The Journal of Headache and Pain

RESEARCH

Open Access

Artificial neural networks applied to somatosensory evoked potentials for migraine classification

Gabriele Sebastianelli[1*†], Daniele Secci[2†], Francesco Casillo[1], Chiara Abagnale[1], Cherubino Di Lorenzo[1], Mariano Serrao[1], Shuu-Jiun Wang[3], Fu-Jung Hsiao[3,4,5] and Gianluca Coppola[1]

Abstract

Background Finding a biomarker to diagnose migraine remains a significant challenge in the headache field. Migraine patients exhibit dynamic and recurrent alterations in the brainstem-thalamo-cortical loop, including reduced thalamocortical activity and abnormal habituation during the interictal phase. Although these insights into migraine pathophysiology have been valuable, they are not currently used in clinical practice. This study aims to evaluate the potential of Artificial Neural Networks (ANNs) in distinguishing migraine patients from healthy individuals using neurophysiological recordings.

Methods We recorded Somatosensory Evoked Potentials (SSEPs) to gather electrophysiological data from low- and high-frequency signal bands in 177 participants, comprising 91 migraine patients (MO) during their interictal period and 86 healthy volunteers (HV). Eleven neurophysiological variables were analyzed, and Principal Component Analysis (PCA) and Forward Feature Selection (FFS) techniques were independently employed to identify relevant variables, refine the feature space, and enhance model interpretability. The ANNs were then trained independently with the features derived from the PCA and FFS to delineate the relationship between electrophysiological inputs and the diagnostic outcome.

Results Both models demonstrated robust performance, achieving over 68% in all the performance metrics (accuracy, sensitivity, specificity, and F1 scores). The classification model trained with FFS-derived features performed better than the model trained with PCA results in distinguishing patients with MO from HV. The model trained with FFS-derived features achieved a median accuracy of 72.8% and an area under the curve (AUC) of 0.79, while the model trained with PCA results showed a median accuracy of 68.9% and an AUC of 0.75.

Conclusion Our findings suggest that ANNs trained with SSEP-derived variables hold promise as a noninvasive tool for migraine classification, offering potential for clinical application and deeper insights into migraine diagnostics.

Keywords Artificial intelligence, Neurophysiology, Evoked potentials, Habituation, Sensitization, Thalamus

†Gabriele Sebastianelli and Daniele Secci contributed equally to the paper.

*Correspondence:
Gabriele Sebastianelli
gabriele.sebastianelli@uniroma1.it

Full list of author information is available at the end of the article

Background

Migraine is highly prevalent worldwide [1], representing one of the leading causes of disability among individuals younger than 50 years old and the first cause in young women [2]. Despite the great interest in finding a biomarker, this remains a challenge in the headache field [3], and migraine diagnosis still relies on clinical criteria [4]. However, clinical criteria do not fully capture migraine heterogeneity, and advanced diagnostic methods based on migraine neurobiology and neurophysiology are necessary to offer deeper phenotyping and personalized treatments [3]. Electrophysiological techniques could be valuable for this purpose, as they allow noninvasive study of neuronal excitability and have proven their utility in understanding migraine pathophysiology. Several studies have provided evidence that individuals with migraine exhibit dynamic and recurrent alterations in the brainstem-thalamo-cortical loop, which varies cyclically and recurrently during the migraine cycle [5]. The most consistent alterations were observed during the interictal phase, with almost all the non-painful evoked potentials. They included dysfunctions in habituation mechanisms from repetitive stimulations and reduced thalamocortical activity, both reverted during the ictal phase when sensitization develops [6]. However, these alterations were often reported as differences at the group level and are not currently used in clinical practice [7].

In the last years, machine learning (ML) algorithms have rapidly evolved leading to breakthroughs in various fields, including headache. The application of ML algorithms to neurophysiological data analysis has yielded promising results, achieving good accuracy in classifying individuals with episodic migraine [8] and chronic migraine [9, 10]. Considering the abnormal processing of sensory stimuli in migraine [11], examining the somatosensory system through evoked potentials, combined with ML techniques, could be an effective approach for classifying individuals with this condition [9, 12, 13].

This study aims to evaluate the potential of Artificial Neural Networks (ANNs) in distinguishing interictal episodic migraine patients from healthy individuals using neurophysiological data obtained from somatosensory evoked potentials (SSEPs) recordings. Applying ANNs to neurophysiology recordings can enhance the understanding of migraine neurophysiology and promote the development of a generalizable tool for classifying individuals with migraine in routine clinical practice.

Materials and methods

Participants

Adults (aged ≥ 18 years and ≤ 65 years) who received a diagnosis of episodic migraine without aura (MO) were recruited from the Headache clinic of the Sapienza University of Rome (Polo Pontino, ICOT, Latina). MO was diagnosed according to the International Classification of Headache Disorders third edition (ICHD-III) [4]. All patients received the headache diary via email at least one month before their screening visit. Key exclusion criteria were more than 15 headache days per month, medication overuse (defined as for ICHD-III), coexistence of other types of primary or secondary headaches, comorbidity with any other neurological or psychiatric disorder, prophylactic migraine medications in the previous three months or regular use of other medications (except for contraceptive pills). Healthy volunteers (HV) with no personal or familial history (first- and second-degree relatives) of migraine, no regular medication intake except for the contraceptive pill, and no other overt medical condition were recruited from medical school students and healthcare professionals at Sapienza University of Rome (Polo Pontino). The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics review board. After receiving a comprehensive study description, all participants provided written informed consent.

SSEPs recordings and analysis

All the patients were recorded during their interictal period (at least 3 days before and after a migraine attack). To reduce variability from hormonal effects on cortical excitability, we managed to schedule recording sessions of female subjects outside their premenstrual or menstrual phases. Participants were instructed to sit relaxed and with their eyes open in a comfortable chair and focus on the induced movement of their thumb. All recordings were conducted in the same laboratory during the afternoon (between 2 p.m. and 6 p.m.). SSEPs were elicited by stimulating the right median nerve at the wrist with constant-current square wave pulses (0.2 ms width, cathode positioned proximally) and a repetition rate of 4.4 Hz. The intensity was set at 1.2 times the motor threshold. The SSEP signals were recorded over the contralateral parietal area (C3', 2 cm posterior to C3 in the International 10–20 system) and referenced to Fz, with the ground electrode positioned on the left wrist. The CED™ power 1401 device (Cambridge Electronic Design Ltd, Cambridge, UK) was used to record the signals, while amplification was performed with the Digitimer™ (Digitimer Ltd, UK) (band-pass 0.05–2500 Hz, Gain 1000).

Low-frequency SSEPs (LF-SSEPs)

Six hundred consecutive sweeps of 50 ms sampled at 5000 Hz were recorded. The Signal™ software package version 4.10 (CED Ltd) was used to analyze all the recordings offline. Artifacts were identified using the Signal™ artifact rejection tool. Any signal amplitude exceeding 90% of the analog-to-digital converter (ADC) range was excluded, and all the rejections were controlled by

visual inspection. Additionally, all the recorded signals were visually inspected, and artifacts, blink, and eye movements were manually rejected. To avoid affecting the habituation calculation, recordings with more than 5 artifacts per 100 sweeps were discarded. However, none of the recordings exceeded this threshold. After removing the artifacts, we obtained at least 95 artifact-free sweeps for every 100 responses, resulting in a total of at least 570 artifact-free evoked responses for each subject. Low-frequency responses (LF-SSEPs) were obtained by digitally filtering the signals between 0 and 450 Hz.

All the artifact-free evoked responses per subject were averaged, and latency and amplitude of the N20, P25, and N33 components were calculated for each subject. The initial 300 traces obtained were also divided into three blocks of 100 sweeps each, and the peak-to-peak amplitude of each block of responses was measured (1st, 2nd, and 3rd block). The habituation was calculated as in our previous works by computing the linear regression between the N20-P25 amplitudes of the first and second blocks (Slope 1–2) and between the first and third blocks (Slope 1–3) [14, 15].

High-frequency oscillations (HFOs)

The extraction of high-frequency oscillations was conducted on the parietal N20 LF-SSEP component using a digital zero-phase shift band-pass filtering between 450 and 750 Hz (Barlett-Hanning window, 51 filter coefficients). Two distinct bursts of HFO were identified based on their decreasing amplitude and frequency: the early pre-synaptic burst (pre-HFO), which occurred during the ascending slope of the conventional N20 component, and the late post-synaptic (post-HFO), which was observed in the descending slope of N20 component and sometimes extended into the ascending slope of the N33 peak. When it was not possible to distinguish between these two components, the burst that appeared before the N20 peak was considered pre-HFO, while the burst occurring after this peak was classified as post-HFO. The stimulus artifact was removed from all traces, and the latency of the negative oscillatory maximum (pre-HFO lat, post-HFO lat) and the maximum peak-to-peak amplitude (pre-HFO amp, post-HFO amp) for both bursts were measured.

