



OPEN ACCESS

EDITED BY

Daniele Corbo,
University of Brescia, Italy

REVIEWED BY

Miguel J. A. Láinez,
Hospital Clínico Universitario de Valencia,
Spain

Gianluca Coppola,
Sapienza University of Rome, Italy

Bing Si,
Binghamton University, United States

*CORRESPONDENCE

Katarina Mitrović
✉ kacam.ftn@gmail.com

RECEIVED 23 November 2022

ACCEPTED 22 May 2023

PUBLISHED 23 June 2023

CITATION

Mitrović K, Petrušić I, Radojičić A,
Daković M and Savić A (2023) Migraine with
aura detection and subtype classification using
machine learning algorithms and
morphometric magnetic resonance imaging
data.

Front. Neurol. 14:1106612.

doi: 10.3389/fneur.2023.1106612

COPYRIGHT

© 2023 Mitrović, Petrušić, Radojičić, Daković
and Savić. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data

Katarina Mitrović^{1*}, Igor Petrušić², Aleksandra Radojičić^{3,4},
Marko Daković² and Andrej Savić⁵

¹Department of Information Technologies, Faculty of Technical Sciences in Čačak, University of Kragujevac, Čačak, Serbia, ²Laboratory for Advanced Analysis of Neuroimages, Faculty of Physical Chemistry, University of Belgrade, Belgrade, Serbia, ³Headache Center, Neurology Clinic, Clinical Center of Serbia, Belgrade, Serbia, ⁴Faculty of Medicine, University of Belgrade, Belgrade, Serbia, ⁵Science and Research Centre, School of Electrical Engineering, University of Belgrade, Belgrade, Serbia

Introduction: Migraine with aura (MwA) is a neurological condition manifested in moderate to severe headaches associated with transient visual and somatosensory symptoms, as well as higher cortical dysfunctions. Considering that about 5% of the world's population suffers from this condition and manifestation could be abundant and characterized by various symptoms, it is of great importance to focus on finding new and advanced techniques for the detection of different phenotypes, which in turn, can allow better diagnosis, classification, and biomarker validation, resulting in tailored treatments of MwA patients.

Methods: This research aimed to test different machine learning techniques to distinguish healthy people from those suffering from MwA, as well as people with simple MwA and those experiencing complex MwA. Magnetic resonance imaging (MRI) post-processed data (cortical thickness, cortical surface area, cortical volume, cortical mean Gaussian curvature, and cortical folding index) was collected from 78 subjects [46 MwA patients (22 simple MwA and 24 complex MwA) and 32 healthy controls] with 340 different features used for the algorithm training.

Results: The results show that an algorithm based on post-processed MRI data yields a high classification accuracy (97%) of MwA patients and precise distinction between simple MwA and complex MwA with an accuracy of 98%. Additionally, the sets of features relevant to the classification were identified. The feature importance ranking indicates the thickness of the left temporal pole, right lingual gyrus, and left pars opercularis as the most prominent markers for MwA classification, while the thickness of left pericalcarine gyrus and left pars opercularis are proposed as the two most important features for the simple and complex MwA classification.

Discussion: This method shows significant potential in the validation of MwA diagnosis and subtype classification, which can tackle and challenge the current treatments of MwA.

KEYWORDS

migraine with aura, machine learning, magnetic resonance imaging, artificial intelligence, classification

1. Introduction

Migraine is one of the most common neurological disorders that affects over a billion people worldwide and manifests as an episodic headache often accompanied by nausea, vomiting and sensitivity to light and/or sound (1, 2). According to the Global Burden of Disease (GBD) study conducted for the period from 1990 to 2019, migraine is the second biggest cause of disability at the fourth level of the GBD scale for people of any gender and age, while among female individuals aged 15–49 migraine ranks first (3, 4). Migraine can be classified into two major subtypes: migraine without aura (MwoA) and migraine with aura (MwA) (5–7). This division is based on a distinct pattern of inheritance in these two subtypes, different disorders and health conditions they cause, variant structural changes in the brain, dissimilar levels of brain activity and different responses to therapies and preventive measures (6, 7). Our research relies on the studies that propose considering MwA and MwoA as two entities and investigating them separately. Nevertheless, it should be noted that many authors do not consider them to be two different entities.

Migraine without aura is present in almost one-third of migraine patients and 5% of people worldwide (8). Typical MwA is characterized by completely reversible visual and somatosensory symptoms and speech disturbances (5). In addition, it has been shown that each migraine attack with aura may have some unique characteristics and differ from other attacks (9). Many studies use magnetic resonance imaging (MRI) as a data source to find important evidence of the impact of MwA on the brain and thereby contribute to endeavors to achieve the most effective diagnosis and treatment of MwA. These studies have made significant findings about the abnormalities in certain brain regions and brain networks associated with migraine (10). Neuroimaging findings in MwA patients have greatly contributed to a better understanding of this disease, but many researchers are still trying to find reliable markers for the diagnosis and MwA treatment (11, 12). The following neuroimaging techniques based on different physical principles are applied to MwA research (10, 11): (1) perfusion-weighted MRI, (2) diffusion-weighted imaging, (3) blood oxygen level-dependent imaging, (4) magnetic resonance spectroscopy, and (5) positron emission tomography. However, two types of MRI data are typically used in migraine classification: functional MRI (fMRI) and structural MRI. Studies based on fMRI have significantly contributed to understanding the brain mechanisms underlying migraine symptoms and processes in the brain during and between migraine attacks (10, 11, 13). Also, the development of structural MRI-based studies arose with the parallel advances in the technology of imaging and the pathophysiologic understanding of migraine (11). Therefore, the research in this area reported structural changes in migraine in white and gray matter and delivered insights into migraine pathophysiology that can provide a useful basis for discovering a reliable biomarker for MwA and even its subtypes (14, 15).

During recent years, artificial intelligence (AI) is increasingly present in various domains of neurological research. Many important findings have been discovered in the study of migraine using machine learning (ML) techniques with MRI data. Recently, this methodology has been widely applied for migraine classification, as well as the identification of brain regions that are

important for migraine diagnosis and treatment (12, 13, 16, 17). The right middle temporal, posterior insula, middle cingulate, left ventromedial prefrontal and bilateral amygdala regions best discriminated the migraine brain from HCs, and 97% classification accuracy was achieved for brain resting state MRI data of migraine patients with over 14 years long disease durations (13). A study using regional cortical thickness, cortical surface area and volume MRI data achieved an accuracy of 68% for migraine and HCs classification, 67% for episodic migraine and HCs classification, 86% for chronic migraine and HCs classification and 84% for chronic and episodic migraine classification (17). The temporal pole, anterior cingulate cortex, superior temporal lobe, entorhinal cortex, medial orbital frontal gyrus and pars triangularis were commonly selected measures for the classification tasks (17). The main potential of ML application in this field is providing an aid in migraine diagnosis and facilitating the process of distinguishing different migraine subtypes (12). This progress and the development of AI techniques and applications are one of the main motivations for our research. ML algorithms based on brain resting-state fMRI and structural MRI have been used to identify brain regions and networks involved in migraine attacks, and/or brain signatures that discriminate migraine patients from healthy controls (10, 13, 16, 17), yet fewer studies focused on implementing such models to solve MwA classification problems.

