinical dataset	Ferroni et al. ⁵⁷	2020	ML	SVM, random optimization	
	Clinical dataset	Lu et al. ⁵³	2022	ML	SVM, DT, MLP	
	Clinical dataset	Gonzalez-Martinez et al. ⁵⁵	2022	ML	RF, Bayesian search optimization method	
	Clinical dataset	Stubberud et al. ⁵⁶	2022	ML, NLP	Multitask Gaussian process model, NLP	
	Clinical dataset	Ciancarelli et al. ⁵⁸	2022	Neural network	ANN	
	Clinical dataset	Martinelli et al. ⁵⁴	2023	ML, neural network	RF, SVM, ANN, adaptive neuro-fuzzy inference system, fuzzy c-means clustering	
	Clinical dataset/MRI	Tso et al. ⁶²	2021	ML	PCA, t-distributed stochastic neighbor embedding, KNN, XGBoost implemented GB DT	
	Migraine attack prediction	MRI, fMRI	Wei et al. ⁵⁹	2023	DL, ML	ResNet34, ResNet50, RexNet50, DenseNet121, 3D ResNet18,, best performance: ResNet-18 /SVM
Multimodal MRI		Wei et al. ⁶⁰	2024	ML	LASSO, LR, SVM-recursive feature elimination for Feature selection / LR, SVM, RF, DT, KNN, MLP elastic network, light GB machine, XGBoost for classification, best performance: RF	
PET		Marino et al. ⁶¹	2023	ML	CBDA	
Wearable device		Siirtola et al. ⁶⁵	2018	ML	QDA, LDA	
Headache diary application/wearable device		Stubberud et al. ⁶⁴	2023	ML	LR, SVM, RF, GB, Adaptive boosting, XGBoost, best performance: RF	
Headache diary application/weather data		Katsuki et al. ⁹	2023	ML, neural network	Generalized linear mixed model, feedforward neural network, XGBoost	
Research		Cortical-evoked potentials in response to repetitive visual/auditory stimulus	Thomas et al. ⁶⁸	2002	Neural network	Neural network model
		Mouse grimace scale	Chiang et al. ⁶⁷	2022	DL	ResNet-18
		Temporal multi-omics profile	Kogelman et al. ⁶⁶	2023	ML	Qlattice

AI, artificial intelligence; ML, machine learning; XGBoost, extreme gradient boosting; LASSO, least absolute shrinkage and selection operator; RF, random forest; GB, gradient boosting; LR, logistic regression; SVM, support vector machine; NL, natural language; DL, deep learning; NLP, natural language processing; ANN, artificial neural network; DT, decision tree; LDA, linear discriminant analysis; QDA, quadratic discriminant analysis; EHR, electronic health records; MRI, magnetic resonance imaging; PCA, principal component analysis; fMRI, functional MRI; CNN, convolutional neural network; LOOCV, leave-one-out cross-validation; MEG, magnetoencephalography; KNN, K-nearest neighbor; EEG, electroencephalography; MLP, multilayer perceptron; BFTree, best first decision tree; SPAARC, sequential pattern-aided adaptive response classification; NBTree, naïve Bayes decision tree; SEP, somatosensory evoked potentials; ECG, electrocardiogram; ClinicalBERT, clinical bidirectional encoder representations from transformers; GPT, generative pre-trained transformer; LLMs, large language models; PET, positron emission tomography; CBDA, Compressive Big Data Analytics.

were selected from the self-reported data at each layer using the least absolute shrinkage and selection operator (LASSO). The model achieved an accuracy of 81% for the test set. The sensitivity and specificity for migraine, TTH, TAC, epicranial headache, and thunderclap headache were 88% and 95%, 69% and 55%, 65% and 46%, 53% and 48%, and 51% and 51%, respectively.¹³

In contrast, Simić et al.⁷ proposed a hybrid system incorporating the Calinski-Harabasz index, Analytical Hierarchy Process, and Weighted Fuzzy C-means Clustering algorithm, an unsupervised ML method. The accuracy rates were 67% for migraine, 74% for TTH, and 86% for other primary headaches, with corresponding F1 scores of 75%, 74%, and 75%, respectively.⁷

The Japanese research group, led by Katsuki, Yamamoto, Sasaki, and Okada, along with other co-authors, has published multiple articles utilizing questionnaires and AI methods. In their first study, published in 2020, they used a combination of questionnaires, unstructured descriptions, and DL methods to classify primary headaches among 848 participants, with 46% diagnosed with migraine, 47% with TTH, and 5% with TAC.¹⁴ Natural language processing (NLP) was employed using the commercial DL framework, Prediction One, and an ANN model was applied. The model achieved an accuracy of 0.7759, a mean precision of 0.8537, a mean recall of 0.6086, and a mean F1 score of 0.6353.

In subsequent studies, the same group used a 17- or 22-item questionnaire along with multiple AI methods to classify five to six different outcomes: migraine and medication-overuse headache (MOH) separately or together as migraine/MOH, TTH, TACs, other primary headaches, and other headaches.¹⁰ Among the 6,058 participants, there were 4,829 cases of migraine, 834 cases of TTH, 78 cases of TACs, 38 cases of other primary headache disorders, and 279 cases of other headaches. The GB classifier yielded the highest c-statistic of 0.88. The c-statistic, equivalent to the AUROC, measures a classification model's ability to discriminate between classes, with higher values indicating better performance. The model's accuracy, sensitivity, specificity, precision, and F1-score were 93.7%, 84.2%, 84.2%, 96.1%, and 84.2%, respectively.

The AI model's performance was compared with that of non-headache specialists, and its usefulness in aiding headache diagnosis was evaluated using data from a study

of 4,000 patients.⁸ The light GB machine classifier achieved the highest c-statistic of 0.9203. The diagnostic accuracy of five non-headache specialists was then compared to that of the AI model using a sample of 50 patients. Without the AI model, the non-specialists' overall diagnostic accuracy was 46%, with a kappa value of 0.212. With the aid of the AI model, their accuracy and kappa value improved significantly to 83.2% and 0.678, respectively. External validation of the AI model's diagnostic performance using a sample of 59 participants demonstrated an overall accuracy of 94.92% and a kappa value of 0.65 (95% confidence interval [95% CI], 0.21–1.00) when compared to the ground truth. The sensitivity, specificity, precision, and F1-score for diagnosing migraines were 98.21%, 66.67%, 98.21%, and 98.21%, respectively.¹¹

The application of the system in pediatric and adolescent populations was also validated. Sasaki et al.¹² used multiple AI models to diagnose 909 participants aged 6 to 17 years, including 234 individuals with migraine. For the test dataset, the model achieved an accuracy of 94.5%, sensitivity of 88.7%, specificity of 96.5%, precision of 90.0%, and an F1-score of 89.4%.