Statistical analysis

Jamovy software (version 2.5.2) was employed for all the classical statistical analyses. Shapiro-Wilk test was used to assess the normal distribution of data. Neurophysiological data were compared between the two groups using an independent sample t-test for normally distributed data and a Mann-Whitney U test for non-normally distributed data. The Chi-square (χ^2) test was used to investigate homogeneity between categorical variables. A p-value < 0.05 was considered statistically significant.

Artificial neural network

From the SSEP analysis, we selected the following 11 neurophysiological features for the model: amplitudes of grand-average N20-P25 and P25-N33, amplitudes of the 1st block, 2nd block, and 3rd block of 100 responses, Slope 1–2, Slope 1–3, pre-HFO latency, pre-HFO amplitude, post-HFO latency, and post-HFO amplitude. With this initial set of 11 features, reducing the dimensionality of the data is crucial before the ANNs development. This process allows the retention of the most critical information, simplifies the model, and reduces the risk of overfitting and the computational burden on the subsequent machine learning algorithms [16]. Two techniques were independently applied to reduce the dimensionality and select the relevant features from our data set: Principal Component Analysis (PCA) and Forward Feature Selection (FFS).

Before applying both techniques, all data were standardized to reduce PCA and FFS sensitivity to the magnitude of the features. The standardization of the features ensures that each feature contributes equally to the model and prevents features with larger magnitudes from dominating the learning process (Eq. 1). This ensures that all features contribute equally to the analysis:

$$X_{standardized} = \frac{X - \mu}{\sigma} \quad (1)$$

where X is the original feature value, μ is the mean, and σ is the standard deviation.

Furthermore, this process helps with ANN's training. Indeed, standardizing the features helps when using activation functions that output values between −1 and 1, such as the tanh function (e.g., hyperbolic tangent activation). These activation functions work best when the input values are centered around zero and within a certain range [17].

Principal component analysis (PCA)

PCA is a widely used statistical technique that transforms a large set of variables into a smaller one that is a linear combination of the original variables. PCA achieves this by identifying the directions (principal components) in which the variance of the data is maximal. Furthermore, these principal components are uncorrelated with each other. This orthogonality property makes the components independent and simplifies the relationships between variables. To achieve this, the PCA involves the computation of eigenvalues and eigenvectors from the covariance matrix of the original variables. The eigenvalues indicate the amount of variance captured by each principal component, while the eigenvectors represent the directions of the principal components in the feature space. Then, it is important to note that PCA does not select

individual features; instead, it reduces the dimensionality of the dataset by defining linear combinations of the original variables. In particular, the method transforms the original features into a new coordinate system where the greatest variance by any projection of the data comes to lie on the first coordinate (principal component), the second greatest variance on the second coordinate, and so on. To select the number of principal components, a common approach is to retain enough components to explain a significant amount of the total variance. For readers interested in further exploring the topic of PCA, they can refer to the work of Joliffe et al. [18].

In this study, the selection of principal components was guided by the goal of maintaining the balance between dimensionality reduction and preserving sufficient variance (95% of the data) to accurately classify whether a patient has migraine or not. The selected principal components were then used as input features to train the ANN. Moreover, to reinforce our decision to use only the selected principal components as inputs for the neural network, we trained the network with different configurations of principal components. Specifically, we began by using only the first principal component, then added the second component, followed by the third, and so on. For each configuration, we evaluated the accuracy of the network in classifying between MO and HV for each configuration. To account for the inherent randomness in the training process, we trained and assessed a total of 100 neural networks per configuration. The goal was to achieve maximum accuracy while using the fewest possible principal components, ensuring that they still explain 95% of the data variance.

Forward feature selection (FFS)

The FFS is a systematic approach for identifying the most relevant features for a classification task. The procedure employed to obtain results using the FFS methodology can be conceptually compared to that of PCA. However, in contrast to PCA, which transforms the original feature values into a new set of uncorrelated components, the FFS methodology directly assists in identifying the most relevant input features for the neural network without applying any transformations to the original feature values. In particular, FFS helps create a simpler, more interpretable, and more effective model, by iteratively adding features that improve model performance. The selected features provide key insights into the variables that most significantly contribute to the classification, enhancing both the model's performance and its practical utility. The process began with an empty model, iteratively adding one feature at a time based on the improvement in a predefined performance metric. In this study, the ANN was used as the model to perform FFS, and the accuracy was used as the predefined performance metric. At each step, the feature that provided the best improvement was selected, and the process was repeated until all features were considered. At the end of this iterative addition, the optimal feature set was identified. To ensure robustness, we repeated the FFS technique a hundred times to evaluate the frequency with which each feature was selected in the optimal feature set. The features that were most frequently selected were then chosen as the final features for the ANN (frequency > 50%). This repeated evaluation helped in identifying the most consistently important features, ensuring that the final model is both reliable and effective in distinguishing between MO and HV. For further reading on FFS, see the referenced study [19].

Artificial neural network design

The ANNs employed in this study is a multilayer perceptron neural network designed to classify MO and HV based on neurophysiological data derived from SSEP recordings. The construction of the ANN model involved the following steps: development of the architecture (input and output layers, hidden layer, and activation functions), training, tuning the hyperparameter, and performance evaluation. The architecture of the ANN consisted of three main components: input, hidden, and output layers. The input layer received the transformed or selected features from the PCA or FFS processes, respectively (Fig. 1). The hidden layer comprised one layer of 50 neurons, which is critical for capturing complex patterns and relationships in the data. The hyperbolic tangent activation function was applied to introduce non-linearity, enhancing the model's ability to learn from the data. The output layer consisted of two neurons with a softmax activation function, producing a probability score that indicates the likelihood of subjects being MO or HV. The training was performed using either the transformed features from PCA or the selected features from FFS. The training process involved optimizing the weights and biases of the network to minimize the error between predicted and actual classifications. This was achieved using the scaled conjugate gradient backpropagation algorithm, which is effective for training large networks and reducing computational complexity. The dataset was then divided into training (65%), validation (20%) and test (15%) sets [16]. The training set was used to learn the model parameters, the validation set was used to monitor performance and prevent overfitting, and the test set was used to evaluate the final model's generalization ability. The cross-entropy performance metric quantified the difference between predicted probabilities and actual binary labels, guiding the optimization process.

Furthermore, hyperparameter tuning was employed to achieve the optimal model performance. Key hyperparameters, such as the number of hidden layers, the number of neurons per layer, learning rate (0.01), batch

