EREDETI KÖZLEMÉNY

ORIGINAL ARTICLE

MRI radiomics based machine learning model of the periaqueductal gray matter in migraine patients

Ismail MESE[1], Rahsan KARACI[2], Ceylan ALTINTAS TASLICAY[3], Cengizhan TASLICAY[4], Gur AKANSEL[5], Saime Fusun DOMAC[2]

[1]Department of Radiology, Health Sciences University, Erenkoy Mental Health and Neurology Training and Research Hospital, Istanbul, Turkey
[2]Department of Neurology, Health Sciences University, Erenkoy Mental Health and Neurology Training and Research Hospital, Istanbul, Turkey
[3]Department of Radiology, Ministry of Health Fethi Sekin City Hospital, Elazig, Turkey
[4]Department of Metallurgical and Materials Engineering, Istanbul Technical University, Istanbul, Turkey
[5]Department of Radiology, Kocaeli University School of Medicine, Kocaeli, Turkey

| English | https://doi.org/10.18071/isz.77.0039 | www.elitmed.hu

Correspondent:
Ismail MESE, MD,
Department of Radiology,
Health Sciences University,
Erenkoy Mental Health and Neurology Training and Research Hospital,
19 Mayıs, Sinan Ercan Cd. No:23, 34736 Kadıköy, Istanbul
Phone: +90 (216) 302 59 59
Fax:+90 (216) 356 04 96,
mobil: +90 (506) 895 16 53,
e-mail: ismail_mese@yahoo.com
https://orcid.org/0000-0002-4429-6996

Érkezett:
2023. március 12.
Elfogadva:
2023. augusztus 18.

Background and purpose – The aim of the study was to investigate the question: Can MRI radiomics analysis of the periaqueductal gray region elucidate the pathophysiological mechanisms underlying various migraine subtypes, and can a machine learning model using these radiomics features accurately differentiate between migraine patients and healthy individuals, as well as between migraine subtypes, including atypical cases with overlapping symptoms?

Methods – The study analyzed initial MRI images of individuals taken after their first migraine diagnosis, and additional MRI scans were acquired from healthy subjects. Radiomics modeling was applied to analyze all the MRI images in the periaqueductal gray region. The dataset was randomized, and oversampling was used if there was class imbalance between groups. The optimal algorithm-based feature selection method was employed to select the most important 5-10 features to differentiate between the two groups. The classification performance of AI algorithms was evaluated using receiver operating characteristic analysis to calculate the area under the curve, classification accuracy, sensitivity, and specificity values. Participants were required to have a confirmed diagnosis of either episodic migraine, probable migraine, or chronic migraine. Patients with aura, those who used migraine-

Migrénes betegek periaqueductalis szürkeállományának MRI radiomikai alapú gépitanulás-modellje

Mese I, MD; Karaci R, MD; Altintas Taslicay C, MD; Taslicay C, Akansel G, Professor; Domac SF, Professor

Háttér és cél – A tanulmány célja a következő kérdés vizsgálata volt: A periaqueductalis szürke régió MRI radiomikai elemzése képes-e megvilágítani a különböző migrénaltípusok hátterében álló patofiziológiai mechanizmusokat, és képes-e az ezeket a radiomikai jellemzőket használó gépi tanulási modell a migrénes betegek és az egészséges egyének pontos megkülönböztetésére, valamint a migrén altípusait, beleértve az átfedő tünetekkel járó atipikus eseteket is?

Módszerek – A vizsgálatban a betegek első migréndiagnózisa után készült kezdeti MRI-felvételeket elemeztük, és további MRI-felvételeket szereztünk be egészséges alanyokról. Radiomikai modellezést alkalmaztunk a periaqueductalis szürke régió összes MRI-felvételének elemzésére. Az adathalmazt randomizáltuk, és túlmintavételt alkalmaztunk, ha a csoportok között osztály-kiegyensúlyozatlanság állt fenn. Az optimális algoritmuson alapuló jellemzőkiválasztási módszert alkalmaztuk a két csoport megkülönböztetésére szolgáló legfontosabb 5-10 jellemző kiválasztására. A mesterséges intelligencia algoritmusok osztályozási teljesítményét a receiver operating characteristic (ROC) vevő működési karakterisztika analízis segítségével értékeltük a görbe alatti terület, az osztályozási pontosság, a szenzitivitás és a specificitás értékeinek kiszámításához. A résztvevőknek rendelkezniük kellett az

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

preventive medication within the past six months, or had chronic illnesses, psychiatric disorders, cerebrovascular conditions, neoplastic diseases, or other headache types were excluded from the study. Additionally, 102 healthy subjects who met the inclusion and exclusion criteria were included.
Results – The algorithm-based information gain method for feature reduction had the best performance among all methods, with the first-order, gray-level size zone matrix, and gray-level co-occurrence matrix classes being the dominant feature classes. The machine learning model correctly classified 82.4% of migraine patients from healthy subjects. Within the migraine group, 74.1% of the episodic migraine-probable migraine patients and 90.5% of the chronic migraine patients were accurately classified. No significant difference was found between probable migraine and episodic migraine patients in terms of the periaqueductal gray region radiomics features. The kNN algorithm showed the best performance for classifying episodic migraine-probable migraine subtypes, while the Random Forest algorithm demonstrated the best performance for classifying the migraine group and chronic migraine subtype.
Conclusion – A radiomics-based machine learning model, utilizing standard MR images obtained during the diagnosis and follow-up of migraine patients, shows promise not only in aiding migraine diagnosis and classification for clinical approach, but also in understanding the neurological mechanisms underlying migraines.