This research applies different ML algorithms to find a classification method that can distinguish HCs from those suffering from MwA, as well as people with simple MwA (MwA-S) and those experiencing complex MwA (MwA-C). Also, we aimed to identify the sets of MRI data-features that are key to these classifications.

2. Materials and methods

2.1. Participants

The study includes HCs and participants with typical episodic MwA. The diagnosis of MwA was based on the third International Classification of Headache Disorders criteria (5). All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The data-collecting protocol was authorized by the Review Board of the Neurology Clinic. The consent of all subjects to participate in the study was mandatory. The following inclusion criteria were applied: (1) 18–55 years of age, (2) suffering from episodic migraines with typical aura for more than 5 years before the enrolment in the study, (3) minimum of two MwA attacks per year, (4) absence of migraine preventive therapy, and (5) right-hand side of body predominance to avoid possible differences in brain regions. Also, the following exclusion criteria were applied: (1) the presence of other types of headaches (except occasional migraine without aura or tension headache), (2) the presence of any other neurological, cardiovascular or metabolic disorder determined through medical history or during a physical examination, (3) reported claustrophobia or inability to perform MRI examination, and (4) structural abnormalities on MRI scan. Also, MwA patients did not experience a migraine 72 h before and after the MRI scan.

2.2. MRI data acquisition and post-processing

The MRI examination was performed on a 3T Scanner (MAGNETOM Skyra, Siemens, Erlangen, Germany). Protocol for MRI examination was: (1) 3D T1 (repetition time (TR) = 2,300 ms, echo time (TE) = 2.98 ms, flip angle = 9°, 130 slices with voxel size $1 \times 1 \times 1 \text{ mm}^3$, acquisition matrix 512×512 and FOV = $256 \times 256 \text{ mm}^2$), (2) 3D FLAIR (TR = 5,000 ms, TE = 398 ms, TI = 1800 ms, flip angle = 120°, acquisition matrix 256×256 , FOV = $256 \times 256 \text{ mm}$) and (3) T2 weighted spin echo [T2W] in an axial plane [TR = 4,800 ms, TE = 92 ms, flip angle (FA) = 90°, acquisition matrix 384×265 , FOV = $256 \times 256 \text{ mm}$, slice thickness = 5 mm]. T2W images were only used to exclude the presence of brain lesions.

Freesurfer (v 6.0) analysis was performed on an HP DL850 server (Intel Xeon 3.2MHz, eight cores, 16GB RAM) using a recon-all script, combining 3D T1 and FLAIR images, for automatic cortical reconstruction and segmentation of brain structures. The average run time (with the parallelization option used) was 6h. Details about Freesurfer and its routines can be found in other studies (18). Cortical parcellation was done according to the Desikan-Killiany Atlas (19). Post-processed MRI data includes the cortical thickness, surface area, volume, mean Gaussian curvature and folding index collected from the left and right brain hemispheres. The data set includes 340 numeric features used for the algorithm training.

2.3. Machine learning

Based on the data collected from the participants in the study, two data sets were created for ML algorithm training. The first data set contains the aforementioned 340 input features and one output that classifies 78 subjects into healthy individuals and MWA patients (MWA classification). The second data set contains identical input features, while the output categorizes 46 MWA patients into those with the MWA-S and MWA-C (MWA subgroup classification). MWA patients with only visual symptoms were labelled as MWA-S, while subjects that experienced additional symptoms such as somatosensory symptoms and/or dysphasia were categorized into the MWA-C subgroup. Values of all features are within their anatomical limits.

The feature selection and ML models were developed in the Python programming language (version 3.8) in the Jupyter Notebook environment. The functions of the Scikit-learn software library for ML, Pandas library for data manipulation and analysis, NumPy library for mathematical functions and Matplotlib library for creating graphs were implemented. The hardware configuration used in this research included an NVIDIA GeForce GTX 1650 Ti GPU, AMD Ryzen 54600H 3.00GHz central processing unit and 8GB of random-access memory.

Feature selection contributes to model simplification by reducing the feature number, decreases the training time, reduces overfitting by enhancing generalization and helps with solving the problem of the dimensionality curse (20). In this paper, the feature selection was performed using the Extremely Randomized Trees (ERT) algorithm. This algorithm creates multiple uncorrelated decision trees over different subsamples of the data set, combines their predictions and returns a result (21). The algorithm evaluates

the importance of each input feature in the data set. The 40 most important features are retained, using a combinatorial search to find optimal sets of features with the aim of error minimization (22, 23). ML classification algorithms were applied to the data set with the selected features.

Several different ML algorithms were applied in this study and hyperparameter tuning was performed for each algorithm. To design a proper configuration of ML algorithms for a specific purpose and refine them for application to a particular data set, it is necessary to tune hyperparameters and explore a range of functions and values. Hyperparameter tuning performed in this study is based on comprehensive research of state-of-the-art hyperparameter optimization rules and their application to the different ML models (24). The following sections provide descriptions of each implemented ML model such as technical details, the range of hyperparameters being tested, and optimal hyperparameters obtained by exhaustive search based on the resulting accuracy. All algorithms were evaluated using leave-one-out cross-validation method. Leave-one-out cross-validation is a special case of k-fold cross-validation commonly applied to data sets with a small number of instances, where the number of folds is equal to the number of instances (25). The study was conducted for MWA and MWA subgroup classification separately using the same methodology. The outcomes of algorithms using different hyperparameter values were compared and the solutions that provided the best results are presented in the Results section of this paper.

2.3.1. Logistic regression

Logistic Regression (LR) showed great potential when predicting the risk of major chronic diseases with a low number of events and simple clinical predictors using a moderate sample size (26). The basic LR equation is as follows:

$$\hat{y} = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}} \quad (1)$$

where \hat{y} represents the estimated output, x_i stands for the data sample, β_0 is the intercept or the constant value where the regression line crosses the vertical axis, β_i is the weight coefficient for input feature x_i that determines the contribution of corresponding input to the accurate output prediction, and n is the number of samples (27). It can be noted that the equation:

$$\hat{y} = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (2)$$

represents the Linear Regression, i.e., the linear model underlying LR. The main LR hyperparameter is the cost function which depends on the regularization method of the penalization, such as L1 or Lasso regularization, L2 or Ridge regularization, and other non-conventional regularization methods that usually combine L1 and L2 (24). Equations for L1 and L2 regularization are as follows:

$$L1 = \lambda \sum_{i=1}^n |\beta_i| \quad (3)$$

$$L2 = \lambda \sum_{i=1}^n \beta_i^2 \tag{4}$$

$$\min(c - 1, f) \tag{9}$$

where λ shows the regularization strength, n is the number of samples, and, β_i represents the weight coefficient (28). In this research, L1 and L2 penalties were tested, and L2-regularized LR provided the best results. The algorithm was trained in 5,000 iterations.