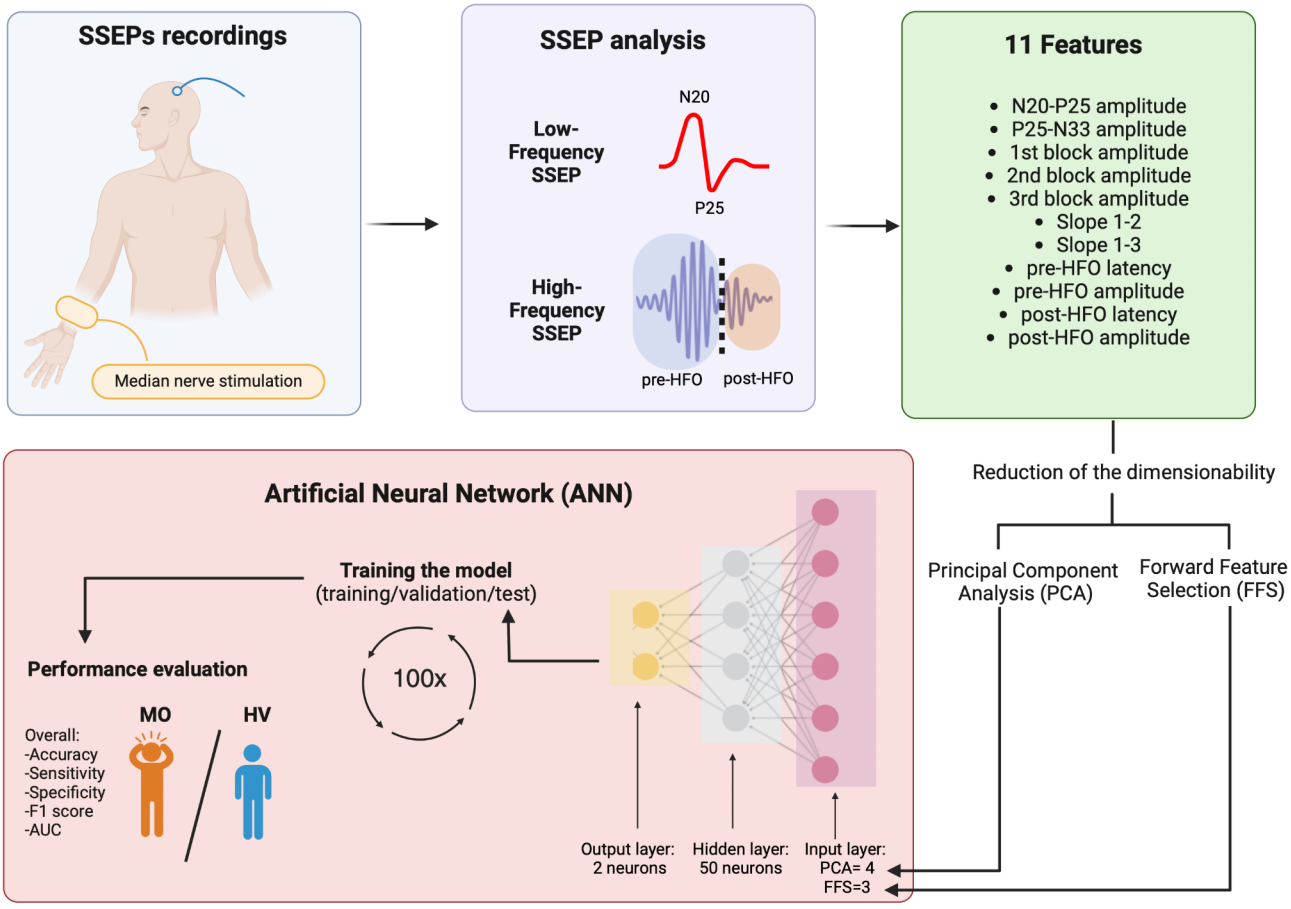

Fig. 1 Pipeline of the Artificial Neural Network (AANs) development. Somatosensory evoked potentials (SSEPs) were recorded by stimulating the median nerve at the wrist in both healthy volunteers (HV) and interictal episodic migraine patients (MO). The recordings were analyzed offline to extract low-frequency responses (LF-SSEPs) and high-frequency oscillations (HFO) from the cortical components of the somatosensory evoked potentials. Eleven features were extracted from the analysis, and two techniques were independently applied to reduce the dimensionality and select the relevant features: Principal Component Analysis (PCA) and Forward Feature Selection (FFS). PCA selected four relevant linear combinations of the features, while FFS selected three relevant features. Two different neural network models were trained with these features transferred to the input layer. The hidden layer comprised 50 neurons, while the output layer comprised 2 neurons. Both models were trained one hundred times by randomly dividing our dataset into training, validation, and test sets for each run. Finally, the performance of both models in classifying HV from MO was evaluated by calculating the median accuracy, the sensitivity (recall), the specificity, and the F1 score. The outcomes were derived by averaging the outputs from 100 neural networks trained on the same dataset. Created in BioRender. Sebastianelli, G. (2025) https://BioRender.com/l67q659

size (18), and the number of epochs (50), were systematically explored using a manual approach. This exhaustive search helped to identify the best combination of hyperparameters that minimize the validation loss and maximize the model's accuracy. Early stopping was also employed to terminate training when the validation loss ceased to improve, thus preventing overfitting. For readers interested in the details of mathematical formulations that explain how ANNs work, they can refer to the work published by Secci et al. [20].

Finally, the outcomes were derived by averaging the outputs from 100 neural networks trained on the same dataset. For each run, the dataset was randomly divided into training, validation, and test sets, introducing variability into the model training process. Additionally, the random initialization of weights further contributed to the variation in each network's performance. Consequently, the neural networks yielded slightly different results. We used more than one network to capture the uncertainty in the results, providing a comprehensive evaluation of the model's robustness. This approach ensures that the results are both reliable and indicative of the model's true performance in distinguishing between MO and HV. To ensure a robust and stable assessment, the ANN's performance was evaluated using key metrics, reporting the median values (derived from the 100 runs) instead of the mean to minimize the impact of outliers and skewed distributions. The metrics used for this evaluation include median accuracy, sensitivity (recall), specificity, and F1 score, calculated as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN}) \;/\; (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (2)$$

$$\text{Sensitivity (Recall)} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (4)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (5)$$

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

where TP (True Positives) are the MO patients correctly identified as MO, TN (True Negatives) are HV correctly identified as HV, FP (False Positives) are HV incorrectly identified as individuals with MO, and FN (False Negatives) are MO patients incorrectly identified as HV. Collectively, these metrics provide a robust evaluation of the ANN's performance in distinguishing between MO and HV. In particular, the median accuracy was reported separately for the training, validation, and test datasets, with confusion matrices used to detail performance within each subset. Additionally, summary tables of all key metrics and the ROC curve illustrated the network's overall performance by evaluating it on the complete dataset (i.e., the combined training, validation, and test set input into the trained neural network). This comprehensive approach provides insights into the model's effectiveness across the entire dataset as well as the subsets, offering a thorough assessment of classification accuracy and reliability.

All data analyses in this study, including PCA, FFS, and ANNs development and evaluation, were conducted using MATLAB (MathWorks Inc., 2024) as a selected environment for code implementation. It was chosen for its extensive suite of toolboxes tailored for data analysis and neural network training.

Table 1 Demographics, clinical and electrophysiological features of MO and HV (mean ± standard deviation)

	MO	HV	*p*-value	Effect size
Participants (n)	91	86		
Female (n)	66 (73%)	44 (51%)	0.003$^{\chi^2}$	
Age	32.94 ± 10.80	31.65 ± 10.89	0.185^{U}	0.12
MHD	3.74 ± 2.99			
SSEP MT (mA)	8.50 ± 2.59	8.96 ± 3.17	0.414^{t}	0.16
N20-P25 (μV)	2.00 ± 1.13	1.93 ± 0.84	0.786^{U}	0.02
P25-N33 (μV)	1.19 ± 0.68	1.00 ± 0.62	**0.035**^{U}	0.18
1st block (μV)	1.92 ± 1.10	2.27 ± 1.00	**0.007**^{U}	0.23
2nd block (μV)	2.19 ± 1.12	2.00 ± 0.85	0.506^{U}	0.06
3rd block (μV)	2.13 ± 1.17	2.06 ± 0.82	0.759^{U}	0.03
Slope 1–2 (μV)	0.26 ± 0.66	-0.27 ± 0.63	**< 0.001**^{U}	0.49
Slope 1–3 (μV)	0.10 ± 0.29	-0.11 ± 0.36	**< 0.001**^{U}	0.34
Pre-HFO lat (ms)	16.26 ± 1.26	16.50 ± 1.84	0.893^{U}	0.01
Post-HFO lat (ms)	23.28 ± 2.72	23.49 ± 2.64	0.827^{U}	0.02
Pre-HFO amp (μV)	0.05 ± 0.03	0.06 ± 0.03	**0.021**^{U}	0.20
Post-HFO amp (μV)	0.05 ± 0.04	0.05 ± 0.03	0.966^{U}	0.00

HV = healthy volunteers; mA = milliampere; MHD = monthly headache days; MO = episodic migraine; ms = millisecond; MT = motor threshold; n = number, SD = standard deviation; SSEP = somatosensory evoked potentials; t = independent sample t-test; U = Mann-Whitney U-test; μV = microvolt, χ^2 = Chi-square test

Results

LF-SSEP and HFOs

This study included 177 participants, including 91 MO and 86 HV. Table 1 provides the demographic, clinical, and electrophysiological characteristics of the participants. The two groups did not differ significantly in terms of age (MO = 32.94 ± 10.80; HV = 31.65 ± 10.89; $p = 0.185$) and motor threshold of the stimulation (HV = 8.96 ± 3.17; MO = 8.50 ± 2.59; $p = 0.414$). However, the MO group had a higher proportion of female participants than the HV (MO = 73% female vs. HV = 51% female, p-value = 0.003). In comparison with HV, individuals with MO showed increased peak-to-peak amplitude for grand average P25-N33 (U = 3195.5, $p = 0.035$), while reduced amplitude for the first block of N20-P25 averaged responses (1st block: U = 2995.00, $p = 0.007$). In contrast to HV, subjects with MO exhibited a progressive increase in the amplitude slope between blocks 1–2 (Slope 1–2: MO = 0.26 ± 0.66; HV= -0.27 ± 0.63; U = 1989.00; p = < 0.001) and blocks 1–3 (Slope 1–3: MO = 0.10 ± 0.29; HV= -0.11 ± 0.36; U = 2573.00; p = < 0.001), indicating deficient habituation. While no differences emerged between the latency of both pre-and post-HFO and the amplitude of post-HFO, individuals with MO exhibited a lower value of the peak-to-peak amplitude of pre-HFO (pre-HFO amplitude: MO = 0.05 ± 0.03; HV = 0.06 ± 0.03; U = 3123.50, $p = 0.021$). No other significant differences emerged between the two groups for the remaining neurophysiological variables (Table 1).