Keywords: migraine, machine learning, radiomics, headache

epizodikus migrén, a valószínűsíthető migrén vagy a krónikus migrén igazolt diagnózisával. A vizsgálatból kizártuk az aurás betegeket, azokat, akik az elmúlt hat hónapban migrénmegelőző gyógyszert szedtek, vagy krónikus betegségben, pszichiátriai rendellenességben, cerebrovascularis betegségben, daganatos betegségben vagy más típusú fejfájásban szenvedtek. Emellett 102 egészséges személyt vontunk be a vizsgálatba, akik megfeleltek a beválasztási és kizárási kritériumoknak.
Eredmények – Az algoritmusalapú információnyerési módszer a jellemzők csökkentésére nyújtotta az összes módszer közül a legjobb teljesítményt, az elsőrendű, a szürke szintű méretzóna mátrix és a szürke szintű együttes mátrix osztályok voltak a domináns jellemzőosztályok. A gépi tanulási modell helyesen különítette el a migrénes betegek 82,4%-át az egészséges alanyoktól. A migrénes csoporton belül az epizodikus migrénes – valószínű migrénes betegek 74,1%-át és a krónikus migrénes betegek 90,5%-át sikerült pontosan osztályozni. A valószínű migrénes és az epizodikus migrénes betegek között nem találtunk szignifikáns különbséget a periaqueductalis szürke régió radiomikai jellemzői tekintetében. A kNN algoritmus mutatta a legjobb teljesítményt az epizodikus migrén – valószínű migrén altípusok meghatározásában, míg a Random Forest algoritmus mutatta a legjobb teljesítményt a migréncsoport és a krónikus migrén altípus meghatározásában.
Következtetés – A migrénes betegek diagnózisa és nyomon követése során nyert standard MR-felvételeket felhasználó radiomikai alapú gépi tanulási modell ígéretesnek bizonyul nemcsak a klinikai megközelítés számára a migrén diagnózisának és osztályozásának segítésében, hanem a migrén hátterében álló neurológiai mechanizmusok megértésében is.

Kulcsszavak: migrén, gépi tanulás, radiomika, fejfájás

Migraine is a widespread multifactorial neurovascular syndrome, mainly characterized by episodes of unilateral headache, with or without aura. Affecting 15.3% of the population, migraine has a higher prevalence in females (20.7%) compared to males (7%) annually[1]. It is ranked as the second leading cause of disability worldwide[2]. Although migraine is traditionally diagnosed clinically, its clinical expression is highly variable[3, 4], and its symptoms can be nonspecific, often encountered in several primary and secondary causes of headache[5]. Common symptoms include throbbing or pulsating headache, sensitivity to light and sound, nausea, vomiting, and visual disturbances known as aura[3–5]. However, the complexity arises from the fact that migraine shares clinical features and overlapping syndromes with various other disorders. For instance, multiple sclerosis[6, 7] may involve vision problems, muscle weakness, and balance issues; acute infarction[8, 9] can present with sudden head-

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

ache, dizziness, and confusion; tension-type headache[10, 11] is characterized by a dull, constant pain on both sides of the head; and epilepsy[12] can cause seizures, loss of consciousness, and temporary confusion. Such similarities can lead to misdiagnosis or delayed diagnosis, resulting in inadequate or inappropriate treatment. A recent systematic review and evidence-based guideline from the American Headache Society recommends using neuroimaging studies in cases where clinical findings are atypical for migraine and symptoms overlap with other disorders[13]. Conventional neuroimaging can be useful in differentiating between primary headaches, such as migraines, and secondary headaches but not in differentiating various types of primary headaches.

The periaqueductal gray matter (PAG) is a critical brain structure involved in the regulation of various physiological functions, such as pain modulation and autonomic control. Moreover, the PAG has extensive connections with other pain-related brain regions, such as the hypothalamus and the trigeminal nucleus caudalis[14]. These connections suggest that the PAG may play a significant role in migraine pathophysiology. Additionally, its known involvement in the central sensitization process supports its role in migraine chronification, which contributes to the transition from episodic to chronic migraine (CM)[15]. Central sensitization is characterized by a heightened response of the nociceptive system to sensory stimuli, potentially leading to an increased frequency and intensity of migraine attacks over time[14, 15].