2.3.2. Linear discriminant analysis

Linear Discriminant Analysis (LDA) is commonly applied to data sets with high dimensionality and a large number of features to reduce dimensionality and determine a feature subspace in which the data samples are separable (29). The objective of LDA is to minimize the variance inside each class and maximize the variance between different classes using the equation of maximization of Fisher’s criterion:

$$\max_{\varphi} J(\varphi) = \max_{\varphi} \frac{\varphi^T S_{bc} \varphi}{\varphi^T S_{wc} \varphi} \tag{5}$$

where φ is an orientation matrix that is determined as the solution of the eigenvalue problem, S_{bc} implies scatter matrices that contain between-class variances:

$$S_{bc} = \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \tag{6}$$

while S_{wc} implies scatter matrices that contain within-class variances:

$$S_{wc} = \sum_{i=1}^c \sum_{x \in X_i} (x - \bar{x}_i)(x - \bar{x}_i)^T \tag{7}$$

The notation c represents the number of classes in the data set, n_i is the number of samples in class i , \bar{x}_i is the mean value of class i , \bar{x} is the mean value of n samples, X_i is the subset of the data set consisted of input samples that belong to class i , and x is a data sample (29, 30). The LDA can also be interpreted as assigning x_i to the class i , $i \in \{1, \dots, c\}$, whose mean is the closest based on the Mahalanobis distance, while also considering the prior probabilities of the class, using the equation:

$$\log P(i|x) = -\frac{1}{2}(x - \bar{x}_i)^T S^{-1}(x - \bar{x}_i) + \log P(i) + C \tag{8}$$

where \bar{x}_i is the mean value of class i , S is the covariance matrix based on the assumption that all classes share the same covariance matrix, $P(i)$ is the class prior probability, and C is a constant term (31). It can be noted that $(x - \bar{x}_i)^T S^{-1}(x - \bar{x}_i)$ represents the Mahalanobis distance.

The first hyperparameter that was tuned within the LDA algorithm is the number of features to be extracted which is calculated as follows:

where c implies the number of classes, whereas f represents the number of features in the data set (24). Further, three types of solvers were tested: Singular Value Decomposition (SVD), Least squares solution, and Eigenvalue Decomposition. The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal (32). SVD solver is based on the assumption that the singular value diagonal covariance matrix can be determined as:

$$S = U^T X V \tag{10}$$

where the rows of U are eigenvectors of $X^T X$, and the columns of V are eigenvectors of XX^T (33). Therefore, the prior probabilities of the class can be obtained without explicitly computing the covariance matrix. Least squares LDA calculates the covariance matrix as $Sw = \bar{x}_i$, where w is the weight vector. The Eigenvalue Decomposition solver is based on the optimization of equations (6) and (7). SVD LDA is the most suitable algorithm for training high-dimensionality data. Each solver was implemented without a shrinkage parameter, which implies using the empirical covariance matrix as an estimate for the covariance matrix, as well as with the Ledoit-Wolf lemma (34). The shrinkage parameter is commonly applied to high-dimensionality problems, where the sample number is notably higher than the feature number, and it may result in improved estimation of covariance matrices. In this research, SVD LDA obtained the best results. It should be noted that an SVD solver does not compute the covariance matrix, hence cannot be used with the covariance matrix shrinkage.

2.3.3. K-nearest neighbors

Within K-Nearest Neighbors (KNN) algorithm, the distance between samples is calculated and the output class is predicted based on the majority vote of the nearest k samples (35). In this study, the distance was measured according to the Euclidean distance equation:

$$D = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \tag{11}$$

where n represents the number of samples, while p and q are two data samples whose distance is being measured (36). The most important KNN hyperparameter is the number of nearest neighbors k (24). The values of hyperparameter k were tested starting with $k=2$, $k=3$ provided the best accuracy result, whereas further rising of the number of neighbors resulted in lower performance of the algorithm.

2.3.4. Classification and regression tree

A decision tree classifier implemented in this research is an optimized version of the Classification and Regression Tree (CART) algorithm. CART is based on binary trees where each test node contains exactly two possible outcomes of the test and leaf nodes represent a predicted outcome (37). At each node, the data set is partitioned recursively creating two subsets:

$$x_{lefti} = \{(x,y)|x_j \leq t_i\} \tag{12}$$

$$x_{righti} = x_i \setminus x_{lefti} \tag{13}$$

where x_i is a subset at node i , x is the input vector, y is the output, j represents a feature, and t_i is a threshold at node i . The suitability of each split option is determined using the splitting criteria, which measures the impurity of the nodes (38). Gini impurity and Shannon information gain are the two main types of impurity functions (24). Gini impurity is calculated as follows:

$$G = \sum_{i=1}^c x_i(1-x_i) \tag{14}$$

where c represents the number of classes in the data set, and x_i denotes the fraction of samples belonging to class i at a specific node (38). Shannon information gain or entropy is calculated using the following equation:

$$E = -\sum_{i=1}^c x_i \log_2 x_i \tag{15}$$

where c is the number of classes, and p_i denotes the fraction of samples belonging to class i at a current node (38). In addition, a strategy used to choose the split at each node can be chosen. Supported strategies are the best split and best random split (24). Another important hyperparameter that can be tuned within the CART algorithm is the number of features to consider when looking for the best split (24). Using the total number of features, calculating the square root, or calculating the binary logarithm of the total number of features are three examples of how to set the maximum number of features. In this study, the CART algorithm was tuned with Gini and entropy impurity functions, best and best random splitting strategies, and using all three previously mentioned values for the maximum number of features. The best results were achieved using the Gini function, best split strategy, and the total number of features as maximum.