However, non-AI methods and rule-based decision systems have also demonstrated impressive results. For example, a web-based headache diagnosis questionnaire validated by telephone interviews showed a sensitivity of 92.6%, a specificity of 94.8%, and a kappa coefficient of 0.875 for diagnosing migraine among 256 participants. For the diagnosis of TTH and probable migraine (PM), the sensitivity, specificity, and kappa coefficients were 78.4%, 98.4%, and 0.809, and 85%, 92.9%, and 0.757, respectively.¹⁵

Computerized systems based on expert opinions have also proven effective. In 2008, Maizels and Wolfe¹⁶ developed a Computerized Headache Assessment Tool (CHAT) using web-based questionnaires with branching questions based on headache frequency, duration, and ICHD criteria. Among 135 participants who completed CHAT and 117 who completed a diagnostic interview, CHAT correctly identified 35/35 cases (100%) of episodic migraine (EM), 42/49 cases (85.7%) of transformed migraine, 11/11 cases of chronic TTH, 2/2 cases of episodic TTH, and 1/1 case of episodic cluster headache (CH). It also identified medication overuse in 43/52 cases (82.7%), with the most common misdiagnosis being transformed migraine or new daily persistent headache.

In another study by Cowan et al.¹⁷, the concordance between a self-administered, computer-based diagnostic engine (CDE) and a semi-structured interview conducted by a headache specialist was assessed. The CDE, developed by the authors using a detailed DT, was completed by 212 participants, who also underwent an interview. For diagnosing migraine and PM, the CDE demonstrated a sensitivity of 90.1% and a specificity of 95.8%, with a concordance rate with SSI of $\kappa=0.83$ (95% CI, 0.75–0.91).

These expert-based systems, built on transparent decision-making processes using ICHD-3 criteria, exhibit high sensitivity and specificity. In contrast, AI operates as a “black-box” system, where the decision-making process is not easily interpretable. While AI models may demonstrate high accuracy, careful interpretation according to current knowledge is necessary, and biases of the data may result in subpar prediction results.¹⁸ Questions remain as to whether current AI offers real advantages beyond being novel and innovative. The challenge remains validating AI models and ensuring their effective application in real-world settings.

2. Natural language

Natural language as a data source in headache research holds significant potential, especially in aiding practitioners and saving time. Many patient interviews are naturally conducted in unstructured language, which doctors traditionally summarize and interpret to make a diagnosis. While structured questionnaires have been used to standardize this natural language, the raw language itself may contain even more valuable information. In this context, natural language includes any unstructured text, such as EHRs and generative large language models (LLMs). However, studies utilizing generative LLMs have predominantly focused on assessing treatment response rather than diagnosis and classification. Three studies were identified in the area of diagnosis and classification, with one integrating questionnaire data and natural language, as previously discussed in the questionnaire section.

Riskin et al.¹⁹ used US claims and EHR data from 2010 to 2012 to compare the efficacy of migraine identification. They defined “Traditional Real-World Evidence (RWE)” as the use of insurance claims or structured EHR data, while “Advanced RWE” was defined as the use of unstructured

EHRs. Although the exact AI-based technology was not specified, an ML algorithm was employed. Based on manual annotation by seven annotators, 2,642 migraine and 6,530 headache-related concepts were identified, and their recall rates were compared. “Traditional RWE” achieved recall rates of 66.6% and 29.6%, while “Advanced RWE” recalled 96.8% and 92.9%, respectively. The superior performance of “Advanced RWE” was consistent across the identification of six migraine-associated symptoms, with F1 scores ranging from 80.7% to 95.6%.

Vandenbussche et al.²⁰ conducted a web-based survey in which 81 migraine and 40 CH patients were asked to describe their headache disorders in detail. NLP was applied to analyze the narrative self-reports, focusing on lexical, semantic, and thematic properties. Lexicon-based sentiment analysis of attack descriptions revealed predominantly negative sentiments. For the classification of migraine and CH using features from the attack descriptions, LR and SVM algorithms demonstrated the best performance, with F1 scores ranging from 0.6 to 0.8. There was a significant difference between Dutch-speaking migraine and CH patients in how they described their disorder. Migraine patients used the Dutch word for “headache” more often, while CH patients more frequently used the word “pain.”

3. Imaging

Numerous studies have employed brain imaging techniques, such as MRI, functional MRI (fMRI), and PET, analyzed with ML and DL methods to differentiate and classify headache disorders, particularly migraine.

Mitrović et al.²¹ analyzed brain MRI data from a cohort including healthy controls (HCs) and patients with migraine with aura (MwA). Cortical thickness, surface area, and volume were compared using various ML methods.²¹ The best classification results were obtained with linear discriminant analysis (LDA), achieving 97% accuracy for MwA. Left temporal pole, right lingual gyrus, and left pars opercularis thickness were notable distinguishing features. Further research used the average Migraine Aura Complexity Score (MACS) from multiple MwA attacks and evaluated its correlation with 340 MRI features.²² Applying ML methods including SVM, a high coefficient of determination (0.89) was achieved, with 26 significant features including left parahippocampal mean Gaussian curvature,

left transverse temporal mean Gaussian curvature, left transverse temporal thickness, and left pars opercularis thickness ($p < 0.01$) strongly correlating with average MACS ($p < 0.05$).

Chong et al.²³ combined questionnaire data with T1-weighted MRI and diffusion tensor imaging, to distinguish between migraine and persistent post-traumatic headache (PTH) attributed to mild traumatic brain injury. A logistic classifier achieved an overall accuracy of 78%, with 97.1% accuracy for migraine and 64.6% for PTH. Critical features contributing to accuracy included responses related to anxiety on sports concussion Assessment Tool and decision-making difficulty on Beck Depression Inventory-13, as well as cortical brain regions such as the bilateral superior temporal gyrus, inferior parietal lobe, posterior cingulate cortex, and fiber tracts like the right anterior thalamic radiations and right superior longitudinal fasciculus. Additional study utilized clinical data, along with MRI measures of brain structure and functional connectivity.²⁴ A classifier using ridge LR on principal components achieved an average accuracy of 72% when using functional connectivity data, and 63.4% without it. In addition, a DL method was developed using a 3D ResNet-18 classifier to automatically identify features that differentiate MRIs of 95 migraine patients, 48 with acute PTH, 49 with persistent PTH, and 532 HCs. The 3D ResNet-18 classifier, an 18-layer CNN based DL architecture for image analysis, adapted for 3D convolutions, achieved an accuracy of 75%, a sensitivity of 66.7%, and a specificity of 83.3% in distinguishing migraine from HCs. The most significant biomarkers identified by the migraine classifier included the caudate, caudal anterior cingulate, superior frontal gyrus, thalamus, and ventral diencephalon.²⁵

Resting-state fMRI has been frequently analyzed in migraine research, utilizing various ML and DL techniques for feature extraction and classification. Several studies compared migraineurs and HCs.