Recent advances in technology have allowed for the collection of large amounts of quantitative data from medical imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT). Through the use of artificial intelligence (AI) algorithms to analyze these data sets, a radiomics-based machine learning model can be created to identify and diagnose pathologies that may not be visible through traditional imaging methods[16–19]. Texture analysis is a method for quantifying patterns in images, similar to radiomics, but it is a less advanced technology in terms of data digging from images than radiomics[20]. In their study, *Zhiye Chen* et al.[21] identified altered MR image texture characteristics of the PAG in episodic migraine (EM) patients compared to healthy subjects, suggesting that these texture characteristics could be considered as imaging biomarkers for EM. Building on this finding, the aim of this study is to investigate the potential of MRI radiomics analysis in evaluating PAG to elucidate the pathophysiological mechanisms underlying migraine and its subtypes, such as EM, probable migraine (PM), and CM. We hypothesize that all migraine patients, including the probable subtype, who lack T2-visible lesions, exhibit MRI radiomics changes in PAG, and that changes are unique for each subtype. A machine learning model can be developed using these radiomics features to differentiate between migraine patients and healthy individuals. Additionally, the model can distinguish each specific subtype not only from healthy subjects but also from other subtypes. Moreover, our study includes atypical migraine cases that show overlapping symptoms with other headache and neurological diseases, which may not be easily detected using traditional questionnaires. This approach aims to develop a more comprehensive diagnostic tool for migraine, ultimately impacting future research and clinical practice by enhancing diagnostic accuracy and early intervention for migraines. This could also lead to personalized treatments for each subtype and improved understanding of migraine pathophysiology.

Methods

Ethic approval

Approval for this study was obtained from the Research Ethics Board of Erenkoy Mental Health and Neurology Training and Research Hospital (Ethics board decision number 2021;03). The board waived the need for written informed consent. We conducted a retrospective review of clinical data and imaging records for migraine patients who received follow-up care at Erenkoy Mental Health and Neurology Training and Research Hospital's Neurology outpatient clinic between January 2020 and February 2022.

Participants

Based on the International Classification of Headache Disorders, 3rd edition (ICHD-3) criteria[22], patients were initially categorized into three groups: EM, PM, and CM. To qualify for inclusion in one of these migraine groups, patients needed to be at least 18 years old and under the care of a neurology department for headache management. They were required to have a confirmed diagnosis of either EM, PM, or CM per ICHD-3 classification, along with initial cranial MRI images in the Picture Archiving and Communication System (PACS) that met technical requirements for evaluating structural changes. Additionally, MRI images had to be devoid of any pathology in the periaqueductal region or other areas. Since diagnosing migraine with aura can be relatively straightforward for healthcare professionals familiar with the condition, particularly when patients present with a clear history and characteristic symptoms, patients with migraine with typical aura were considered out of scope for this study. Additionally patients could not have used migraine-preventive medication within the past six months and had to be free of chronic illnesses, psychiatric disorders, cerebrovascular conditions, neoplastic diseases, or other headache types. Patients could also not have a history of alcohol, nicotine, or substance abuse.

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

Healthy volunteers who were admitted to the neurology outpatient clinic were examined and confirmed to have no neurological or other diseases through physical examination, prior imaging, and history. We obtained cranial MRI images from these volunteers, and to be included as a healthy subject, they must have never had any primary headache disorders or other types of headaches in the past year. Their MRI images had to be technically eligible for evaluation of structural changes and free of any pathology in the periaqueductal area. They could not have chronic disorders, psychiatric diseases, cerebrovascular disorders, neoplastic diseases, or a history of alcohol, nicotine, or substance abuse. The exclusion criteria for both groups included a history of head or major body trauma, regular use of psychoactive or hormonal medication, lack of medical information, and non-diagnostic MRI studies for radiologic evaluation. We collected and evaluated general demographic and headache information from both groups.

Tracking changes over time

We retrospectively tracked migraine patients in PM subgroup for over 2 years, using examination notes and headache diaries collected every 3-6 months. We aimed to gain a deeper understanding of the natural progression of the disorder. Additionally, we sought to exclude other disorders that mimic migraine within the probable migraine subtype initially.

One hundred twenty-two patients with migraines met our inclusion criteria. We excluded 11 patients due to their diagnoses changing over time PM to other diagnoses. In addıtıon, four patients were excluded because their images were not suitable for radiologic evaluation. We also excluded one patient who had a history of head trauma, one patient who regularly used psychoactive or hormonal medication, and three patients who lacked medical information. Definitively, among the remaining 102 migraine patients, 56 had EM, 25 had PM, and 21 had CM. The study also included 102 healthy subjects who met our inclusion and exclusion criteria **(Figure 1)**.

Procedure

We utilized the initial MRI images of individuals taken after their first diagnosis of migraine. All MRI examinations were performed using a 1.5 Tesla MRI scanner (Magnetom Avanto TIM) at our institution. T1-weighted (T1W) images were acquired with an echo time (TE) range of 12-14 milliseconds and repetition time (TR) range of 400-600 milliseconds, with a slice thickness of 5-6 millimeters, an interslice gap of 0.5-1 millimeter, and a field of view (FOV) of 200-230 millimeters.

Radiomics modeling

Reproducibility and generalizability issues
To maintain reproducibility and generalizability in radiomics modeling, we followed previously defined distinct guidelines[23–25]. We applied their suggestions as much as possible to avoid common mistakes. We also made an attempt to avoid information leakage that might occur during any step of the pipeline, including early image preprocessing. The data split was performed correctly.

In our study, we adhered to strict guidelines for sample size determination based on existing radiomics research[23]. The literature recommended a sample size of at least five to ten times the minimum number of selected features[23, 26]. Following this guideline, we initially began with smaller sample sizes and we used a sample size of 50, which is at least five times the number of selected features. Consequently, we continued the study with a sample size that provided sufficient power for identifying specific patterns using radiomics analysis. This sample size was more feasible in terms of time and resource constraints compared to analyzing a larger cohort, while still encompassing a diverse range of disease characteristics.