2.3.5. Naive Bayes

Naive Bayes (NB) is a classification technique based on Bayes' theorem (39). A frequency table for each attribute is generated to calculate the posterior probability. These tables show the number of occurrences of each attribute value in each possible class. The frequency tables are converted into likelihood tables by calculating the ratios of class and overall frequencies. Further, the class and predictor prior probabilities are computed. Finally, the posterior probability is calculated as follows:

$$P(i|x) = \frac{P(i) \prod_{i=1}^f P(x_i|i)}{P(x)} \tag{16}$$

where f is the number of features, x is the input vector, i is the observed output class, $i \in \{1, \dots, c\}$, c is the number of classes, $P(i)$ is

class prior probability, $P(x)$ is predictor prior probability, and $P(x|i)$ is the likelihood (40). When equation (16) is applied to each class, the class with the highest probability is chosen as the final result. The types of NB classifiers mainly differ by the assumptions they make regarding the distribution of the likelihood. In this paper, the Gaussian NB classifier was implemented, where the likelihood is calculated as follows:

$$P(x|i) = \frac{1}{\sqrt{2\pi} s_{x_i}^2} e^{-\frac{(x-\bar{x}_i)^2}{2s_{x_i}^2}} \tag{17}$$

where $s_{x_i}^2$ denotes the variance, and \bar{x}_i is the mean of the input x for the observed class i (41). For the NB algorithm, no hyperparameter needs to be tuned (24).

2.3.6. Support vectors machine

Support Vector Machine (SVM) is an algorithm that transforms the problem space into a multidimensional space in order to make the problem linearly separable and to divide the classes using a hyperplane (42). The hyperplane is identified based on the margins between data samples from different classes, and it is used as a partition boundary for the classification.

The goal of SVM implementation used in this study is to find the solution for the following minimization problem:

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - \bar{1}^T \alpha \tag{18}$$

subject to $0 \leq \alpha_i \leq C, i = 1, \dots, n$

$$y^T \alpha = 0$$

where $\bar{1}$ is the vector of all ones, α_i are the dual coefficients, C is the upper bound, n is the number of samples, and Q is a positive semidefinite matrix $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, where $K(x_i, x_j)$ is the kernel (42). Kernel functions aim to measure the similarity between data samples x_i and x_j , and kernel type is one of the crucial SVM hyperparameters (24). Radial basis function (RBF), polynomial, and sigmoid kernel are tested in this research and their equations are listed below, respectively, (43):

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0 \tag{19}$$

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \tag{20}$$

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \tag{21}$$

Variables γ (gamma), r (coef0), and d (degree) are the hyperparameter of the kernels. The gamma defines the hyperplane depth and it can be tuned in all the above-mentioned kernels. The

coef0 is an independent term that can be used in polynomial and sigmoid functions. The degree hyperparameter of polynomial kernels determines the flexibility of the separation line.

The gamma hyperparameter was tuned using the following equations:

$$\gamma = 0.001 \quad (22)$$

$$\gamma = \frac{1}{f * s_x^2} \quad (23)$$

$$\gamma = \frac{1}{f} \quad (24)$$

where f represents the number of features, and s_x^2 is the variance of the input. Independent term coef0 was set to the value zero. The degree hyperparameter was set to a value of three. C-SVM with RBF kernel, regularization hyperparameter value of 100, L2 squared penalty, and gamma value calculated according to equation (24) obtained the best results.

2.3.7. Random forest

Random forest (RF) is an improved bagging algorithm in which a set of loosely correlated decision trees is created (44). The training is conducted over the bagged trees separately. At each split, the algorithm is considering a random subset of predictors, therefore avoiding the overfitting problem and providing more reliable results.

The main hyperparameter of the RF algorithm is the number of estimators or trees to be generated whose results are combined into the final prediction (24). The algorithm was trained with 100, 200, and 300 trees, where raising the number of trees led to better results. Further increase in the number of trees was limited by the computer power that was available.

Another important hyperparameter is the maximal number of features to be considered in each tree. In this research, the value of this hyperparameter was calculated as follows:

$$m = \sqrt{f} \quad (25)$$

where f is the number of features (44). The decision trees were tested with Gini and entropy impurity functions that measure the quality of a split. These functions are mathematically presented with equations (14) and (15). RF with 300 decision trees and entropy function obtained the highest accuracy in comparison to other hyperparameter values.

3. Results

This study was based on MRI post-processed data collected from 78 subjects (46 MwA patients and 32 HCs). Groups were balanced for age (MwA = 36.56 ± 9.03 (21–54 years range) vs. HCs = 35.67 ± 8.98 (19–55 years range), $p = 0.669$) and sex (MwA = 72% females vs. HCs = 70% females, $p = 1.000$). The study involved 22 MwA-S and 24 MwA-C patients. Subgroup characteristics including demographic data and aura features are shown in Table 1.

In this study, two data sets were created on which feature selection and ML algorithms were applied. Each data set addresses one classification task: MwA vs. HCs and MwA subgroup classification. Table 2 summarizes the classification results of healthy individuals and MwA sufferers as well as the classification of MwA-S and MwA-C (MwA subgroup) for seven ML algorithms (LR, LDA, KNN, CART, NB, SVM, and RF). The accuracies of the presented results are based on the leave-one-out cross-validation. For both classification problems, the best results were achieved by the LDA algorithm. The highest accuracy of 97% was achieved for MwA classification, while in the case of MwA subgroup prediction, an accuracy of 98% was obtained. The majority of algorithms (except SVM) achieved better results when classifying subgroups.

The importance and ranking of the 10 most prominent features for the LDA algorithm in relation to the average feature importance using all algorithms are presented in Tables 3, 4 for MwA and MwA subgroup classification, respectively. Feature importance was calculated using the ERT algorithm, resulting in assigning the importance score to each of the 340 features, which added up to a value of 100. When using the LDA algorithm that yielded the best classification performance, the feature importance ranking indicates the thickness of the left temporal pole, right lingual gyrus and left pars opercularis as the most prominent markers for MwA classification (Figure 1), while the thickness of left pericalcarine gyrus and left pars

TABLE 1 Participant characteristics for MwA subgroup classification.

Variable	Participants		Statistics
	MwA-S patients (n=22)	MwA-C patients (n=24)	
Age, mean \pm SD (range)	37.64 \pm 8.41 (19–55)	33.88 \pm 9.28 (20–55)	$p = 0.158$
Sex, number of females (%)	15 (68)	17 (71)	$p = 1.000$
Migraine frequency per year, mean \pm SD (range)	4.91 \pm 4.33 (2–20)	7.75 \pm 8.57 (1–30)	$p = 0.160$
Visual symptoms, number of patients (%)	22 (100)	24 (100)	/
Somatosensory symptoms, number of patients (%)	0 (0)	24 (100)	/
Dysphasic symptoms, number of patients (%)	0 (0)	24 (100)	/
Average aura duration (minutes), mean \pm SD (range)	30.23 \pm 16.65 (10–60)	41.04 \pm 14.96 (20–90)	$p = 0.025$

MwA, migraine with aura; MwA-S, simple migraine with aura; MwA-C, complex migraine with aura; SD, standard deviation.

TABLE 2 Classification accuracies across all machine learning algorithms tested.