Performance of the model trained with PCA results

Figure 2a illustrates the variance explained by each principal component, showing that the first four components were required to capture the 95% of the total variance. Additionally, as depicted in Fig. 2b, these four components represented the lowest number needed to achieve the highest accuracy. This was also confirmed by the evidence that the inclusion of additional components beyond the fourth introduced a minimal modification of the accuracy (Fig. 2b). The result of Fig. 2b applies specifically to the test dataset, ensuring that the evaluation reflects the model's performance on data that was never seen during training. The resulting median accuracy across the one hundred ANNs runs was approximately 65%, with a relatively narrow uncertainty interval. This indicates that, despite the variability across different neural network training runs, the model consistently achieved a moderate level of accuracy when evaluated on unseen data, reflecting the stability and robustness of the selected principal components for classification.

a)

Principal component	Variance explained (%)
1	51.72
2	26.25
3	11.95
4	5.06
5	2.27
6	1.64
7	1.10
8	0.01
9	0.00
10	0.00
11	0.00

b)

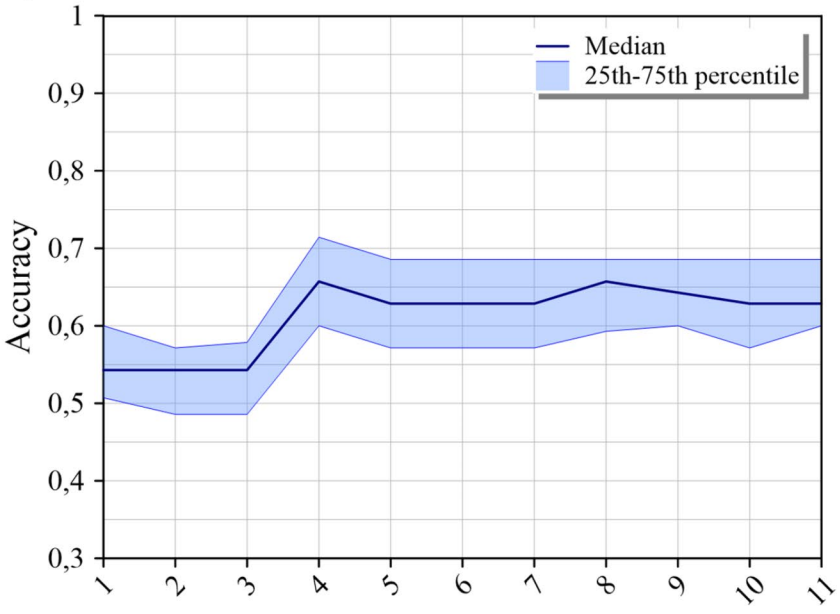

Fig. 2 **a**) Explained variance (%) by each component. The sum of the variance of the first four components accounts for almost 95% of the total variance. **b**) Number of principal components vs. accuracy. The value of the accuracy is based on the test dataset in order to evaluate the network's robustness in terms of generalization capability

Figure 3 presents the confusion matrices for the results obtained by the ANN configured using only the first four principal components. The matrices provide a comprehensive assessment of the model's classification ability, showing the distribution of correct and incorrect classifications within each dataset (training, validation, and test). Moreover, we analyzed the ROC curve to evaluate the overall performance (complete dataset) of the ANN classifier and obtained a final AUC value of 0.75 in identifying individuals with MO (Fig. 3).

Finally, Table 2 summarizes the key performance metrics for the ANN trained with the four features derived from the PCA, specifically referring to the overall performance evaluated on the complete dataset. The classification model based on PCA-selected features achieved an overall sensitivity of 71.5% and a specificity of 68% in identifying subjects with MO, with an F1 score of 69.7%. The median sensitivity in identifying HV was 68%, indicating that the model performs similarly in identifying HV and MO, with a slight advantage for MO. The standard deviation of the sensitivity was relatively low (9.2% for HV and 7% for MO), which indicates a consistent recall across different runs. The model demonstrated reasonably good performance, achieving a median overall accuracy of 68.9% in distinguishing MO from HV. The low variance of the accuracy (5.1%) indicates stable performance across different runs.

Performance of the model trained with FFS results

Figure 4 presents a bar plot illustrating how often each feature was selected by the FFS in the "best dataset"—the dataset that yielded the highest accuracy for each trained ANN. Given the inherent randomness of the training process, the features included in the optimal configuration identified by FFS can vary. This analysis provides insights into the most consistently selected features, highlighting their relevance in achieving high model performance.

Three parameters (Slope 1–2, P25-N33, and pre-HFO amp) showed a 50% or greater frequency to be included in the "best dataset" and were selected as input for training the ANN classification model (Fig. 4). Then, after defining these inputs, we trained and evaluated an additional one hundred neural networks, similar to the approach with PCA, to assess the robustness of the ANN's generalization capabilities. Figure 5 shows the confusion matrices detailing the classifier's accuracy across training, validation, and test datasets by illustrating correct and incorrect classifications. Furthermore, in Fig. 5, the ROC curve is presented with a median AUC value of 0.79 for the class "MO".

The overall key metrics of the classification model based on FFS-selected features are reported in Table 3. The overall sensitivity was 71.5%, with a specificity of 73.8% in identifying subjects with MO and an F1 score of 72.7%. The median sensitivity in identifying HV was 73.8%, indicating that the model performs similarly in identifying HV and MO, with a slight advantage for HV.