Tu et al.²⁶ examined 70 migraine without aura (MwoA) patients and 46 matched HCs, identifying abnormal functional connectivity within the visual network (VN), default mode network (DMN), sensorimotor network (SMN), and fronto-parietal networks that distinguished migraineurs from HCs using an SVM model with 93% sensitivity and 89% specificity. The model was validated on an independent cohort of 19 MwoA patients and 19 additional con-

trols, achieving 84% sensitivity and specificity. To verify specificity, the model was tested on 18 MwoA patients and 76 non-migraine pain patients (with chronic lower back pain and fibromyalgia), demonstrating 78% sensitivity and 76% specificity for distinguishing migraineurs from non-migraineurs.

Nie et al.²⁷ applied both unsupervised and supervised ML techniques. Using an automatic segmentation algorithm, K-means clustering combined with hierarchical clustering identified 17 dynamic functional connectome patterns (DFCPs).²⁷ SVM was used to select optimal features from static functional connectivity strength and DFCP features and to classify migraine patients and HCs.²⁸

Chong et al.²⁹ used diagonal quadratic discriminant analysis (QDA), an ML algorithm to analyze functional connections from 33 seeded pain-related regions of 58 migraine patients and 50 HCs. Notably, those with a disease duration of more than 14 years were classified more accurately (96.7% vs. 82.1%).

MwA was also examined in several studies. Fernandes et al.³⁰ used Gaussian Process Classifier to differentiate between ictal and interictal periods in two patients with MwA.

Yang et al.³¹ analyzed the amplitude of low-frequency fluctuations, regional homogeneity, and regional functional correlation strength to distinguish 21 patients with MwoA, 15 with MwA, and 28 HCs. SVM classifier achieved an accuracy of 83.67%, whereas a CNN approach based on the Inception module improved accuracy to 86.18%.

4. Electrophysiology and magnetoencephalography

Wavelet data from electrophysiology studies, including EEG and SEP, have also been utilized for the diagnosis and classification of migraine. Analyzing these data often requires transformations, such as Fourier transformation, to process the complex signals. MEG has also been employed in the analysis of headache disorders and, in this review, is included in this section due to its time-dependent data acquisition characteristics. Studies utilizing EEG and MEG signals have been conducted to differentiate migraine from other conditions.

Hsiao et al.³² conducted multiple studies utilizing MEG. In 2022, resting-state MEG data from 70 HCs, 100 chronic migraine (CM) patients, 35 EM patients, and 35 FM pa-

tients were analyzed to calculate source-based oscillatory connectivity in relevant cortical regions.³² Using a SVM classifier, a model was developed to identify CM. The salience, SMN, and parts of the DMNs were key features differentiating CM from HCs, with classification performance showing an accuracy of $\geq 86.8\%$ and an area under the curve (AUC) of ≥ 0.9 . When comparing CM to EM, the model achieved an accuracy of 94.5% and an AUC of 0.96, and for CM versus FM, an accuracy of 89.1% and an AUC of 0.91. In 2023, resting-state MEG data of 70 HCs, 100 CM, 40 CM with FM, 35 FM, 30 chronic TTH, and 75 EM were analyzed.³³ Features were extracted and classified using ML algorithms including DT, discriminant analysis, naïve Bayes classifiers, SVM, and K-nearest neighbor (KNN). The best classification model distinguished CM from HCs with an accuracy of over 92.6% and an AUC of over 0.93. When validating CM classification against other groups, accuracy exceeded 75.7%, with an AUC greater than 0.8.

Although EEG is not routinely recommended in headache practice, its application in headache research has persisted.³⁴ EEG signals have been utilized to classify HCs, migraine patients, CM patients,^{35,36} and to differentiate between MwA and MwoA.³⁷ EEG signals were recorded during resting state, visual or auditory stimulation tasks, or non-painful, painful, and repetitive painful electrical stimulation. Various signal processing techniques were applied, such as the tunable Q-factor wavelet transform method to decompose EEG signals into sub-bands³⁸ and segmentation of a 3-minute EEG into 120 1-second segments, generating 325 functional connectivity values between electrode pairs.³⁷ Most studies employed ML models. However, in one study, EEG signals were transformed into scalogram-spectrogram images and classified using CNN architectures, including AlexNet, ResNet50, and SqueezeNet.³⁶

Akben et al.³⁹ in 2012 compared different flash stimulation frequencies (2 Hz, 4 Hz, and 6 Hz) and durations (2 seconds, 4 seconds, 6 seconds, and 10 seconds) to determine the most effective conditions for detecting migraine. EEG was recorded during flash stimulation in 15 migraine patients and 15 HCs. The power spectral density estimate was computed, and a multilayer perceptron (MLP) neural network was used for classification. The study found that a 4 Hz flash stimulation frequency and an 8-second duration were most effective in detecting migraine, particularly at

the beta band of the T5-T3 channel.

In another study by Cao et al.⁴⁰, a wearable, wireless EEG device (Mindo-4S) was used to record EEG signals from the prefrontal (Fpz) and occipital (O1, Oz, O2) regions to differentiate 40 MwoA patients from 40 HCs. EEGs from interictal, pre-ictal, ictal, and post-ictal phases were processed, and a binary classification model was developed using LDA, KNN, MLP, Bayesian classifier, and SVM. The SVM demonstrated the highest accuracy ($76\% \pm 4\%$) for classifying interictal and pre-ictal phases using prefrontal EEG complexity.

Chiang et al.⁴¹ analyzed the electrocardiogram (ECG) data of 17,840 participants with MwA and 22,162 participants with MwoA, excluding those with a history of atrial fibrillation (AF). The team employed an AI-ECG algorithm, developed using a CNN-based approach, to calculate the probability of concurrent paroxysmal or impending AF in ECGs showing normal sinus rhythm. The AF prediction model output was significantly higher in the MwA group compared to the MwoA group (mean [standard deviation], 7.3% [15.0%] vs. 5.6% [12.4%]; mean difference [95% CI], 1.7% [1.5%–2.0%]; $p < 0.001$). These differences remained significant even after adjusting for vascular comorbidities, suggesting a higher probability of concurrent paroxysmal or impending AF in individuals with MwA compared to those with MwoA.

Although not as extensively researched, SEP have also been investigated in the context of migraine. Zhu et al.⁴² utilized SEP data to differentiate between 42 migraine patients (29 in the interictal phase and 13 in the ictal phase) and 15 HCs. The right median nerve SEPs were recorded, and features in both the time and frequency domains were selected through a feature selection method. The data were then classified using various ML algorithms, including RF, XGBoost trees, SVM, KNN, MLP, LDA, and LR. The classification accuracies for distinguishing HCs, ictal, and interictal phases ranged from 51.2% to 72.4%. After model and feature selection, the accuracy improved to 89.7% for HC-ictal, 88.7% for HC-interictal, 80.2% for ictal-interictal, and 73.3% for HC-ictal-interictal classification. Interestingly, a tested CNN-based model showed lower performance compared to the ML-based models.