Classification problem, %	Machine learning algorithm						
	LR	LDA	KNN	CART	NB	SVM	RF
HC vs. MwA (MwA detection)	87.18	97.44	78.21	88.46	88.46	89.74	85.90
MwA-S vs. MwA-C (MwA subgroups)	87.23	97.87	82.98	91.49	89.36	82.98	89.36

HC, healthy controls; MwA, migraine with aura; MwA-S, simple migraine with aura; MwA-C, complex migraine with aura; LR, linear regression; LDA, linear discriminant analysis; KNN, K-nearest neighbors; CART, decision tree; NB, Naive Bayes; SVM, support vectors machine; RF, random forest.

TABLE 3 Feature importance for MwA classification.

Cortex area	LDA		All algorithms	
	Rank	Importance (%)	Rank	Importance (%) Mean+SD
Left temporal pole thickness	1	1.47	4	0.94 ± 0.52
Right lingual gyrus thickness	2	1.23	1	1.31 ± 0.23
Left pars opercularis thickness	3	1.21	2	1.06 ± 0.56
Left temporal pole volume	4	1.17	3	0.98 ± 0.21
Left parahippocampal gyrus thickness	5	1.06	6	0.91 ± 0.12
Left caudal anterior cingulate thickness	6	0.97	5	0.92 ± 0.09
Left pars opercularis mean Gaussian curvature	7	0.94	8	0.84 ± 0.40
Left middle temporal gyrus thickness	8	0.93	9	0.46 ± 0.34
Left fusiform gyrus folding index	9	0.90	7	0.91 ± 0.25
Left inferior parietal folding index	10	0.86	10	0.28 ± 0.37

LDA, linear discriminant analysis; SD, standard deviation; Importance, feature importance calculated using the Extremely Randomized Trees algorithm (assigning the importance score to each feature which adds up to a value of 100); All algorithms/Importance, average feature importance using all algorithms and standard deviation (Linear Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Naive Bayes, Support Vectors Machine, and Random Forest).

opercularis are proposed as the two most important features for the simple and complex MwA classification (Figure 2). Figures 1, 2 highlight the cortical features that notably stand out compared to the other features based on their importance value for the LDA algorithm, which can be seen in Tables 1, 2 for the MwA and MwA subgroup classification. The thickness of the left temporal pole, right lingual gyrus, and left pars opercularis recognized as the most prominent markers for MwA classification had high importance levels that are greater than 1.2 (1.47, 1.23, and 1.21 respectively). The thickness of the left pericalcarine gyrus and left pars opercularis as the two most important features for the MwA subgroup classification had high importance levels that are greater than 0.9 (1.68 and 1.13 respectively).

4. Discussion

The focus of this study was finding new and advanced techniques for the detection of different phenotypes, which in turn, can allow better classification, marker validation and more optimized treatment of MwA patients. The main contribution of this work was finding an ML algorithm that performs the MwA vs. HC and MwA-S vs. MwA-C subgroups classification with high accuracy. This study showed that for both classification tasks the LDA algorithm has the best potential. In addition, sets of the most important features that contribute to accurate MwA detection and MwA subgroup classification were identified.

The development of neuroimaging techniques is in constant progress and multiple functional and anatomical imaging characteristics for aid in migraine diagnosis have been found, but research efforts are still being directed toward finding reliable biomarkers (11). The feature selection conducted in this research aims to contribute to the efforts in finding a marker that can improve the MwA diagnosis. Prior research concluded that the presence of aura is related to the regional distribution of cortical thickness and surface area abnormalities (45). Present results confirmed earlier findings on the existence of changes in the thickness of certain areas of the cortex in MwA patients (15, 46–48). This study also supports previous findings which indicate different thicknesses in the cerebral cortex when comparing MwA subgroups and demonstrate the cerebral cortex as a hallmark for the investigation of the complex MwA pathophysiology, where a further sub-phenotypes investigation is suggested (49). Moreover, some other features of the cortex, such as the folding index, may play a significant part in the differentiation between MwA-S and MwA-C subgroups (50).

Previous studies elaborate on the great potential of ML algorithms in determining brain aberrations that are specific to the migraine that could eventually be developed into computer-aided diagnostic tools based on MRI markers (12, 16, 17), which this study also investigates. Moreover, previous studies propose LDA, SVM and decision tree algorithms with 10-fold cross-validation to assess different types of migraine classification accuracy (13, 17), which were included and tested in this research. Studies that focused

TABLE 4 Feature importance for MWA subgroup classification.

Cortex area	LDA		All algorithms	
	Rank	Importance (%)	Rank	Importance (%) Mean±SD
Left pericalcarine thickness	1	1.68	1	1.95 ± 0.3
Left pars opercularis thickness	2	1.13	7	0.40 ± 0.51
Right parahippocampal gyrus thickness	3	0.83	9	0.37 ± 0.48
Right lingual gyrus surface area	4	0.82	5	0.45 ± 0.43
Right transverse temporal gyrus thickness	5	0.82	6	0.43 ± 0.41
Right paracentral lobule mean Gaussian curvature	6	0.80	8	0.39 ± 0.37
Right lateral occipital folding index	7	0.80	2	0.69 ± 0.1
Left parahippocampal gyrus folding index	8	0.75	4	0.53 ± 0.38
Left parahippocampal gyrus mean Gaussian curvature	9	0.75	10	0.19 ± 0.32
Right entorhinal volume	10	0.74	3	0.61 ± 0.42

LDA, linear discriminant analysis; SD, standard deviation; Importance, feature importance calculated using the Extremely Randomized Trees algorithm (assigning the importance score to each feature which adds up to a value of 100); All algorithms/Importance, average feature importance using all algorithms and standard deviation (Linear Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Naive Bayes, Support Vectors Machine, and Random Forest).

on distinguishing MwoA and MWA patients proposed the SVM algorithm and achieved 84% accuracy when using cerebral blood flow imaging markers (51), while 85% accuracy was reached when using electroencephalography data (52). In our research, the LDA algorithm outperformed other algorithms for both classification tasks, while KNN achieved the lowest accuracy. Also, it can be noted that the SVM algorithm resulted in noticeably lower accuracy compared to LDA. The study resulted in a promising accuracy of over 97% for both classification tasks. Also, the results indicate that most algorithms had better MWA subgroup classification accuracy. Previous studies using neural networks in migraine classification have achieved high accuracy (from 91 to 99%), however, these studies are predominantly based on the classification of MwoA and MWA patients (53–55). An approach that combines functional MRI data and inception module-based convolutional neural networks achieve up to 99% accuracy when discriminating HCs and migraine patients, as well as migraine subtypes - MwoA and MWA (54). Another study performs the classification of seven types of migraines and reaches 98% accuracy using an artificial neural network (55). However, these studies do not focus on HCs and MWA or MWA subtype classification problems, which is the main goal of this study. Current work employs a leave-one-out classification performance validation, without data scaling which confirms the applicability of the developed model for further clinical use with novel MRI data.