Preprocessing steps - standardization, bias field correction filter
Axial T1-weighted images were uploaded to 3D Slicer 4.10.2[27] in DICOM (Digital Imaging and Communications in Medicine) format. The uploaded images were

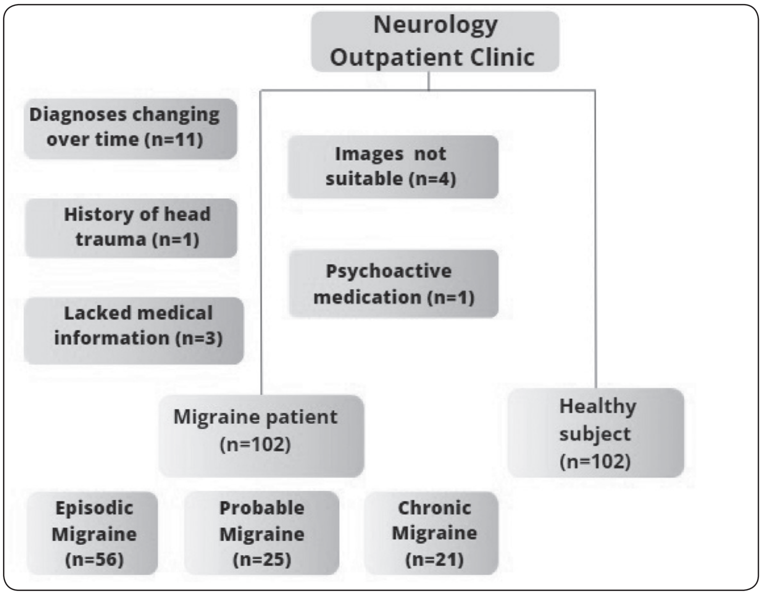

Figure 1. *Ineligible participants, final tally of migraine sufferers and subtype classifications*

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

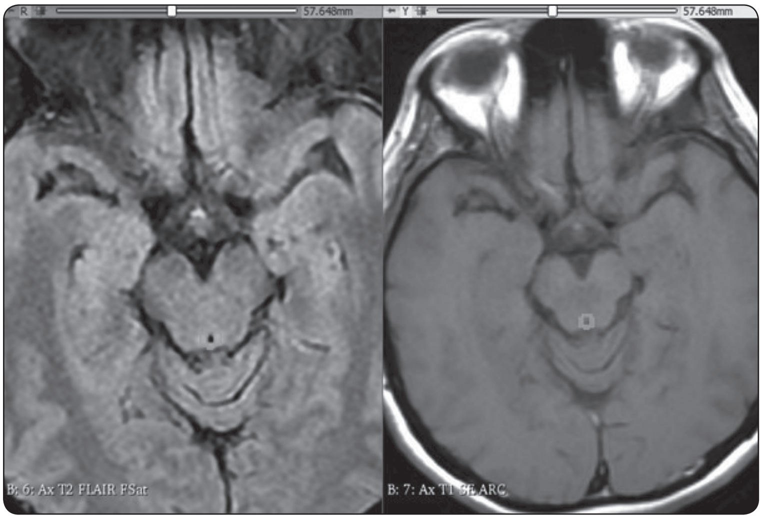

Figure 2. *Manual segmentation of periaquaductal gray matter in 3D slicer software*

standardized using the N4ITK bias field correction filter and ±3 sigma normalization.

PAG segmentation, discretization, pixel size resampling, and feature extraction

PAG segmentation was performed manually by a single reader with ten years of experience **(Figure 2)**. The segmentation was carried out to include the entire 3-dimensional volume. In case of any lack of certainty, additional sequences were loaded into the software to make the segmentation area clear. In cases where clear segmentation could not be done even with additional sequences, it was supervised by a senior reader with 25 years of experience. Voxel sizes were also resampled to 1x1x1 mm³. We used a fixed bin width value of 25 to ensure adequate discretization in gray levels. This value was determined by the average number of gray-scale ranges obtained by histogram analysis after segmentation in filtered images. Fourteen shape-based semantic features and 93 textural features, including 18 first-order, 24 Gray-Level Co-Occurrence Matrix, 16 Gray-Level Run Length Matrix, 16 Gray-Level Size Zone Matrix, 14 Gray-Level Dependence Matrix, and 5 Neighboring Gray Tone Difference Matrix features were extracted from each segmented volume using the radiomics extension of the software.

Post-processing steps - randomization of the data set, oversampling, feature reduction

Before creating the model, randomization was performed to ensure the homogenization of the dataset. If there was class imbalance between groups, the SMOTE (Synthetic minority oversampling technique) method was planned to be used to increase the size of the minority group. We then studied algorithm-based feature selection methods in our work. Different algorithm-based feature reduction methods were compared according to their helpfulness in the diagnosis procedure. Using the optimal algorithm-based method as a feature reduction method, a classification was made by selecting the most important 5-10 features to differentiate between the two groups.