5. Wearables and other devices

De Brouwer et al.⁴³ utilized the Empatica E4 wearable device (Empatica Inc., Cambridge, MA, USA) along with a custom-made application to maintain a diary of headache-specific data. The device employed data-driven ML algorithms to detect activity, stress, and sleep events. Individual headache attacks were classified based on a knowledge-based classification system, focusing on migraine, CH, and TTH. A total of 133 headache attacks from 14 migraine and four CH patients were analyzed. The strict application of ICHD-3 criteria resulted in the classification of eight out of 98 MwoA attacks and 0 out of 35 CH attacks. However, an adapted version of the criteria, which modified the headache duration for treated and terminated episodes, improved classification to 28 out of 98 MwoA attacks and 17 out of 35 CH attacks. The device also collected data on activities and stress events, which were confirmed in 46% and 59% of cases, respectively, indicating the potential link between headache and physiological data, although further improvement is warranted.

Functional near-infrared spectroscopy was employed to measure changes in hemoglobin levels in the prefrontal cortex during a mental arithmetic task, with the data used to classify 13 HCs, nine CM patients, and 12 MOH patients.⁴⁴ ML techniques, including LDA and QDA, were applied in both direct and stepwise classifications. The resulting model achieved a sensitivity of 100% and a specificity of 75% in classifying CM patients.

The statistical application of AI is particularly well-suited for use in classification tasks, especially when applied to data sources such as brain imaging, electrophysiology, wearable devices, or other measurable inputs. These data sources provide numerous inputs, and the diagnosis of headache disorders offers clearly defined target labels, facilitating the use of AI in generating accurate classifications. As demonstrated in the studies presented, these methods often yield favorable accuracies and show significant potential. However, the application of these AI methods in real-world clinical settings remains uncertain. A meta-analysis on the real-world accuracy of wearable activity trackers for detecting COVID-19, AF, and falls reported sensitivities of 79.5%, 94.2%, and 81.9%, and specificities of 76.8%, 95.3%, and 62.5%, respectively.⁴⁵ Notably, the highest accuracy was observed in detecting AF, which

is primarily diagnosed using wavelet-transformed data from ECG signals. In contrast, the gold standard for diagnosing headache disorders is patient interviews, and interpreting headache diagnoses classified by complex wavelet data presents significant challenges. Additionally, randomized controlled studies are limited in demonstrating the benefits of AI or comparing with gold standard methods.¹⁸ Also most studies, except for the ECG study by Chiang et al.⁴⁶, involved a small number of participants, raising concerns about the generalizability of these AI applications to broader populations.

ARTIFICIAL INTELLIGENCE APPLICATIONS IN THE ASSESSMENT OF TREATMENT EFFICACY AND RESPONSE IN HEADACHE DISORDERS

Assessing treatment response is a crucial aspect of clinical practice. Identifying responders and non-responders helps avoid ineffective therapies and minimize adverse effects, which is the core principle of precision medicine. This is particularly important when treating patients with headache disorders, especially CM, using costly therapies such as OnabotulinumtoxinA and anti-calcitonin gene-related peptide monoclonal antibodies (anti-CGRP mAb), where non-responders can have significant implications. AI methods have been increasingly utilized to assess or predict the need for treatment, evaluate treatment response, and identify potential good responders.

1. Questionnaire/survey

Ashina et al.⁴⁷ conducted a web-based survey involving 31,529 out of 61,826 individuals (51.0%) who had sought medical care for migraine in the previous 12 months. Using ML techniques, including RF and LASSO, the study identified 13 sociodemographic and clinical factors most strongly associated with seeking medical care for migraine. Among these, higher interictal burden, disability, and allodynia were particularly significant factors.

2. Natural language

NLP of EHRs, generative LLMs have been utilized to assess treatment response, evaluate current treatment status, and analyze patient feedback.

Hindiye et al.⁴⁸ constructed a migraine outcome model based on headache severity (mild, moderate, severe), headache descriptors (pulsating, debilitating, stabbing), headache progression, and associated symptoms (nausea, vomiting, photophobia, and phonophobia). Each data element was weighted to define a 10-point scale. EHR data from 2018 to 2020 were reviewed, and trained annotators assigned scores. The accuracy of “traditional approaches” and “advanced approaches” was compared. From 2,006 encounters, the average F1 score for automated extraction was 92.0% for AI applied to unstructured data (advanced approach).

Guo et al.⁴⁹ developed a platform-independent text classification system to automatically detect and analyze self-reported migraine-related posts. Texts from Twitter and Reddit were manually labeled, and six transformer-based models were used to classify posts as positive if at least one sentence within the post was identified as self-reporting. The best system achieved an F1 score of 0.9 on Twitter and 0.93 on Reddit, demonstrating minimal bias. Treatment-related information and associated sentiments were also analyzed. This study suggests the potential for analyzing treatment response based on real-time, real-world self-reports, outside of traditional hospital settings or headache diaries, which could reduce recall bias.

Chiang et al.⁵⁰ performed a retrospective cross-sectional study from two tertiary headache referral centers. A total of 1,915 neurology consultation notes written by 15 specialized clinicians between 2012 and 2022 were extracted. Four NLP frameworks were applied to generate answers and extract headache frequency. Among these, the generative pre-trained transformer 2 (GPT-2) generative model showed the best performance, with an accuracy of 0.92 (95% CI, 0.91–0.93) and an R^2 score of 0.89 (95% CI, 0.87–0.90). All GPT-2-based models outperformed the ClinicalBERT (Bidirectional Encoder Representation from Transformers) model in terms of exact matching accuracy.

Li et al.⁵¹ provided 30 migraine-related queries, including evaluation, definition, testing, diagnosis, treatment, follow-up, prognosis, and special population considerations, to five LLMs (ChatGPT-3.5, ChatGPT-4.0, Google Bard, Meta Llama2, and Anthropic Claude2). The answers were randomly ordered and rated by neurologists.⁵¹ Although the difference in performance was not statistically significant, ChatGPT-4.0 received the highest accuracy ratings,

whereas Google Bard had a relatively higher proportion of ‘poor’ ratings. Notably, there were erroneous recommendations, such as proposing hemispherectomy for persistent and severe migraine by ChatGPT-3.5.

This study highlights the need for caution among clinicians, researchers, and potential patients when using LLMs for medical purposes. These erroneous recommendations are not just incorrect; they have the potential to cause patient harm. Therefore, the use of LLMs must be managed with caution and public awareness, and further research is warranted.

Another significant caution regarding the use of LLMs for medical advice arises from a study by Moskatel and Zhang.⁵² They queried ChatGPT-3.5 on the efficacy of 47 medications for the prevention of migraine and evaluated its responses and citations. The assessments of 33 medications were found to be unreliable, with 66% (76/115) of the citations being hallucinations and 5% (6/115) being erroneous.