In our study, the LDA algorithm yielded the thickness of the left temporal pole, right lingual gyrus and left pars opercularis as the most prominent markers for MWA classification. The left temporal pole is already marked as an important cortical feature of migraineurs' brain in several previous studies (56–58), indicating that aberrant function, connectivity and structure of the temporal pole might contribute to clinical abnormalities in migraine patients (57). Furthermore, it is suggested that increased glucose metabolism in the left temporal pole compared to healthy individuals during olfactory stimuli might reflect the unique role of the temporal pole in odor hypersensitivity and odor-triggered

migraine (58) and thus can be considered as a potential target for treatment (59). Also, previous studies demonstrated an important role of the extrastriate visual cortex, including the lingual gyrus, in the MWA pathophysiology and point out that mitochondrial dysfunction might be only present in MWA relative to MwoA and HCs (60). Moreover, it was reported that glutamate levels were increased while gamma-aminobutyric acid (GABA) levels were decreased in the visual cortex in MWA patients, suggesting disturbances in the cortical excitatory-inhibitory balance, which could predispose the cortex to CSD and aura (61, 62). Another study compared healthy controls with MWA patients and found results that indicate significant differences in thickness of several brain areas between HC and S-MWA and between HC and C-MWA (15). The study reported abnormal thickness of MWA patients compared to HCs in the right high-level visual-information-processing areas, including the lingual gyrus, which is also confirmed by our results.

Regarding the simple and complex MWA classification, the thickness of the left pericalcarine gyrus and left pars opercularis cortex are proposed as the two most important features, although their role in the complexity of aura manifestation is unknown and could not be marked as specific features for the complexity of MWA because changes in the function and structure of the left pericalcarine gyrus and left pars opercularis cortex are also noted in the MwoA (63, 64), although thicker cortex of the left pericalcarine gyrus is demonstrated in MWA-C patients (50). Increased cortical thickness of the calcarine area of the left hemisphere was discovered in both MWA subtypes in contrast to HCs, while no significant differences emerged among MWA-S and MWA-C (15). Anyhow, given that all cortex features had a low percentage of importance in the classification, both for MWA classification and subclassification, but accuracy was very high, it can suggest that MWA is not a disease of one brain region yet a functional disease of the neural network with multiple structural changes of the cerebral cortex. Finally, it should be kept in mind that this is the first study, to the best of our knowledge, that tries to

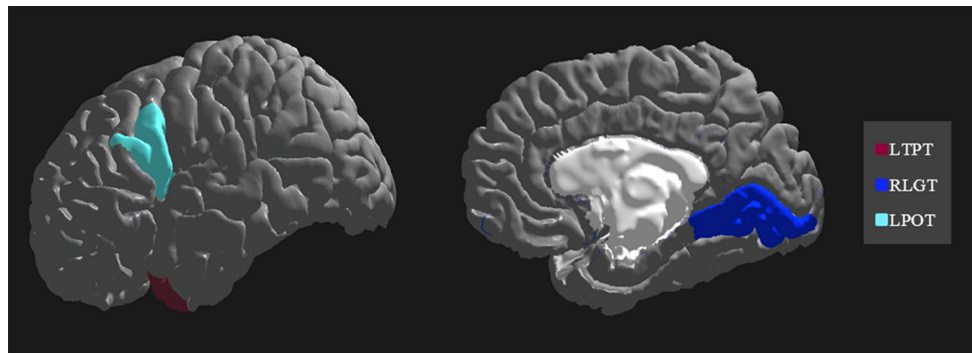


FIGURE 1

Top features for migraine with aura classification using the Linear Discriminant Analysis algorithm (LTPT, left temporal pole thickness; RLGT, right lingual gyrus thickness; LPOT, left pars opercularis thickness).

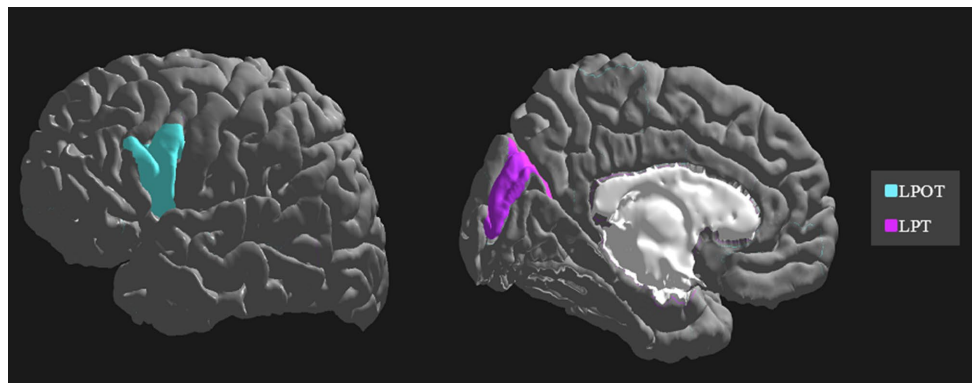


FIGURE 2

Top features for migraine with aura subgroup classification using the Linear Discriminant Analysis algorithm (LPOT, left pars opercularis thickness; LPT, left pericalcarine thickness).

classify MwA patients and HCs, as well as MwA-S and MwA-C subgroups, using several cortex features derived from structural neuroimaging by state-of-the-art post-processing techniques. Hence, our results should be validated in future multicentric studies where a larger sample of MwA patients will yield more solid conclusions. Also, this is the first step in classifying the complexity of MwA based on morphometric MRI data and future work should include a finer classification based on the Migraine Aura Complexity Score (MACS) (49).

A possible limitation of the current study is the lack of prospectively collected information about the lateralization of MwA attacks which could potentially play an influencing factor in neuroimaging studies. However, throughout the medical history of selected MwA patients, there was no report suggesting that the patient had clear lateralization of visual or somatosensory symptoms. Moreover, future studies should include a group of patients who have only migraines without aura to investigate specific morphometric features from neuroimages related only to MwA-S and MwA-C subgroups. In addition, new knowledge and better accuracy might be obtained by combining fMRI and structural MRI modalities (12). ML models based on the combined data could be implemented and tested in future studies.

5. Conclusion

This study shows that high accuracy can be reached when using the LDA algorithm for classifying individuals into HCs and MwA patients, as well as for MwA subtype classification. Furthermore, results indicate the existence of an abnormality in MwA patient thickness, surface area, volume, mean Gaussian curvature and/or folding index of particular cortex areas concerning HCs, as well as when comparing MwA-S and MwA-C patients. Also, this method shows significant potential in the validation of MwA diagnosis and subtype classification, which can tackle and challenge the current treatments of MwA.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of the Neurology Clinic,

Clinical Center of Serbia. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KM, IP, and AS conceptualized and designed the study. IP and AR performed the data acquisition. All authors contributed to the data analysis and interpretation. KM, IP, and AS drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Funding

This research was partly supported by Ministry of education, science and technological development, Republic of Serbia (contract number for IP and MD: 451-03-68/2022-14/200146, contract number for AS: 451-03-68/2022-14/200103).