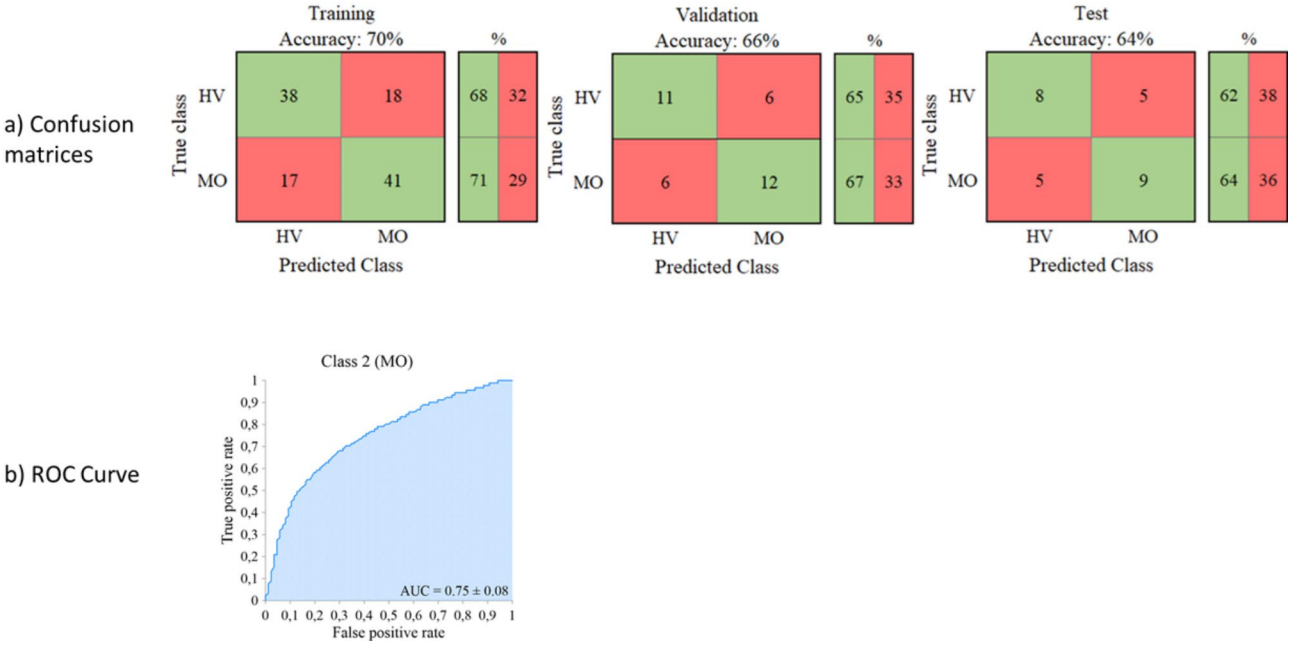

Fig. 3 **a**) Confusion matrices for a detailed summary of the ANN's performance trained with the transformed features from Principal Component Analysis (PCA). The matrix is structured such that the rows represent the true classes (HV or MO), and the columns represent the predicted classes. The diagonal elements (in green) reflect the number of correctly classified cases (true negatives and true positives: TN - TP), while the off-diagonal elements (in red) capture the number of misclassifications (false negatives and false positives: FN - FP). A row summary displays the percentages of correctly and incorrectly classified observations for each true class. For example, considering the training dataset, in the row corresponding to the "HV" class, the percentage of HV correctly classified is shown along with the percentage of HV misclassified as having MO (TN = 68%; FP = 32%). Similarly, for the "MO" class, the row summary indicates the percentage of individuals correctly identified as having MO (TP = 71%) and those misclassified as HV (FN = 29%). **b**) Area under the curve (AUC) for the overall performance of the ANN trained with the transformed features from PCA. The ROC curve is a graphical representation that illustrates the performance of a binary classifier across different threshold settings. The ROC curve plots the True Positive Rate (TPR) on the y-axis against the False Positive Rate (FPR) on the x-axis for all possible threshold values. As the threshold varies, the trade-off between correctly classifying positive cases and misclassifying negative cases as positive is visualized. An ideal classifier would achieve a point near the top-left corner of the ROC plot, where the TPR is high, and the FPR is low. Furthermore, the AUC provides a single metric that summarizes the classifier's overall performance. An AUC of 1 represents perfect classification, where the classifier achieves a high TPR with a low FPR across all thresholds. An AUC of 0.5, however, indicates performance no better than random chance

Table 2 Overall performance of the ANN trained with the transformed features from principal component analysis (PCA) in terms of the median and standard deviation (SD) metrics obtained by running 100 neural networks, using the first four principal components as input and dividing the dataset into training (65%), validation (20%) and test (15%) sets

	MO	HV
Median Accuracy ± SD	0.689 ± 0.05	
Median Sensitivity (Recall) ± SD	0.715 ± 0.070	0.680 ± 0.092
Median Specificity ± SD	0.680 ± 0.092	0.715 ± 0.070
Median F1 Score ± SD	0.697 ± 0.052	0.682 ± 0.061

The standard deviation of the sensitivity was relatively low (8.7% for HV and 5.2% for MO), which indicates a consistent recall across different runs. The model showed a median accuracy of almost 73%, with a low variance of 3.5%, which reveals stable performance across different runs. Notably, the FFS results showed slightly higher percentages than the model trained with PCA, indicating that FFS overperformed the PCA method in correctly identifying HV individuals and MO patients.

Discussion

Over the past decade, ML algorithms have experienced rapid growth, primarily driven by significant computing power and technology advancements. This progress has enabled the creation of more sophisticated and complex models, leading to breakthroughs in various fields, including healthcare. As a result, ML has transitioned from theoretical research to practical applications, significantly impacting how we solve complex problems and analyze vast amounts of data [21]. In the headache field, particularly migraine, ML techniques have been used to improve the diagnostic accuracy of the non-specialist [22], identify imaging biomarkers [23], detect neurophysiological changes before the attacks [24], classify migraine subtypes [8, 9, 10, 13, 25, 26], and predict treatment response [27, 28]. One area of interest has been using neurophysiological data to classify migraine patients. Hsiao et al. showed the utility of sensory-evoked oscillations from electroencephalogram activity in distinguishing between chronic migraine patients and healthy controls. Although the small sample size and the absence

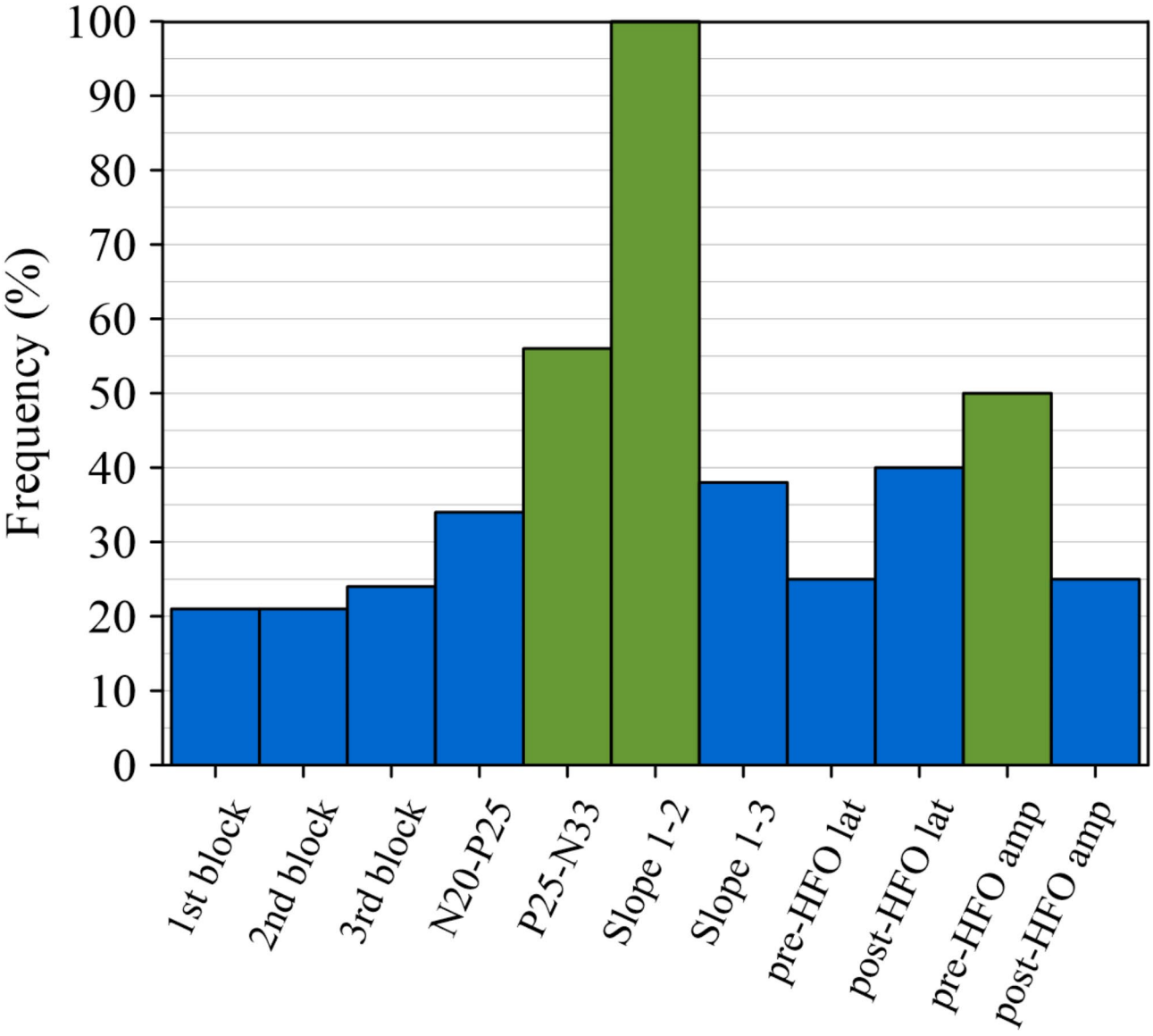

Fig. 4 Feature selection frequency of the 11 considered variables obtained by training 100 neural networks. The green columns indicate the selected features for training the ANN

of episodic migraine patients, they achieved an accuracy of 87.5% with an AUC of 0.84 with the best model [9]. Recently, the same group developed a classification model based on data obtained on the combination of somatosensory evoked responses and scores of psychometric assessments. This classification model performed well in differentiating chronic migraine patients (accuracy = 81.8%, AUC = 0.86) and episodic migraine patients from healthy controls (accuracy = 77.5%, AUC = 0.84). However, it is important to note that when the model used only the features of SSEP responses and connectivity, its performance dropped, reaching an accuracy of 72.3% with an AUC of 0.742 in distinguishing episodic migraine patients from healthy controls [13].