3. Clinical dataset

Lu et al.⁵³ evaluated 610 migraine patients, including 326 who responded to non-steroidal anti-inflammatory drugs (NSAIDs) and those who did not. They extracted potential predictors among demographic and clinical features using multivariable LR analysis.⁵³ The SVM, DT, and MLP algorithms were used to predict NSAID responsiveness, with the AUC for the test cohort ranging from 0.712 to 0.744 across the three ML methods. Significant predictors identified included disease duration, headache intensity, frequency, anxiety, depression, and sleep disorders.

Martinelli et al.⁵⁴ attempted to predict treatment response to OnabotulinumtoxinA in patients with CM and high-frequency episodic migraine. Among the 212 enrolled patients, 35 were classified as excellent responders and 38 as non-responders. The Relif Family feature selection algorithm was used to select demographic and clinical data, which were then analyzed using various ML methods. Although ML methods failed to distinguish good responders from non-responders overall, the RF algorithm in the high-frequency EM group achieved a high classification accuracy of 85.71%. Key predictors of response in the high-frequency EM group included age at migraine onset, opioid use, anxiety subscore on the Hospital Anxiety

and Depression Scale, and Migraine Disability Assessment (MIDAS) score.

Gonzalez-Martinez et al.⁵⁵'s team utilized prospectively collected multicenter dataset of 712 migraine patients receiving anti-CGRP mAb therapies to predict treatment response. The study population was predominantly female (93%), with 84% having CM. A RF-based approach was employed, with hyperparameters selected using a Bayesian search optimization method. Prediction models at 6, 9, and 12 months utilized variables such as headache days per month at each time point and their reduction, migraine days per month at baseline and 3 months, and headache impact test (HIT-6) scores. The F1 scores of the models ranged from 0.70 to 0.97, with AUROC values between 0.87 and 0.98. A calculator tool was subsequently developed and made available online (<https://portal.brainguard.life/tools/cgrp.php>).

Stubberud et al.⁵⁶ utilized clinical data from a retrospective cohort of 1,446 CM patients to estimate individual treatment effects across 10 classes of preventive therapies, including OnabotulinumtoxinA, flunarizine, candesartan, serotonin-noradrenaline reuptake inhibitors, topiramate, tricyclic antidepressants, acupuncture, valproate, beta blockers, and serotonergic agents. The analysis was performed using a causal multitask Gaussian process model. Data were collected through automated extraction using NLP of Microsoft Word template-based clinical records, achieving an accuracy of 90.73% compared to manual extraction. Individual treatment effects were then used to rank the preventive therapies for machine-guided prescription. The machine prescription policy was estimated to reduce time-to-response by 35% (3.750 months; 95% CI, 3.507–3.993; $p < 0.0001$) compared with expert guidelines, with no substantive increase in expense per patient.

Ferroni et al.⁵⁷'s research utilized a dataset of 777 migraine patients with 21% (162) of whom reported MO lasting for at least 2 years, to predict the risk of developing MO. The team developed a customized ML-based decision support system combining SVM and Random Optimization (RO-MO), which was compared to a baseline SVM model. The final RO-MO decision support system, incorporating the top four models, achieved a c-statistic of 0.83, with sensitivity and specificity of 0.69 and 0.87, respectively, and an accuracy of 0.87. LR analysis confirmed the system's effectiveness in predicting MO, with odds ratios of 5.7 and 21.0

for patients classified as probably (three predictors positive) and definitely at risk of MO (four predictors positive), respectively.

Ciancarelli et al.⁵⁸ used ANN to predict the effect of EMG-biofeedback treatment in 20 CM patients. The ANN predicted post-treatment MIDAS scores with 75% accuracy. A significant correlation between NOx (nitrite and nitrate) levels and MIDAS ($R = -0.675$, $p = 0.011$) suggested that higher nitric oxide levels pre-treatment were associated with lower post-treatment MIDAS scores, particularly when peroxide levels are within a specific range (116–205 U/mL).

4. Imaging

Wei et al.⁵⁹ evaluated 111 migraine patients, of whom 62 were responders to NSAIDs and 49 were non-responders. Their 3D-T1 weighted images were analyzed using DL with the ResNet-18 model demonstrated the best accuracy of 0.78. In a subsequent study, the static functional connectivity was compared among 35 NSAID-responsive episodic MwoA patients, 35 NSAID-non-responsive MwoA patients, and 33 HCs. Clinical characteristics and functional network connectivity features were applied to a SVM model to classify NSAID responsiveness, yielding a sensitivity of 0.88, specificity of 0.89, and an AUROC of 0.93. NSAID-responsive patients exhibited reduced connectivity between the DMN and VN, as well as between the SMN and VN, while showing enhanced VN-auditory network connections.

In a follow-up study, the team compared 59 NSAID responders with 59 non-responders among migraine patients, using propensity score matching.⁶⁰ Multimodal MRI was employed to extract percentage amplitude oscillations and gray matter volume from six brain areas, with multiple ML models applied. The RF model, which had the lowest predictive residuals, was selected. The model metrics in the training and testing groups were as follows: AUROC 0.982/0.711, sensitivity 0.976/0.667, and F1 score 0.930/0.649. The choice of AI algorithm is noteworthy. ResNet-18, a CNN based DL architecture, is advantageous for direct image analysis. When features extracted from MRI were used, ML methods were applied. Marino et al.⁶¹ utilized Compressive Big Data Analytics (CBDA), a semi-supervised ML technique, to identify predictive migraine biomarkers at the molecular level using a PET

dataset from 38 migraine patients and 23 HCs. The CBDA method classified migraineurs from HCs with accuracy, sensitivity, and specificity above 90% for both whole-brain and region-of-interest analyses. The putamen was identified as the most predictive region for migraine, particularly regarding μ -opioid and D2/D3 dopamine receptors.

Tso et al.⁶² predicted verapamil responsiveness in 708 CH and probable CH patients, comprising 317 episodic and 391 chronic cases, using 72 clinical features from 410 patients and imaging data from 194 patients. Non-linear dimensionality reduction techniques, including principal component analysis and t-distributed stochastic neighbor embedding, were applied to the clinical data, identifying two large clusters. KNN was then used to define these clusters. The voxel-based morphometry analysis revealed a gray matter cluster in lobule VI of the cerebellum (-4, -66, -20) that exhibited increased gray matter concentration in verapamil non-responders compared with responders ($p=0.008$). The XGBoost-implemented GB DT was used to predict verapamil response, achieving AUROC of 0.689 on cross-validation (95% CI, 0.651–0.710) and 0.621 on held-out data.

While there are still relatively few studies and the results have not yet been particularly compelling, the potential for utilizing AI in this area has been demonstrated. Further research and development are needed to refine these methods and make them more accessible for clinical application in the future.