Acknowledgments

The authors would like to acknowledge the study program Intelligent Systems of the Multidisciplinary PhD studies at the

University of Belgrade for providing the framework for this collaboration. KM is affiliated with the study program Intelligent Systems as a PhD student, and AS as a lecturer and PhD thesis mentor.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer is currently organizing a research topic with the author IP.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Amiri P, Kazeminasab S, Nejadghaderi SA, Mohammadinasab R, Pourfathi H, Araj-Khodaei M, et al. Migraine: a review on its history, global epidemiology, risk factors, and comorbidities. *Front Neurol.* (2022) 12:800605. doi: 10.3389/fneur.2021.800605
- Goadsby PJ. *Pain: the person, the science, the clinical interface.* Melbourne: BPA Print Group (2015). 109 p.
- Steiner TJ, Stovner LJ, Jensen R, Uluduz D, Katsarava Z. Migraine remains second among the world's causes of disability, and first among young women: findings from GBD2019. *J Headache Pain.* (2020) 21:137. doi: 10.1186/s10194-020-01208-0
- Safiri S, Pourfathi H, Eagan A, Mansournia MA, Khodayari MT, Sullman MJ, et al. Global, regional, and national burden of migraine in 204 countries and territories, 1990 to 2019. *Pain.* (2022) 163:e293–309. doi: 10.1097/j.pain.0000000000002275
- Headache Classification Committee of the International Headache Society (IHS). The international classification of headache disorders. *Cephalalgia.* (2018) 38:1–211. doi: 10.1177/0333102417738202
- Hansen JM, Charles A. Differences in treatment response between migraine with aura and migraine without aura: lessons from clinical practice and RCTs. *J Headache Pain.* (2019) 20:96. doi: 10.1186/s10194-019-1046-4
- Kincses ZT, Veréb D, Faragó P, Tóth E, Kocsis K, Kincses B, et al. Are migraine with and without aura really different entities? *Front Neurol.* (2019) 10:982. doi: 10.3389/fneur.2019.00982
- Rasmussen BK, Olesen J. Migraine with aura and migraine without aura: an epidemiological study. *Cephalalgia.* (1992) 12:221–8. doi: 10.1046/j.1468-2982.1992.1204221.x
- Hansen JM, Goadsby PJ, Charles AC. Variability of clinical features in attacks of migraine with aura. *Cephalalgia.* (2016) 36:216–24. doi: 10.1177/0333102415584601
- Schwedt TJ, Chong CD. Functional imaging and migraine: new connections? *Curr Opin Neurol.* (2015) 28:265. doi: 10.1097/WCO.0000000000000194
- Cutrer FM, Black DF. Imaging findings of migraine. Headache: the journal of head and face. *Pain.* (2006) 46:1095–107. doi: 10.1111/j.1526-4610.2006.00503.x
- Rocca MA, Harrer JU, Filippi M. Are machine learning approaches the future to study patients with migraine? *Neurology.* (2020) 94:291–2. doi: 10.1212/WNL.00000000000008956
- 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) 37:828–44. doi: 10.1177/0333102416652091
- Petrusic I, Dakovic M, Zidverc-Trajkovic J. Subcortical volume changes in migraine with Aura. *J Clin Neurol.* (2019) 15:e34. doi: 10.3988/jcn.2019.15.4.448
- Abagnale C, di Renzo A, Sebastianelli G, Casillo F, Tinelli E, Giuliani G, et al. Whole brain surface-based morphometry and tract-based spatial statistics in migraine with aura patients: difference between pure visual and complex auras. *Front Hum Neurosci.* (2023) 17:1146302. doi: 10.3389/fnhum.2023.1146302
- Messina R, Filippi M. What we gain from machine learning studies in headache patients. *Front Neurol.* (2020) 11:221. doi: 10.3389/fneur.2020.00221
- Schwedt TJ, Chong CD, Wu T, Gaw N, Fu Y, Li J. Accurate classification of chronic migraine via brain magnetic resonance imaging. *Headache.* (2015) 55:762–77. doi: 10.1111/head.12584
- Fischl B. FreeSurfer. *NeuroImage.* (2012) 62:774–81. doi: 10.1016/j.neuroimage.2012.01.021
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage.* (2006) 31:968–80. doi: 10.1016/j.neuroimage.2006.01.021
- Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. *J. Big Data.* (2020) 7:1–26. doi: 10.1186/s40537-020-00327-4
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* (2006) 63:3–42. doi: 10.1007/s10994-006-6226-1
- Hua J, Xiong Z, Lowey J, Suh E, Dougherty ER. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics.* (2005) 21:1509–15. doi: 10.1093/bioinformatics/bti171
- Somol P, Novovicová J, Pudil P. Efficient feature subset selection and subset size optimization. *Pattern Recogn Recent Adv.* (2010) 1:75–98. doi: 10.5772/9356
- Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing.* (2020) 20:295–316. doi: 10.1016/j.neucom.2020.07.061
- Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn.* (2015) 48:2839–46. doi: 10.1016/j.patcog.2015.03.009
- Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol.* (2020) 122:56–69. doi: 10.1016/j.jclinepi.2020.03.002
- Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied logistic regression.* Hoboken: John Wiley & Sons (2013). 7 p.
- Owen AB. A robust hybrid of lasso and ridge regression In: JS Verducci, X Shen and J Lafferty, editors. *Prediction and discovery. Contemporary mathematics* (2007). American Mathematical Society. 59–72.
- Xanthopoulos P, Pardalos PM, Trafalis TB. Linear discriminant analysis. *Robust Data Mining.* (2013) 1:27–33. doi: 10.1007/978-1-4419-9878-1_4
- Sharma A, Paliwal KK. Linear discriminant analysis for the small sample size problem: an overview. *Int J Mach Learn Cybern.* (2015) 6:443–54. doi: 10.1007/s13042-013-0226-9