Zhu and colleagues used data from SSEP recordings, such as latency, amplitude, and spectral power features, to develop a model to distinguish healthy controls from migraine patients during the interictal and ictal phases. Their model achieved over 88% accuracy in classifying migraine ictal or interictal versus healthy volunteers, with 73.3% accuracy for healthy volunteers-ictal-interictal classification tasks [8]. Similar to our study, they employed SSEP recordings and similar machine learning techniques, including PCA and FFS for feature selection. However, it is important to note that they used a different data collection and extraction method to augment the number of data and compensate for the low sample size. Indeed, they recorded SSEP by collecting 425 to 2000 independent trials for each subject, averaged every

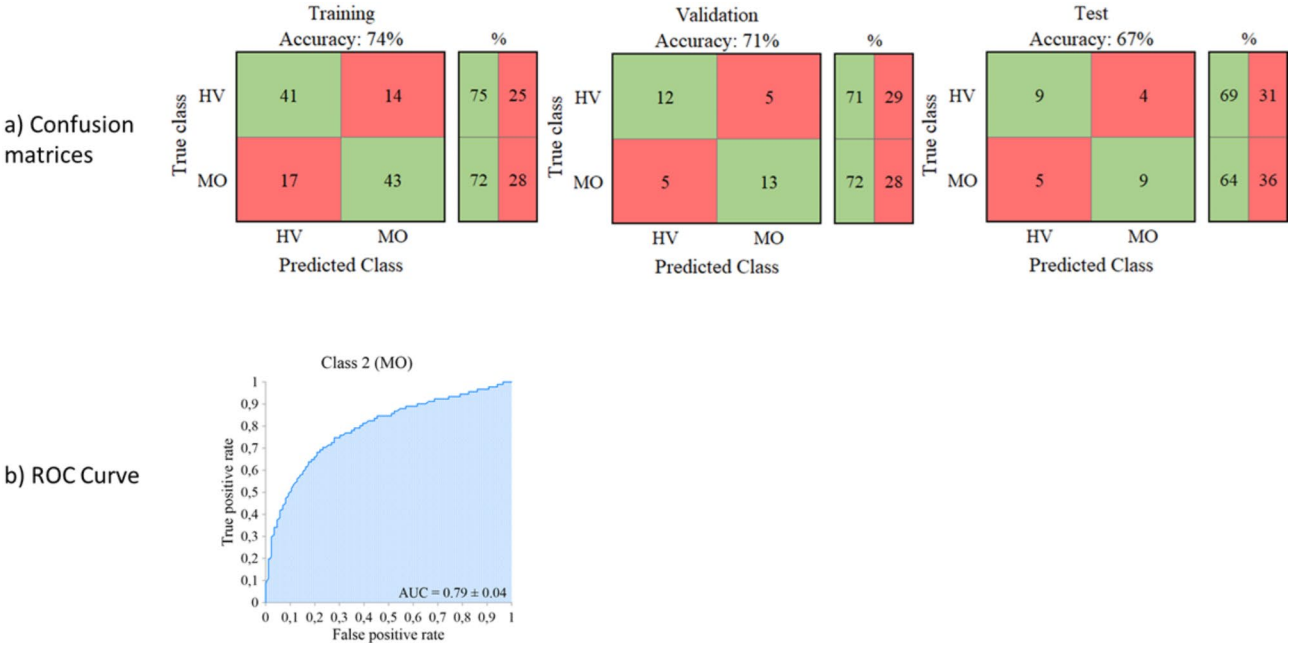

Fig. 5 **a**) Confusion matrices for a detailed summary of the ANN's performance trained with the Forward Feature Selection (FFS). The matrix is structured such that the rows represent the true classes (HV or MO), and the columns represent the predicted classes. The diagonal elements (in green) reflect the number of correctly classified cases (true negatives and true positives: TN and TP), while the off-diagonal elements (in red) capture the number of misclassifications false negatives and false positives: FN - FP). A row summary displays the percentages of correctly and incorrectly classified observations for each true class. For instance, considering the training dataset, in the "HV" class row, the percentage of correctly classified HV is displayed alongside the percentage misclassified as having MO (TN = 75%; FP = 25%). Similarly, for the "MO" class, the row summary reflects the percentage of individuals correctly identified as having MO (TP = 72%), and those misclassified as HV (FN = 28%). **b**) Area under the curve (AUC) for the overall performance of the ANN trained with the features selected with the FFS

Table 3 Overall performance of the ANN trained with the selected features from forward feature selection (FFS) in terms of the median and standard deviation (SD) metrics obtained by running 100 neural networks, using the three selected features as input and dividing the dataset into training (65%), validation (20%) and test (15%) sets

	MO	HV
Median Accuracy ± SD	0.728 ± 0.035	
Median Sensitivity (Recall) ± SD	0.715 ± 0.052	0.738 ± 0.087
Median Specificity ± SD	0.738 ± 0.087	0.715 ± 0.052
Median F1 Score ± SD	0.727 ± 0.036	0.725 ± 0.056

40 consecutive sweeps, and labeled the resulting waveform as a healthy volunteer, ictal, or interictal, depending on the subject under study. This resulted in a total of 325 ictal, 534 interictal, and 323 HV samples from a baseline sample of 42 migraine patients, 29 interictal and 13 ictal patients, and 15 healthy volunteers [8]. This could be the reason why we obtained lower values in the performance metrics and different selected features. However, we believe a single-subject analysis is preferred when designing a clinical migraine classification system.

In this study, we aimed to evaluate the potential of ANNs in distinguishing migraine patients from healthy individuals using neurophysiological features derived from SSEP recordings. We found that the ANN model trained with the FFS results outperformed the model trained with PCA results. Indeed, the model trained with PCA results showed a median accuracy of 68.9% with an AUC of 0.75. In contrast, the model trained with FFS results achieved a median accuracy of almost 73% with an AUC of 0.79. These different performances could be due to the fact that unless these techniques are both used for handling high-dimensional data, their objectives and methodologies differ significantly. PCA aims to reduce dimensionality by transforming the original correlated features into a smaller set of uncorrelated principal components, capturing the maximum variance in the data. This transformation enhances interpretability and computational efficiency by focusing on the directions of greatest variance [18]. In contrast, FFS focuses on improving model performance by iteratively selecting the most relevant subset of original features based on a predefined performance metric, such as accuracy [19]. One could hypothesize that the PCA's reliance on linear combinations of original features might have limited the ANN's ability to capture more complex, non-linear patterns specific to migraine pathophysiology, which the FFS could capture [29]. Indeed, the FFS directly identified three parameters (Slope 1–2, P25-N33, and pre-HFO amp) that were selected with a frequency of 50% or greater and were used as input for training the ANN

classification model. The slope between blocks 1–2 (i.e., habituation) and the amplitude of the pre-HFO represent some of the well-established neurophysiological alterations of sensory response characteristics of migraine patients [6, 30, 31], and our results support that this pattern of responses could also help distinguish MO from HV.

In particular, Slope 1–2 (the linear regression between the N20-P25 amplitude of the first and second blocks, i.e., habituation) was selected as a feature with a frequency of 100% in the best configuration. In contrast, even if statistically significant, the slope between the first and third blocks (Slope 1–3) was not selected as a relevant feature, with a percentage of being selected under 40% (Fig. 4). It is not yet clear whether the best methodology for calculating the slope is considering the linear regression between the first and second blocks or between the first and third blocks. Ozkul and al. Uckardes observed a clear-cut lack of habituation as early as the 2nd block of 100 averaged responses, which continued in the 3rd and 4th blocks [32]. Our present results could add an important point in favor of Slope 1–2 being more representative of migraine pathophysiology than Slope 1–3.