Post-processing steps - creating model with reduced features

For AI algorithm-based classifications, we used Orange data mining software version 3.24 (https://orange.biolab.si/)[28]. The Support Vector Machine (SVM), logistic regression, k-Nearest Neighbor (kNN), Naive Bayes, decision tree, Random forest, CN2 rule indicer, and neural network AI algorithms were present in this software. To determine the model with the highest diagnostic value in differentiating between the migraine and healthy subjects, we performed a classification using reduced radiomics features obtained from the T1W sequence with various AI algorithms. We used stratified 10-fold cross-validation as an internal validation technique in this comparison-classification. We also developed additional models to differentiate between subtypes and healthy subjects, as well as between the various subtypes themselves. We determined a common AI algorithm that simultaneously provides high values for area-under-curve (AUC), sensitivity, and specificity based on these models. The performance of the algorithms was reported using multiple metrics, including AUC, accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. Confusion matrices were presented for classification tasks, and class imbalance was addressed by including related metrics such as F1-measure. **Figure 3** summarizes the steps involved in creating a model.

Statistical analysis

The normality of radiomic features was determined using the Shapiro-Wilk test. Numeric variables, such as age were assessed using a one-way ANOVA test, while categorical variables such as gender were assessed using the Chi-square test (IBM SPSS 20.0 software package). A p-value of less than 0.05 was considered statistically significant.

We evaluated the classification performance of AI algorithms using receiver operating characteristic (ROC) analysis to calculate the area under the curve (AUC), classification accuracy, sensitivity, and specificity values. We used these metrics to evaluate and compare the effectiveness of various AI algorithms in classifying different types of migraines based on the available data.

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

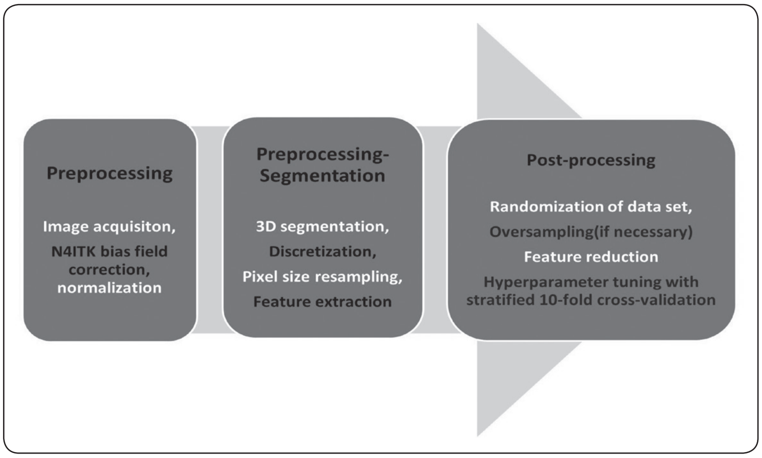

Figure 3. *The steps of creating a model*

Table 1. *Comparison of different feature reduction methods; best possible AUC, sensivity and specity values with different AI algorithms*

Feature Reduction Method	AUC	Sensivity	Specificity
Information gain – 10 feature	0.873	0.823	0.774
Information gain – 5 feature	0.853	0.794	0.794
Gini – 10 feature	0.839	0.823	0.764
Relief – 10 feature	0.812	0.813	0.813
ANOVA – 10 feature	0.835	0.735	0.578

The AUC and ROC are two common metrics used to evaluate the performance of classification models in machine learning. The ROC curve is a graphical representation of the trade-off between sensitivity (true positive rate) and specificity (true negative rate) at various threshold settings. The AUC is the area under the ROC curve and provides a single value that summarizes the overall performance of the classification model. A higher AUC value indicates better classification performance, with an AUC of 1.0 representing a perfect classifier, while an AUC of 0.5 suggests no better performance than random chance[29].

Results

The demographic characteristics of the patient groups were as follows: 71 out of 102 (69.6%) migraine patients were female, and 31 (21.4%) were male (chi-square test value: 7.529, 1 degree of freedom, and p-value < 0.01). In the EM group, there were 40 women and 16 men, while in the PM group, there were 16 women and 9 men, while in the CM group there were 15 women and 6 men. The mean age was 38.7 ± 10.9 for the EM group, 35.7 ± 11.1 for the PM group and 39.1 ± 14.5 for the CM group. To test for differences among the groups, a one-way ANOVA test was performed. The result of the ANOVA test showed that there was no significant difference in mean age among the three groups ($F(2, 98) = 1.666$, $p = 0.195$). The peak prevalence for female patients was in the third, fourth, and fifth decades of life, while for male patients, it was in the third and fourth decades.

The algorithm-based information "gain" method for feature reduction had the best performance among all methods. This finding was revealed based on AUC, sensitivity, and specificity obtained by different feature reduction methods in combination with different AI algorithms. The best possible AUC, sensitivity, and specificity values with different AI algorithms are reported in **Table 1** for the algorithm-based feature reduction methods. A comparison was also made by selecting the most important 5 and 10 features, and the latter yielded better results **(Table 1)**. The first-order, gray-level size zone matrix, and gray-level co-occurrence matrix classes were the dominant feature classes for the information gain model.

In the classification with reduced features, the Random Forest algorithm performed the best and correctly classified 82.4% (84 out of 102 labeled segmentation data) of the migraine group with an AUC value of 0.873. The sensitivity and specificity, accuracy, positive predictive value, negative predictive value, and F1 value for predicting the migraine group were 82.4%, 77.5%, 84.4%, 78.5%, 81.4%, and 0.804, respectively. **Table 2** shows the detailed performances of machine learning algorithms and confusion matrices for detecting migraine. **Figure 4** shows the comparison results of ROC curves.