ARTIFICIAL INTELLIGENCE APPLICATIONS IN MIGRAINE ATTACK PREDICTION

1. Forecasting migraine attack

Migraine sufferers often have a strong desire to predict both the onset and intensity of a migraine attack. Despite knowing that acute-phase migraine medication should be taken immediately when a headache begins (as reported by 184 out of 207 participants), many delay treatment. This hesitation is largely due to the desire to confirm whether the headache is indeed a migraine (68.7%) and to reserve medication for cases that develop into severe migraine attacks (46.2%).⁶³ The application of AI holds great potential in forecasting migraine attacks, given its strength in classification and prediction.

In a study by Stubberud et al.⁶⁴, 18 migraine patients were prospectively included, completing 388 headache diary entries and self-administering app-based biofeedback sessions that wirelessly measured heart rate, peripheral skin temperature, and muscle tension. The primary outcome was the presence or absence of any headache on the day following a completed headache diary entry and biofeedback session. The RF model was the top-performing model in the out-of-sample test set, achieving an AUROC of 0.62, with accuracy, sensitivity, and specificity of 0.56, 0.0, and 1.0, respectively. A GB classifier showed similar results. Using SHapley Additive exPlanations, the most important features for predicting the next day's headache were identified as premonitory symptoms (craving, swelling, and feeling cold), the amount of sleep, the presence and intensity of headache, the impact of the headache on daily functioning, the length of the biofeedback session, and mean heart rate.

Siirtola et al.⁶⁵ utilized wearable sensors from the wrist-worn Empatica E4 device, along with sleep data, to predict migraine attacks. Data from seven participants, including headache diaries and sleep metrics, were used. The wearable device collected data from a 3D accelerometer, thermometer, electrodermal activity sensor (galvanic skin response), and photoplethysmography sensor (measuring blood volume, heart rate, and heart rate variability). Features were derived by comparing nights before a migraine attack to nights without an attack, and nights before a day without a migraine were also compared with each other. QDA and LDA were used as classifiers, with QDA producing better results than LDA. The personal model outperformed the balanced user-independent model, with accuracy for detecting attacks one night prior exceeding 82% in five individuals, while accuracy varied significantly, ranging from 60.4% to 69.6% in the other two individuals.

Katsuki et al.⁹ utilized a smartphone application to collect hourly headache occurrences from 4,375 migraine sufferers, integrating this data with local weather information. The variables were analyzed using a generalized linear mixed model, feedforward neural network, and XGBoost. The study found that headache occurrences were associated with lower barometric pressure ($p<0.001$, gain=3.9) and significant decreases in barometric pressure ($p<0.001$, gain=11.7), higher barometric pressure at 6 a.m. ($p<0.001$, gain=4.6), higher humidity ($p<0.001$, gain=7.1), and in-

creased rainfall ($p < 0.001$, gain=3.1).

Further research is needed to enhance accuracy, ease of use, and generalizability, but the significant patient demand and industrial potential underscore the importance of this field.

ARTIFICIAL INTELLIGENCE APPLICATION IN RESEARCH OF HEADACHE DISORDERS

1. Basic research

Kogelman et al.⁶⁶ collected temporal multi-omics profiles from 24 migraine patients during spontaneous migraine attacks, 2 hours after triptan treatment, during headache-free periods, and after a cold-pressor test. Relevant metabolites were evaluated using an ML method based on symbolic regression, QLattice.⁶⁶ The study detected lower cortisol levels, higher sumatriptan levels, and elevated glutamine levels following treatment. Changes in sumatriptan levels were correlated with changes in GNA1 and VIPR2 gene expression, both of which are known to regulate cAMP levels.

Chiang et al.⁶⁷ developed a DL model for the mouse grimace scale (MGS) called DeepMGS, utilizing the ResNet-18 architecture. This model automatically crops mouse face images, predicts action unit scores and total scores on the MGS, and infers the presence of pain. The system was tested on six migraine and six control mice, with performance compared to human scorers. The model achieved an accuracy of 70% to 90% and demonstrated a high correlation with human scorers in total MGS score (correlation coefficient=0.83).

Thomas et al.⁶⁸ used a neural network model to replicate the neurophysiological dysfunction observed in migraine sufferers, specifically analyzing cortical-evoked potentials in response to repetitive visual and auditory stimuli. They developed normal and migraine synapse models for comparison. Upon repetitive presentation of stimuli at 40 dB and 70 dB input levels, the migraine model exhibited sensitization, with higher potentiating synapse strength resulting in a greater output.

2. Imaging

Hong et al.⁶⁹ developed a system for the segmentation of

deep white matter hyperintensities (WMHs) using a deep neural network based on the U-Net architecture. The model, applied to 148 migraine patients, comprised two networks: the first identified potential deep WMH candidates, and the second reduced false positives among these candidates. The models achieved a true positive rate of 0.88, a false discovery rate of 0.13, and an F1 score of 0.88 for segmenting deep WMHs.

CONCLUSIONS AND FUTURE PERSPECTIVES

The application of AI in the field of headache disorders is on the rise and has shown promising results. However, significant challenges remain in improving accuracy, generalizability and validation, ease of application, and linking findings to clinical relevance. Further research is needed in areas such as digital twins, which have been suggested as a potential tool in migraine management but have yet to be thoroughly explored.⁷⁰

The appropriate use of AI holds great potential to enhance diagnosis, treatment, and research processes in the headache field. However, it is important to recognize that DL, ML, and various supervised and unsupervised methods do not always produce optimal results. No single approach—whether ML, DL, or supervised/unsupervised methods—is inherently superior to the other. Therefore, selecting the most appropriate method with careful consideration of study design is recommended. Caution is necessary when interpreting results, particularly with generative AI models such as LLMs.

AVAILABILITY OF DATA AND MATERIAL

The data presented in this study are available upon reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: WL, MKC; Data curation: WL, MKC; Formal analysis: WL, MKC; Investigation: WL, MKC; Methodology: WL; Writing—original draft: WL; Writing—review & editing: WL, MKC.

CONFLICT OF INTEREST

Wonwoo Lee was involved as a site investigator in a multicenter trial sponsored by Eli Lilly and Co., WhanIn Pharm Co. Ltd., and Handok-Teva. He has received lecture honoraria from Abbott and SK chemical in the past 24 months. Min Kyung Chu was a site investigator for a multicenter trial sponsored by Allergan Korea, Biohaven Pharmaceuticals, and Lundbeck Korea. He has received lecture honoraria from Allergan Korea, Handok-Teva, Eli Lilly and Company, and Yuyu Pharmaceutical Company in the past 24 months. Additionally, he received grants from Yonsei University College of Medicine (6-2021-0229), the Korea Health Industry Development Institute (KHIDI) (HV22C0106), and National Research Foundation of Korea (2022R1A2C1091767).

FUNDING STATEMENT

Not applicable.

ACKNOWLEDGMENTS

Grammatical error revision was supported by ChatGPT-4o.