Several electrophysiological studies have shown that episodic migraine patients interictally are characterized by deficient habituation to repetitive stimulations, with almost all the non-painful evoked potentials [6], suggesting an altered processing of sensory information. This was further supported by the evidence that the magnitude of the habituation deficit directly correlated with the worsening of the disease [33]. Different hypotheses were proposed to explain the habituation deficit, including lower cortical activation levels due to abnormal thalamic control [31]. In particular, the study of HFOs has shown that thalamo-cortical activity is reduced in migraine patients between attacks while normalized during attacks [30]. This was further supported by the evidence of dynamic microstructural changes in bilateral thalami during the interictal period, which normalized during attacks [34]. Additionally, functional MRI studies revealed alteration in thalamocortical networks with abnormal connections of the posterior thalamus with the visual cortex and the precuneus [35] and reduced functional connectivity between the default mode network and the visuospatial system [36]. Then, anatomic or functional disconnection of the thalamus from its regulatory structures (monoaminergic brainstem nuclei) can cause "thalamo-cortical dysrhythmia", causing these cortical oscillatory dysfunctions [37]. In accordance with this hypothesis, we found a reduced amplitude for the first block of N20-P25 averaged responses and a reduced amplitude of pre-HFO, suggesting a decreased level of thalamocortical activation along with reduced pre-activation of the somatosensory cortex. However, only the pre-HFO amplitude was selected as a relevant feature approximately half of the time, while the amplitude of the first block was not selected as a relevant feature for improving the model. This suggests that, while both are valuable for understanding the migraine pathophysiology with a significant difference from a group level, only the amplitude of the pre-HFO could have a clinical value in distinguishing MO from HV.

Interestingly, we found that MO had an increased amplitude of P25-N33 compared to HV, which was selected by the FSS as a relevant feature for the classification model. We hypothesize that this result could be due to an inhibition deficit. In support of this interpretation, Valeriani et al. found a disinhibited recovery of the P24 and N30 SSEP components after paired-pulse stimulation in children with migraine [38]. The origin of the P25-N33 complex has long been debated. It was hypothesized that both the somatosensory and rostral-region of the supplementary motor area play a part in developing this potential [39, 40, 41]. The role of the somatosensory cortex in migraine pathophysiology is well-known. Previous studies showed interictal structural changes in the somatosensory cortex of migraine patients [42] along with alterations in the regional cerebral blood flow [43], while the supplementary motor area seems to contribute to stress-related migraine pain perception [44]. Additionally, face pain, such as that experienced in trigeminal neuralgia, is associated with increased P27-N30 parieto-frontal amplitudes [45]. We argue that the recurrence of migraine pain may elicit neuroplastic alterations in the sensorimotor cortex.

In conclusion, we showed that the ANN trained with SSEP-derived variables holds promise as a noninvasive tool for migraine classification, offering potential for clinical application and deeper insights into migraine diagnostics.

We acknowledge that our study has some limitations. First, while the sample size was adequate for this analysis, increasing it could significantly enhance the network's generalization and improve the model's performance. This was demonstrated in our study by incorporating the dataset initially used as the test set into the training and validation sets, resulting in a four-percentage point improvement in overall accuracy. Secondly, the homogeneity of the migraine group (i.e., only episodic migraine without aura) may limit the applicability of the results to other migraine subtypes. Then, testing the model in classifying episodic migraine regarding the different phases of the migraine cycle or recognizing chronic migraine patients or patients with other forms of primary headaches would be crucial in improving the model's generalizability and validating its applicability in clinical settings.

Furthermore, the sex distribution was not balanced between the two groups (MO group with a higher

proportion of female participants compared to the HV group). While this imbalance may theoretically influence the results, several factors mitigate its potential impact. First, the dataset used for the ANN development was randomly split into training, validation and test sets. Moreover, this process was carried out one hundred times (the results are presented as the average of 100 ANNs), ensuring that the sex-related bias was diluted across all these procedures. It is also important to note that all the neural networks were trained without any information about the sex of the subjects. Then, it seems implausible that this imbalance has influenced the training of the ANNs. However, to further support the robustness of our findings and to scientifically address the concern that the sex imbalance could have biased our findings, we tested the performance of the ANN in distinguishing MO and HV depending on the sex (a detailed description of this analysis is provided in the Supplementary Materials). We found that the ANN correctly identified migraine patients with almost the same percentage in both sexes (median sensitivity of 0.72 for male MO and 0.71 for female MO) (Figure S1). This result confirms that the network generalizes well regardless of sex, further validating its robustness in distinguishing MO from HV patients.

Future studies should validate our findings in a different context. Additionally, exploring temporal data using sophisticated neural networks (e.g., Long-Short Term Memory) could provide insights into how neurophysiological markers and ANN performance change over time, particularly in response to treatment. Finally, integrating other modalities, such as imaging or genetic or biological data, could potentially enhance the predictive power of ANNs in migraine classification.

Conclusions

In this study, we provided evidence that the ANNs model trained with SSEP-derived variables satisfactorily distinguished MO patients from HV, achieving over 68% in all the performance metrics (accuracy, sensitivity, specificity, and F1 scores). The balanced sensitivity and specificity across classes suggest that the model is equally adept at identifying MO and HV. This is crucial for the clinical utility of such models, as it reduces the likelihood of misclassification, which could lead to inappropriate treatment decisions. Then, the ANNs model trained with SSEP-derived variables holds promise as a noninvasive tool for migraine classification. However, the overall accuracy (68.9 or 72.8%) suggests that while the model is robust, it still has potential for refinement. Future research could explore integrating additional neurophysiological techniques or clinical data, more advanced feature selection methods, or different neural network architectures to enhance model performance. Furthermore, extending this approach to a different primary headache could improve its applicability in the clinical setting.

Abbreviations

ANNs	Artificial Neural Networks
FP	False positives
FN	False negatives
FFS	Forward Feature Selection
HFOs	High-frequency oscillations
HVs	Healthy volunteers
ICHD-III	International Classification of Headache Disorders third edition
LF-SSEPs	Low-frequency responses
ML	Machine Learning
MO	Episodic migraine without aura
PCA	Principal Component Analysis
SSEPs	Somatosensory evoked potentials
TP	True positives
TN	True negatives

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s10194-025-01989-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

G.S. and D.S. contributed equally to the paper. G.S., D.S., and G.C. contributed substantially to protocol development, interpretation of data, statistics, drafting and reviewing the manuscript. D.S. performed the design and evaluation of the artificial neural network. C.A., M.S., and C.D.L. contributed to participant enrolment. F.C., G.S., and G.C. performed the electrophysiological recordings. S.J.W. and F.J.H. were implied in the revision of the manuscript and data interpretation. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive external funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests relevant to this research.

Ethics approval

All the participants provided written informed consent to participate in the study, which was approved by the local ethics committee.

Author details

[1]Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome Polo Pontino ICOT, Via Franco Faggiana 1668, Latina 04100, Italy
[2]Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, Parma 43124, Italy
[3]Brain Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan
[4]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[5]Department of Neurology, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan

Received: 17 January 2025 / Accepted: 25 February 2025
Published online: 04 April 2025