We first conducted an ANOVA analysis and compared four groups: EM, PM, CM, and healthy subjects. We obtained a significant overall difference among the groups ($F(3, 96) = 8.47$, $p = 0.002$). Further pairwise comparisons were performed to understand the differences between specific groups. The comparisons revealed that there were no significant differences between the EM and PM groups ($p= 0.22$). However, we found significant differences between the EM group and both the CM group and healthy subjects ($p = 0.01$). Similarly, the PM group showed significant differences when compared to the CM group and healthy subjects ($p = 0.01$). The CM group and the healthy subjects also exhibited significant differences between them ($p = 0.007$). Given the lack of significant differences between the EM and PM subtypes, it was decided to merge these two groups into a single group for

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

Table 2. *Performance metrics for predicting migraine group*

Model	AUC	F1	Precision	Sensivity	Specificity	Confusion Matrix	
						True Positive	False Positive
						False Negative	True Negative
kNN	0.805	0.747	0.714	0.784	0.686	80	22
						32	70
Tree	0.79	0.736	0.747	0.725	0.754	74	28
						25	77
SVM	0.786	0.77	0.785	0.754	0.794	77	25
						21	81
Random Forest	0.873	0.803	0.785	0.823	0.774	84	18
						23	79
Neural Network	0.819	0.759	0.745	0.774	0.735	79	23
						27	75
Naive Bayes	0.828	0.766	0.759	0.774	0.754	79	23
						25	77
Logistic Regression	0.659	0.019	1.0	0.009	1.0	1	101
						0	102
CN2 rule inducer	0.832	0.714	0.744	0.686	0.764	70	32
						24	78

SVM: The Support Vector Machine, kNN: k-Nearest Neighbor

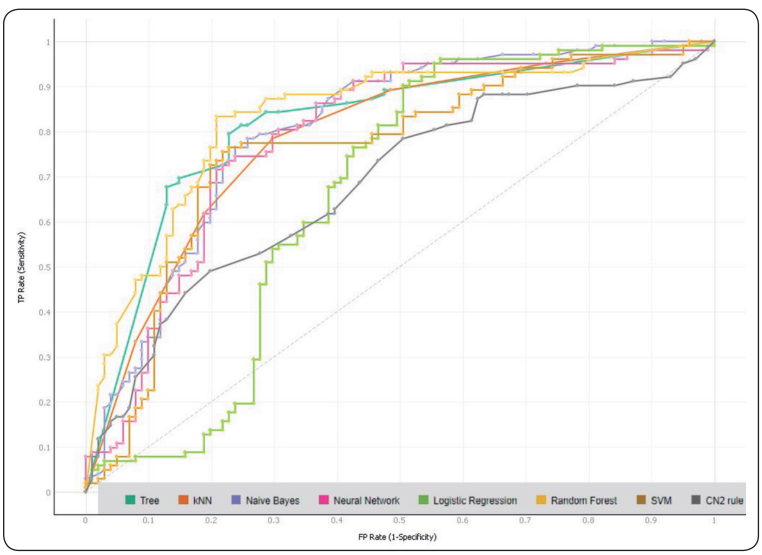

Figure 4. *Comparison of ROC curves*

further analysis. The new combined group, referred to as the Episodic-Probable (EM-PM) subtype, will be further compared to the CM group and healthy subjects to better understand the patterns and differences in the data.

The kNN algorithm correctly classified 74.1% (60 out of 81 labeled segmentation data) of the EM-PM group with an AUC value of 0.848. The sensitivity and specificity, accuracy, positive predictive value, negative predictive value, and F1 value for predicting the migraine group were 74.1%, 78.4%, 76.5%, 73.2%, 79.2%, and 0.736, respectively. **Table 3** shows the detailed performance of machine learning algorithms for detecting EM-PM. The Random Forest algorithm correctly classified 90.5% (19 out of 21 labeled segmentation data) of the CM group with an AUC value of 0.983. The sensitivity and specificity, accuracy, positive predictive value, negative predictive value, and F1 value for predicting the migraine group were 90.5%, 95.1%, 97.2%, 79.2%, 98.0%, and 0.844, respectively. **Table 3** shows the detailed performance of machine learning algorithms for detecting CM migraine.

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

Table 3. *Performance metrics for predicting episodic-probable migraine group and chronic migraine group*

Predicting episodic-probable migraine group						Predicting chronic migraine group				
Model	AUC	F1	Pre	Sensivity	Specificity	AUC	F1	Pre	Sensivity	Specificity
kNN	0.847	0.736	0.731	0.74	0.784	0.944	0.714	0.714	0.714	0.941
Tree	0.802	0.732	0.737	0.728	0.794	0.97	0.869	0.8	0.952	0.95
SVM	0.79	0.711	0.646	0.79	0.656	0.937	**	**	**	1.0
RF	0.802	0.746	0.729	0.765	0.774	0.982	0.844	0.791	0.904	0.95
NN	0.807	0.712	0.666	0.765	0.696	0.937	0.296	0.666	0.19	0.98
NB	0.847	0.771	0.733	0.814	0.764	0.915	0.68	0.586	0.809	0.882
LR	0.807	0.714	0.643	0.802	0.647	0.678	**	**	**	1.0
CN2	0.791	0.678	0.644	0.716	0.686	0.965	0.808	0.73	0.904	0.931

*The Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Random Forest (RF), Neural Network (NN), Logistic Regression (LR), Naive Bayes (NB), CN2 rule inducer (CN2), Precision (Pre).
**Could not be measured (Low sampling).