31. Ghojogh B, Crowley M. Linear and quadratic discriminant analysis: tutorial. (2019). Available at: <https://arxiv.org/abs/1906.02590> (Accessed May 15, 2023)
32. Wall ME, Dyck PA, Brettin TS. SVDMAN - singular value decomposition analysis of microarray data. *Bioinformatics*. (2001) 17:566–8. doi: 10.1093/bioinformatics/17.6.566
33. Golub GH, Van Loan CF. Matrix computations. *Matrix Comput.* (2013):76. doi: 10.56021/9781421407944
34. Ledoit O, Wolf M. Honey, I shrunk the sample covariance matrix. *UPF Econ Bus Work Paper*. (2003):691. doi: 10.2139/ssrn.433840
35. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. (1967) 13:21–7. doi: 10.1109/IT.1967.1053964
36. Danielsson PE. Euclidean distance mapping. *Comput Graphics Image Process*. (1980) 14:227–48. doi: 10.1016/0146-664X(80)90054-4
37. Breiman L. *Classification and regression trees*. Wadsworth International Group. 1st ed Belmont, California, USA: Routledge (1984).
38. Breiman L. Technical note: some properties of splitting criteria. *Mach Learn*. (1996) 24:41–7. doi: 10.1007/BF00117831
39. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philos Trans R Soc Lond*. (1763) 31:370–418. doi: 10.1098/rstl.1763.0053
40. Murty MN, Devi VS. Pattern recognition: an algorithmic approach. *Pattern Recogn*. (2011):93. doi: 10.1007/978-0-85729-495-1
41. Lou W, Wang X, Chen F, Chen Y, Jiang B, Zhang H. Sequence based prediction of DNA-binding proteins based on hybrid feature selection using random forest and Gaussian naive Bayes. *PLoS One*. (2014) 9:e86703. doi: 10.1371/journal.pone.0086703
42. Cortes C, Vapnik V. Support-vector networks. *Machine Learn*. (1995) 20:273–97. doi: 10.1007/BF00994018
43. Vapnik V. *The nature of statistical learning theory*. New York, USA: Springer Science & Business Media (1999). 144–146.
44. James G, Witten D, Hastie T, Tibshirani R. *Statistical learning. An introduction to statistical learning: with applications in R (2nd)*. New York, USA: Springer Texts (2021). 343–345
45. Messina R, Rocca MA, Colombo B, Valsasina P, Horsfield MA, Copetti M, et al. Cortical abnormalities in patients with migraine: a surface-based analysis. *Radiology*. (2013) 268:170–80. doi: 10.1148/radiol.13122004
46. DaSilva AF, Granziera C, Snyder J, Hadjikhani N. Thickening in the somatosensory cortex of patients with migraine. *Neurology*. (2007) 69:1990–5. doi: 10.1212/01.wnl.0000291618.32247.2d
47. Granziera C, DaSilva AF, Snyder J, Tuch DS, Hadjikhani N. Anatomical alterations of the visual motion processing network in migraine with and without aura. *PLoS Med*. (2006) 3:e402. doi: 10.1371/journal.pmed.0030402
48. Petrusic I, Viana M, Dakovic M, Goadsby P J, Zidverc-Trajkovic J. Proposal for a migraine aura complexity score. *Cephalalgia*. (2019) 39:732–41. doi: 10.1177/0333102418815487
49. Petrusic I, Viana M, Dakovic M, Zidverc-Trajkovic J. Application of the migraine Aura complexity score (MACS): clinical and neuroimaging study. *Front Neurol*. (2019) 10:1112. doi: 10.3389/fneur.2019.01112
50. Petrusic I, Dakovic M, Kacar K, Zidverc-Trajkovic J. Migraine with aura: surface-based analysis of the cerebral cortex with magnetic resonance imaging. *Korean J Radiol*. (2018) 19:767–76. doi: 10.3348/kjr.2018.19.4.767
51. Fu T, Liu L, Huang X, Zhang D, Gao Y, Yin X, et al. Cerebral blood flow alterations in migraine patients with and without aura: an arterial spin labeling study. *J Headache Pain*. (2022) 23:131–9. doi: 10.1186/s10194-022-01501-0
52. 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) 48:403–12. doi: 10.1007/s10439-019-02357-3
53. De la Hoz R, Rúa Ascar Z, Manuel J. (2014) Análisis de modelos de redes neuronales artificiales, para un sistema de diagnósticos de migrañas con aura y sin aura. Master's thesis. Universidad Simón Bolívar
54. Yang H, Zhang J, Liu Q, Wang Y. Multimodal MRI-based classification of migraine: using deep learning convolutional neural network. *Biomed Eng Online*. (2018) 17:1–4. doi: 10.1186/s12938-018-0587-0
55. Sanchez-Sanchez PA, García-González JR, Ascar JM. Automatic migraine classification using artificial neural networks. *F1000Research*. (2020):9. doi: 10.12688/f1000research.23181.2
56. Coppola G, di Renzo A, Tinelli E, Iacovelli E, Lepre C, di Lorenzo C, et al. Evidence for brain morphometric changes during the migraine cycle: a magnetic resonance-based morphometry study. *Cephalalgia*. (2015) 35:783–91. doi: 10.1177/0333102414559732
57. Moulton EA, Becerra L, Maleki N, Pendse G, Tully S, Hargreaves R, et al. Painful heat reveals hyperexcitability of the temporal pole in interictal and ictal migraine states. *Cereb Cortex*. (2011) 21:435–48. doi: 10.1093/cercor/bhq109
58. Demarquay G, Royet JP, Mick G, Ryvlin P. Olfactory hypersensitivity in migraineurs: a H(2)(15)O-PET study. *Cephalalgia*. (2008) 28:1069–80. doi: 10.1111/j.1468-2982.2008.01672.x
59. Cortese F, Pierelli F, Bove I, di Lorenzo C, Evangelista M, Perrotta A, et al. Anodal transcranial direct current stimulation over the left temporal pole restores normal visual evoked potential habituation in interictal migraineurs. *J Headache Pain*. (2017) 18:70–9. doi: 10.1186/s10194-017-0778-2
60. Sándor PS, Dydak U, Schoenen J, Kollias SS, Hess K, Boesiger P, et al. MR-spectroscopic imaging during visual stimulation in subgroups of migraine with aura. *Cephalalgia*. (2005) 25:507–18. doi: 10.1111/j.1468-2982.2005.00900.x
61. Zielman R, Wijnen JP, Webb A, Onderwater GL, Ronen I, Ferrari MD, et al. Cortical glutamate in migraine. *Brain*. (2017) 140:1859–71. doi: 10.1093/brain/awx130
62. Bridge H, Stagg CJ, Near J, Lau CI, Zisner A, Cader MZ. Altered neurochemical coupling in the occipital cortex in migraine with visual aura. *Cephalalgia*. (2015) 35:1025–30. doi: 10.1177/0333102414566860
63. Wei HL, Zhou X, Chen YC, Yu YS, Guo X, Zhou GP, et al. Impaired intrinsic functional connectivity between the thalamus and visual cortex in migraine without aura. *J Headache Pain*. (2019) 20:116–9. doi: 10.1186/s10194-019-1065-1
64. Amin FM, de Icco R, al-Karagholi MAM, Raghava J, Wolfram F, Larsson HB, et al. Investigation of cortical thickness and volume during spontaneous attacks of migraine without aura: a 3-tesla MRI study. *J Headache Pain*. (2021) 22:98–9. doi: 10.1186/s10194-021-01312-9