References

1. Steiner TJ, Stovner LJ (2023) Global epidemiology of migraine and its implications for public health and health policy. Nat Rev Neurol 19:109–117. https://doi.org/10.1038/s41582-022-00763-1
2. Steiner TJ, Stovner LJ, Jensen R et al (2020) Migraine remains second among the world's causes of disability, and first among young women: findings from GBD2019. J Headache Pain 21:137. https://doi.org/10.1186/s10194-020-01208-0
3. Ashina M, Terwindt GM, Al-Karagholi MA-M et al (2021) Migraine: disease characterisation, biomarkers, and precision medicine. Lancet 397:1496–1504. https://doi.org/10.1016/S0140-6736(20)32162-0
4. (2018) Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition. Cephalalgia 38:1–211. https://doi.org/10.1177/0333102417738202
5. Puledda F, Viganò A, Sebastianelli G et al (2023) Electrophysiological findings in migraine May reflect abnormal synaptic plasticity mechanisms: A narrative review. Cephalalgia 43:3331024231195780. https://doi.org/10.1177/03331024231195780
6. Coppola G, Di Lorenzo C, Schoenen J, Pierelli F (2013) Habituation and sensitization in primary headaches. J Headache Pain 14:65. https://doi.org/10.1186/1129-2377-14-65
7. Orrù G, Pettersson-Yeo W, Marquand AF et al (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 36:1140–1152. https://doi.org/10.1016/j.neubiorev.2012.01.004
8. Zhu B, Coppola G, Shoaran M (2019) Migraine classification using somatosensory evoked potentials. Cephalalgia 39:1143–1155. https://doi.org/10.1177/0333102419839975
9. Hsiao F-J, Chen W-T, Wang Y-F et al (2023) Identification of patients with chronic migraine by using sensory-evoked oscillations from the electroencephalogram classifier. Cephalalgia 43:3331024231176074. https://doi.org/10.1177/03331024231176074
10. Hsiao F-J, Chen W-T, Pan L-LH et al (2022) Resting-state Magnetoencephalographic oscillatory connectivity to identify patients with chronic migraine using machine learning. J Headache Pain 23:130. https://doi.org/10.1186/s10194-022-01500-1
11. de Tommaso M, Ambrosini A, Brighina F et al (2014) Altered processing of sensory stimuli in patients with migraine. Nat Rev Neurol 10:144–155. https://doi.org/10.1038/nrneurol.2014.14
12. Hsiao F-J, Chen W-T, Pan L-LH et al (2022) Dynamic brainstem and somatosensory cortical excitability during migraine cycles. J Headache Pain 23:21. https://doi.org/10.1186/s10194-022-01392-1
13. Hsiao F-J, Chen W-T, Liu H-Y et al (2024) Altered brainstem–cortex activation and interaction in migraine patients: somatosensory evoked EEG responses with machine learning. J Headache Pain 25:185. https://doi.org/10.1186/s10194-024-01892-2
14. Sebastianelli G, Abagnale C, Casillo F et al (2022) Bimodal sensory integration in migraine: A study of the effect of visual stimulation on somatosensory evoked cortical responses. Cephalalgia 42:654–662. https://doi.org/10.1177/03331024221075073
15. Sebastianelli G, Casillo F, Abagnale C et al (2023) Central sensitization mechanisms in chronic migraine with medication overuse headache: a study of thalamocortical activation and lateral cortical Inhibition. Cephalalgia 43:3331024231202240. https://doi.org/10.1177/03331024231202240
16. Petrušić I, Savić AA, Mitrović K et al (2024) Machine learning classification Meets migraine: recommendations for study evaluation. J Headache Pain 25:215. https://doi.org/10.1186/s10194-024-01924-x
17. Shanker M, Hu MY, Hung MS (1996) Effect of data standardization on neural network training. Omega 24:385–397. https://doi.org/10.1016/0305-0483(96)00010-2
18. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philosophical Trans Royal Soc A: Math Phys Eng Sci 374:20150202. https://doi.org/10.1098/rsta.2015.0202
19. Liu H, Yu L (2005) Toward integrating feature selection algorithms for classification and clustering. IEEE Trans Knowl Data Eng 17:491–502. https://doi.org/10.1109/TKDE.2005.66
20. Secci D, Molino L, Zanini A (2022) Contaminant source identification in groundwater by means of artificial neural network. J Hydrol 611:128003. https://doi.org/10.1016/j.jhydrol.2022.128003
21. Hinton G (2018) Deep Learning—A technology with the potential to transform health care. JAMA 320:1101–1102. https://doi.org/10.1001/jama.2018.11100
22. Katsuki M, Shimazu T, Kikui S et al (2023) Developing an artificial intelligence-based headache diagnostic model and its utility for non-specialists' diagnostic accuracy. Cephalalgia 43:03331024231156925. https://doi.org/10.1177/03331024231156925
23. Messina R, Sudre CH, Wei DY et al (2023) Biomarkers of migraine and cluster headache: differences and similarities. Ann Neurol 93:729–742. https://doi.org/10.1002/ana.26583
24. Cao Z, Lai K-L, Lin C-T et al (2018) Exploring resting-state EEG complexity before migraine attacks. Cephalalgia 38:1296–1306. https://doi.org/10.1177/0333102417733953
25. Schwedt TJ, Chong CD, Wu T et al (2015) Accurate classification of chronic migraine via brain magnetic resonance imaging. Headache 55:762–777. https://doi.org/10.1111/head.12584
26. Chong CD, Gaw N, Fu Y et al (2017) Migraine classification using magnetic resonance imaging resting-state functional connectivity data. Cephalalgia 37:828–844. https://doi.org/10.1177/0333102416652091
27. Gonzalez-Martinez A, Pagán J, Sanz-García A et al (2022) 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 29:3102–3111. https://doi.org/10.1111/ene.15458
28. Romozzi M, Lokhandwala A, Vollono C et al (2024) An evolving machine-learning-based algorithm to early predict response to anti-CGRP monoclonal antibodies in patients with migraine. Cephalalgia 44:03331024241262751. https://doi.org/10.1177/03331024241262751
29. Monahan AH (2000) Nonlinear principal component analysis by neural networks: theory and application to the Lorenz system. J Clim 13:821–835. https://doi.org/10.1175/1520-0442(2000)013<0821:NPCABN>2.0.CO;2
30. Coppola G, Vandenheede M, Di Clemente L et al (2005) Somatosensory evoked high-frequency oscillations reflecting thalamo-cortical activity are decreased in migraine patients between attacks. Brain 128:98–103. https://doi.org/10.1093/brain/awh334
31. Coppola G, Pierelli F, Schoenen J (2007) Is the cerebral cortex hyperexcitable or hyperresponsive in migraine? Cephalalgia 27:1427–1439. https://doi.org/10.1111/j.1468-2982.2007.01500.x
32. Ozkul Y, Uckardes A (2002) Median nerve somatosensory evoked potentials in migraine. Eur J Neurol 9:227–232. https://doi.org/10.1046/j.1468-1331.2002.00387.x
33. Restuccia D, Vollono C, Virdis D et al (2014) Patterns of habituation and clinical fluctuations in migraine. Cephalalgia 34:201–210. https://doi.org/10.1177/0333102413508241
34. Coppola G, Tinelli E, Lepre C et al (2014) Dynamic changes in thalamic microstructure of migraine without aura patients: a diffusion tensor magnetic resonance imaging study. Eur J Neurol 21:287–e13. https://doi.org/10.1111/ene.12296
35. Tu Y, Fu Z, Zeng F et al (2019) Abnormal thalamocortical network dynamics in migraine. Neurology 92:e2706–e2716. https://doi.org/10.1212/WNL.0000000000007607
36. Coppola G, Di Renzo A, Tinelli E et al (2016) Thalamo-cortical network activity between migraine attacks: insights from MRI-based microstructural and functional resting-state network correlation analysis. J Headache Pain 17:100. https://doi.org/10.1186/s10194-016-0693-y
37. Llinás RR, Ribary U, Jeanmonod D et al (1999) Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. Proc Natl Acad Sci U S A 96:15222–15227. https://doi.org/10.1073/pnas.96.26.15222
38. Valeriani M, Rinalduzzi S, Vigevano F (2005) Multilevel somatosensory system disinhibition in children with migraine. Pain 118:137–144. https://doi.org/10.1016/j.pain.2005.08.026
39. Allison T, McCarthy G, Luby M et al (1996) Localization of functional regions of human mesial cortex by somatosensory evoked potential recording and by cortical stimulation. Electroencephalogr Clin Neurophysiol 100:126–140. https://doi.org/10.1016/0013-4694(95)00226-x

40. Rossini PM, Babiloni F, Babiloni C et al (1997) Topography of spatially enhanced human shortlatency somatosensory evoked potentials. NeuroReport 8:991
41. Barba C, Valeriani M, Colicchio G, Mauguière F (2005) Short and middle-latency median nerve (MN) SEPs recorded by depth electrodes in human pre-SMA and SMA-proper. Clin Neurophysiol 116:2664–2674. https://doi.org/10.1016/j.clinph.2005.07.022
42. DaSilva AFM, Granziera C, Snyder J, Hadjikhani N (2007) Thickening in the somatosensory cortex of patients with migraine. Neurology 69:1990–1995. https://doi.org/10.1212/01.wnl.0000291618.32247.2d
43. Hodkinson DJ, Veggeberg R, Wilcox SL et al (2015) Primary somatosensory cortices contain altered patterns of regional cerebral blood flow in the interictal phase of migraine. PLoS ONE 10:e0137971. https://doi.org/10.1371/journal.pone.0137971
44. Kökönyei G, Galambos A, Kocsel N et al (2021) Inter-individual differences in pain anticipation and pain perception in migraine: neural correlates of migraine frequency and cortisol-to-dehydroepiandrosterone sulfate (DHEA-S) ratio. PLoS ONE 16:e0261570. https://doi.org/10.1371/journal.pone.0261570
45. Tinazzi M, Valeriani M, Moretto G et al (2004) Plastic interactions between hand and face cortical representations in patients with trigeminal neuralgia: a somatosensory-evoked potentials study. Neuroscience 127:769–776. https://doi.org/10.1016/j.neuroscience.2004.05.020

Publisher's note