Discussion

Our machine learning model correctly differentiated 82.4% (84 of 102) of the migraine patients from healthy subjects. Within the migraine group, it correctly classified 74.1% (60 of 81) of the EM-PM patients and 90.5% (19 of 21) of the CM patients. Our research expands on the work of Zhiye Chen et al.,[21] who identified altered MR image texture characteristics of the PAG in EM as potential imaging biomarkers. By extending the analysis to all migraine subtypes, including the probable subtype, and atypical cases with overlapping symptoms, our study provides a more comprehensive understanding of the neurological mechanisms underlying migraines.

Interestingly, we found no significant difference between PM and EM patients in terms of PAG radiomics features. This suggests that the neurological mechanisms affecting the PAG might be similar between these two subtypes, despite the differences in their clinical presentation. This finding could further help refine the classification of migraine subtypes and contribute to a better understanding of the similarities and differences in their underlying pathophysiology.

This study has several important implications for future research and clinical practice in the field of migraine management. First, our machine learning model, which uses MRI radiomics analysis to differentiate between migraine patients, their subtypes, and healthy subjects, has demonstrated high accuracy. This enhanced diagnostic accuracy has the potential to enable healthcare professionals to more effectively identify and classify migraine cases, including the probable migraine group with overlapping symptoms from other diseases. As a result, patients can receive a more precise diagnosis, reducing the risk of misdiagnosis. Second, improved diagnostic accuracy allows healthcare professionals to intervene earlier in the course of the disease. Early intervention has the potential to prevent the progression of migraines, reduce the impact of the condition on the patients' quality of life, and decrease the burden on healthcare systems. Lastly, understanding the neurological mechanisms underlying migraine and its subtypes can inform the development of personalized treatment strategies. By tailoring treatments to each patient's specific subtype and neurological profile, healthcare professionals can optimize the effectiveness of interventions, leading to better patient outcomes and overall management of migraines.

Another finding in our study was that more females were afflicted by migraine than males ($p = 0.006$). This was the expected finding in migraine patients and was also the case within subtypes[30, 31]. In agreement with previous reports, the mean age in the CM subgroup (39.1) was higher than the EM subgroup (37.8) and PM subgroup (35.7), but the difference was not statistically significant ($F(2, 98) = 1.666$, $p = 0.195$), probably due to the small sample size ($n = 102$). In a study conducted by *Buse* et al.,[32] the researchers compared the sociodemographic and comorbidity profiles of CM and EM sufferers. The study had a large sample size, including 11,944 participants with 8,219 diagnosed with EM and 1,491 with CM. The mean age of participants in the EM group was 43.9 years, while for the CM group, it was 42.6 years. However, this study did not report on probable migraine (PM) patients or the presence of aura. On the other hand, *Bigal* et al.[33] focused on estimating the prevalence and burden of CM in the general population. Their study surveyed a total of 120,000 households and screened 24,000 individuals. The resulting sample consisted of 1,134 patients with EM, 410 with CM, and 2,317 with PM. The mean age of participants was 45.3 years for EM, 46.4 years for CM, and 46.2 years for PM. Similar to the study by Buse et al., the presence of aura was not specifically reported.

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

When comparing the mean ages of participants across the studies, it can be observed that the mean ages in our study groups were generally lower than those in the other studies[32, 33]. The reason of this could be attributed to various factors such as differences in study population, recruitment criteria (including the exclusion of patients with aura), sample size, and study design, which may have influenced age distribution and representation of the general migraine patient population.

Chen et al.[21] evaluated the contribution of texture parameters of the PAG in migraine patients and found that angular second moment (ASD), inverse difference moment (IDM), and entropy scores were significantly different in the EM group compared to healthy subjects. However, their number of patients was small, and they only used texture analysis as the method of analysis. In contrast, our study included four times as many patients and evaluated over 100 features obtained by radiomics analysis. Moreover, we developed a machine learning model based on these features to differentiate between patient groups.

Another study reported disrupted functional connectivity of the PAG subregions in EM, which the authors attributed to possible altered PAG function[34]. They hypothesized that PAG dysfunction is associated with PAG texture inhomogeneity. Although texture heterogeneity is rare in conventional MRI, several studies have explored this topic[21, 35]. Iron deposition has been proposed as a probable contributor to this heterogeneity[36, 37]. Our study's results enable us to identify differences between CM and EM-PM patients from healthy subjects, as well as between each subtype in terms of PAG radiomics features. However, we cannot conclude that PAG dysfunction is associated with PAG texture inhomogeneity. Examining control regions other than migraine-specific regions could help determine if the algorithm can effectively identify migraines based on PAG dysfunction, rather than other atypical features of images.

Feature selection methods are crucial in radiomics for identifying the most informative and relevant features from high-dimensional imaging data[38]. These methods are employed to select the optimal features for a given learning algorithm. They contribute to improved model performance, reduced computational complexity, and prevention of overfitting by eliminating redundant or irrelevant features. As a result, they lead to more accurate and interpretable radiomic models, which facilitate better clinical decision-making and personalized patient care[38].