REFERENCES

- Gautam R, Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. *J Med Syst* 2020;44:49.
- Woldeamanuel YW, Cowan RP. Computerized migraine diagnostic tools: a systematic review. *Ther Adv Chronic Dis* 2022;13:20406223211065235.
- Hoehndorf R, Queralt-Rosinach N. Data Science and symbolic AI: synergies, challenges and opportunities. *Data Sci* 2017;1:27-38.
- van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *Int J Man Mach Stud* 1978;10:313-322.
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- Liu F, Bao G, Yan M, Lin G. A decision support system for primary headache developed through machine learning. *PeerJ* 2022;10:e12743.
- Simić S, Villar JR, Calvo-Rolle JL, Sekulić SR, Simić SD, Simić D. An application of a hybrid intelligent system for diagnosing primary headaches. *Int J Environ Res Public Health* 2021;18:1890.
- Katsuki M, Shimazu T, Kikui S, et al. Developing an artificial intelligence-based headache diagnostic model and its utility for non-specialists' diagnostic accuracy. *Cephalalgia* 2023;43:3331024231156925.
- Katsuki M, Tatsumoto M, Kimoto K, et al. Investigating the effects of weather on headache occurrence using a smartphone application and artificial intelligence: a retrospective observational cross-sectional study. *Headache* 2023;63:585-600.
- Katsuki M, Matsumori Y, Kawamura S, et al. Developing an artificial intelligence-based diagnostic model of headaches from a dataset of clinic patients' records. *Headache* 2023;63:1097-1108.
- Okada M, Katsuki M, Shimazu T, et al. Preliminary external validation results of the artificial intelligence-based headache diagnostic model: a multicenter prospective observational study. *Life (Basel)* 2024;14:744.
- Sasaki S, Katsuki M, Kawahara J, et al. Developing an artificial intelligence-based pediatric and adolescent migraine diagnostic model. *Cureus* 2023;15:e44415.
- Kwon J, Lee H, Cho S, Chung CS, Lee MJ, Park H. Machine learning-based automated classification of headache disorders using patient-reported questionnaires. *Sci Rep* 2020;10:14062.
- Katsuki M, Narita N, Matsumori Y, et al. Preliminary development of a deep learning-based automated primary headache diagnosis model using Japanese natural language processing of medical questionnaire. *Surg Neurol Int* 2020;11:475.
- Kim KM, Kim AR, Lee W, Jang BH, Heo K, Chu MK. Development and validation of a web-based headache diagnosis questionnaire. *Sci Rep* 2022;12:7032.
- Maizels M, Wolfe WJ. An expert system for headache diagnosis: the Computerized Headache Assessment tool (CHAT). *Headache* 2008;48:72-78.
- Cowan RP, Rapoport AM, Blythe J, et al. Diagnostic accuracy of an artificial intelligence online engine in migraine: a multi-center study. *Headache* 2022;62:870-882.
- Khan B, Fatima H, Qureshi A, et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomed Mater Devices* 2023;1:731-738.
- Riskin D, Cady R, Shroff A, Hindiyeh NA, Smith T, Kymes S. Using artificial intelligence to identify patients with migraine and associated symptoms and conditions within electronic health records. *BMC Med Inform Decis Mak* 2023;23:121.
- Vandenbussche N, Van Hee C, Hoste V, Paemeleire K. Using

- natural language processing to automatically classify written self-reported narratives by patients with migraine or cluster headache. *J Headache Pain* 2022;23:129.
21. Mitrović K, Petrušić I, Radojičić A, Daković M, Savić A. Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data. *Front Neurol* 2023;14:1106612.
 22. Mitrović K, Savić AM, Radojičić A, Daković M, Petrušić I. Machine learning approach for Migraine Aura Complexity Score prediction based on magnetic resonance imaging data. *J Headache Pain* 2023;24:169.
 23. Chong CD, Berisha V, Ross K, Kahn M, Dumkrieger G, Schwedt TJ. Distinguishing persistent post-traumatic headache from migraine: classification based on clinical symptoms and brain structural MRI data. *Cephalalgia* 2021;41:943-955.
 24. Dumkrieger G, Chong CD, Ross K, Berisha V, Schwedt TJ. The value of brain MRI functional connectivity data in a machine learning classifier for distinguishing migraine from persistent post-traumatic headache. *Front Pain Res (Lausanne)* 2023;3:1012831.
 25. Rahman Siddiquee MM, Shah J, Chong C, et al. Headache classification and automatic biomarker extraction from structural MRIs using deep learning. *Brain Commun* 2022;5:fcac311.
 26. Tu Y, Zeng F, Lan L, et al. An fMRI-based neural marker for migraine without aura. *Neurology* 2020;94:e741-e751.
 27. Nie W, Zeng W, Yang J, et al. Extraction and analysis of dynamic functional connectome patterns in migraine sufferers: a resting-state fMRI study. *Comput Math Methods Med* 2021;2021:6614520.
 28. Nie W, Zeng W, Yang J, Zhao L, Shi Y. Classification of migraine using static functional connectivity strength and dynamic functional connectome patterns: a resting-state fMRI study. *Brain Sci* 2023;13:596.
 29. Chong CD, Gaw N, Fu Y, Li J, Wu T, Schwedt TJ. Migraine classification using magnetic resonance imaging resting-state functional connectivity data. *Cephalalgia* 2017;37:828-844.
 30. Fernandes O Jr, Ramos LR, Acchar MC, Sanchez TA. Migraine aura discrimination using machine learning: an fMRI study during ictal and interictal periods. *Med Biol Eng Comput* 2024;62:2545-2556.
 31. Yang H, Zhang J, Liu Q, Wang Y. Multimodal MRI-based classification of migraine: using deep learning convolutional neural network. *Biomed Eng Online* 2018;17:138.
 32. Hsiao FJ, Chen WT, Pan LH, et al. Resting-state magnetoencephalographic oscillatory connectivity to identify patients with chronic migraine using machine learning. *J Headache Pain* 2022;23:130.
 33. Hsiao FJ, Chen WT, Wu YT, et al. Characteristic oscillatory brain networks for predicting patients with chronic migraine. *J Headache Pain* 2023;24:139.
 34. Langer-Gould AM, Anderson WE, Armstrong MJ, et al. The American Academy of Neurology's top five choosing wisely recommendations. *Neurology* 2013;81:1004-1011.
 35. Hsiao FJ, Chen WT, Wang YF, et al. Identification of patients with chronic migraine by using sensory-evoked oscillations from the electroencephalogram classifier. *Cephalalgia* 2023;43:3331024231176074.
 36. Orhanbulucu F, Latifoğlu F, Baydemir R. A new hybrid approach based on time frequency images and deep learning methods for diagnosis of migraine disease and investigation of stimulus effect. *Diagnostics (Basel)* 2023;13:1887.
 37. Frid A, Shor M, Shifrin A, Yarnitsky D, Granovsky Y. A biomarker for discriminating between migraine with and without aura: machine learning on functional connectivity on resting-state EEGs. *Ann Biomed Eng* 2020;48:403-412.
 38. Aslan Z. Migraine detection from EEG signals using tunable Q-factor wavelet transform and ensemble learning techniques. *Phys Eng Sci Med* 2021;44:1201-1212.
 39. Akben SB, Subasi A, Tuncel D. Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network. *J Med Syst* 2012;36:925-931.
 40. Cao Z, Lai KL, Lin CT, Chuang CH, Chou CC, Wang SJ. Exploring resting-state EEG complexity before migraine attacks. *Cephalalgia* 2018;38:1296-1306.
 41. Chiang CC, Chhabra N, Chao CJ, et al. Migraine with aura associates with a higher artificial intelligence: ECG atrial fibrillation prediction model output compared to migraine without aura in both women and men. *Headache* 2022;62:939-951.
 42. Zhu B, Coppola G, Shoran M. Migraine classification using somatosensory evoked potentials. *Cephalalgia* 2019;39:1143-1155.
 43. De Brouwer M, Vandenbussche N, Steenwinckel B, et al. mBrain: towards the continuous follow-up and headache classification of primary headache disorder patients. *BMC Med Inform Decis Mak* 2022;22:87.
 44. Chen WT, Hsieh CY, Liu YH, Cheong PL, Wang YM, Sun CW. Migraine classification by machine learning with functional near-infrared spectroscopy during the mental arithmetic task. *Sci Rep* 2022;12:14590.
 45. Singh B, Chastin S, Miatke A, et al. Real-world accuracy of wearable activity trackers for detecting medical conditions:

- systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2024;12:e56972.
46. Chiang CC, Schwedt TJ, Dodick DW. Exploring the association between migraine and atrial fibrillation utilizing a novel artificial intelligence-ECG algorithm. *Headache* 2022;62:933-934.
 47. Ashina S, Muenzel EJ, Nicholson RA, et al. Machine learning identifies factors most associated with seeking medical care for migraine: results of the OVERCOME (US) study. *Headache* 2024;64:1027-1039.
 48. Hindiyeh NA, Riskin D, Alexander K, Cady R, Kymes S. Development and validation of a novel model for characterizing migraine outcomes within real-world data. *J Headache Pain* 2022;23:124.
 49. Guo Y, Rajwal S, Lakamana S, et al. Generalizable natural language processing framework for migraine reporting from social media. *AMIA Jt Summits Transl Sci Proc* 2023;2023:261-270.
 50. Chiang CC, Luo M, Dumkrieger G, et al. A large language model-based generative natural language processing framework fine-tuned on clinical notes accurately extracts headache frequency from electronic health records. *Headache* 2024;64:400-409.
 51. Li L, Li P, Wang K, Zhang L, Ji H, Zhao H. Benchmarking state-of-the-art large language models for migraine patient education: performance comparison of responses to common queries. *J Med Internet Res* 2024;26:e55927.
 52. Moskatel LS, Zhang N. The utility of ChatGPT in the assessment of literature on the prevention of migraine: an observational, qualitative study. *Front Neurol* 2023;14:1225223.
 53. Lu ZX, Dong BQ, Wei HL, Chen L. Prediction and associated factors of non-steroidal anti-inflammatory drugs efficacy in migraine treatment. *Front Pharmacol* 2022;13:1002080.
 54. Martinelli D, Pocora MM, De Icco R, et al. Searching for the predictors of response to BoNT-A in migraine using machine learning approaches. *Toxins (Basel)* 2023;15:364.
 55. Gonzalez-Martinez A, Pagán J, Sanz-García A, et al. Machine-learning-based approach for predicting response to anti-calcitonin gene-related peptide (CGRP) receptor or ligand antibody treatment in patients with migraine: a multicenter Spanish study. *Eur J Neurol* 2022;29:3102-3111.
 56. Stubberud A, Gray R, Tronvik E, Matharu M, Nachev P. Machine prescription for chronic migraine. *Brain Commun* 2022;4:fcac059.
 57. Ferroni P, Zanzotto FM, Scarpato N, et al. Machine learning approach to predict medication overuse in migraine patients. *Comput Struct Biotechnol J* 2020;18:1487-1496.
 58. Ciancarelli I, Morone G, Tozzi Ciancarelli MG, et al. Identification of determinants of biofeedback treatment's efficacy in treating migraine and oxidative stress by ARIANNA (ARTificial Intelligent Assistant for Neural Network Analysis). *Healthcare (Basel)* 2022;10:941.
 59. Wei HL, Wei C, Feng Y, et al. Predicting the efficacy of non-steroidal anti-inflammatory drugs in migraine using deep learning and three-dimensional T1-weighted images. *iScience* 2023;26:108107.
 60. Wei HL, Yu YS, Wang MY, et al. Exploring potential neuroimaging biomarkers for the response to non-steroidal anti-inflammatory drugs in episodic migraine. *J Headache Pain* 2024;25:104.
 61. Marino S, Jassar H, Kim DJ, et al. Classifying migraine using PET compressive big data analytics of brain's μ -opioid and D2/D3 dopamine neurotransmission. *Front Pharmacol* 2023;14:1173596.
 62. Tso AR, Brudfors M, Danno D, et al. Machine phenotyping of cluster headache and its response to verapamil. *Brain* 2021;144:655-664.
 63. Baron EP, Markowitz SY, Lettich A, et al. Triptan education and improving knowledge for optimal migraine treatment: an observational study. *Headache* 2014;54:686-697.
 64. Stubberud A, Ingvaldsen SH, Brenner E, et al. Forecasting migraine with machine learning based on mobile phone diary and wearable data. *Cephalalgia* 2023;43:3331024231169244.
 65. Siirtola P, Koskimäki H, Mönttinen H, Rönning J. Using sleep time data from wearable sensors for early detection of migraine attacks. *Sensors (Basel)* 2018;18:1374.
 66. Kogelman LJA, Falkenberg K, Ottosson F, et al. Multi-omic analyses of triptan-treated migraine attacks gives insight into molecular mechanisms. *Sci Rep* 2023;13:12395.
 67. Chiang CY, Chen YP, Tzeng HR, Chang MH, Chiou LC, Pei YC. Deep learning-based grimace scoring is comparable to human scoring in a mouse migraine model. *J Pers Med* 2022;12:851.
 68. Thomas E, Sándor PS, Ambrosini A, Schoenen J. A neural network model of sensitization of evoked cortical responses in migraine. *Cephalalgia* 2002;22:48-53.
 69. Hong J, Park BY, Lee MJ, Chung CS, Cha J, Park H. Two-step deep neural network for segmentation of deep white matter hyperintensities in migraineurs. *Comput Methods Programs Biomed* 2020;183:105065.
 70. Gazerani P. Intelligent digital twins for personalized migraine care. *J Pers Med* 2023;13:1255.