We compared different algorithm-based feature selection methods and found that "gain" method was most useful in aiding diagnosis **(Table 1)**. We attempted to achieve better results by changing the selected feature numbers. The selection of the 10 most important features was superior to that of 5 **(Table 1)**. "Gain" method demonstrated that the first-order statistics, gray-level size zone matrix, and gray-level co-occurrence matrix classes were the dominant feature classes for differentiating the migraine group from healthy subjects. First-order statistics features describe the distribution of individual voxel values without concern for spatial relationships. These are histogram-based properties that report the mean, median, maximum, and minimum values of the voxel intensities on the image, as well as their skewness, kurtosis, uniformity, and entropy[39]. First-order statistics, especially entropy, proved highly valuable for differentiation. According to the literature, "gain" method calculates the decrease in entropy or surprise resulting from a dataset transformation[40]. Based on this, we consider "gain" as the most valuable method for reducing features.

The AI algorithms employed in this study leveraged advanced machine learning techniques, allowing computers to analyze large volumes of data typically obtained through radiomics, texture analysis, or other data mining techniques. By identifying underlying patterns and making informed predictions, these algorithms can adapt and improve over time as they encounter more data, refining their ability to make accurate predictions[41] In radiomics, various AI algorithms, such as Support Vector Machines (SVM) and Random Forest (RF), are utilized to analyze the selected features for tasks like classification, regression, and segmentation. These algorithms are chosen based on their ability to handle high-dimensional data, nonlinear relationships, and robustness against overfitting[42]. Support Vector Machines are effective in handling small to moderate-sized datasets with high dimensionality, while Random Forest, an ensemble learning algorithm, can manage missing data and high-dimensional datasets[43]. The use of these AI algorithms in radiomics allows for more accurate and efficient extraction of quantitative information from medical images, ultimately aiding in early detection, diagnosis, prognosis, and treatment planning for various diseases.

Among the various AI algorithms used in our study, Random Forest exhibited the best diagnostic performance for classifying the migraine group **(Table 2** and **Table 3)**. Meanwhile, kNN demonstrated the best performance for classifying migraines with the EM-PM subtype **(Table 3)**. The Random Forest algorithm, introduced in 2001, has gained considerable popularity in both regression and classification tasks[44]. It follows specific rules for tree growing, tree combination, self-testing, and post-processing, and is known for being robust to overfitting, stable in the presence of outliers, and more efficient in high-dimensional parameter spaces than other AI algorithms[45].

The generalizability of the findings of the present study is subject to a few limitations. First, it is a retrospective single-center study with a small sample size. Another important limitation is the significantly lower number of CM patients than that of EM-PM patients. The main reason

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

for this is that CM has a relatively low incidence[30–33]. Second, MR images were obtained from participants over a period of 15 months. Nevertheless, we aimed to standardize the images by equalizing the voxel sizes and ensuring the homogeneity of the signal intensities by excluding the extreme values with filtration to reduce the effect of these differences. To minimize variabilities and effects, all image datasets in our study underwent a normalization procedure. Third, the segmentation of images was performed manually. Manual segmentation was performed by a single person, and interobserver evaluation was not performed. However, we aimed to be objective about segmentation due to the identifiability of the area we are interested in, and the special interest of the supervising author in this subject and the field. Fourth, the use of '10-fold cross-validation' as an internal validation technique for evaluating the performance of AI algorithms is less ideal than the 'hold-out technique', which uses a separate validation group. Nevertheless, this technique has been well-accepted in the field of radiomics, and in one of the rare reviews on validation[24], this technique was largely preferred for radiomics modeling of kidney tumors (9 of 20 kidney tumors). We know of no study on the selection of a validation method to be used in the study of migraine patients. Fifth, we only used T1W images to evaluate texture parameters. T1W is generally considered better for depicting normal anatomy. Further studies may be designed as multi-center studies, and balanced groups may be created. In addition, the performance of AI algorithms should be confirmed in independent external datasets with intra- and interobserver correlations. Separate and comparative studies with other sequences such as T2-weighted image and susceptibility-weighted image for texture analysis may further our understanding. Future research should also consider including and comparing various patient groups, such as those with tension-type headaches, cluster headaches, and epilepsy, in addition to migraine patients.

A radiomics-based machine learning model, utilizing standard MR images obtained during the diagnosis and follow-up of migraine patients, shows promise not only in aiding migraine diagnosis and classification for clinical approach, but also in understanding the neurological mechanisms underlying migraines. Nevertheless, to establish its true diagnostic value, further research and validation are required.

References

1. *Burch R, Rizzoli P, Loder E*. The prevalence and impact of migraine and severe headache in the United States: Figures and trends from government health studies. Headache 2018;58:496-505. https://doi.org/10.1111/head.13281
2. *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;21:4-7. https://doi.org/10.1186/s10194-020-01208-0
3. *Kalra AA, Elliott D*. Acute migraine: Current treatment and emerging therapies. Ther Clin Risk Manag 2007;3:449-59.
4. *Kirchmann M*. Migraine with aura: New understanding from clinical epidemiologic studies. Curr Opin Neurol 2006;19:286-93. https://doi.org/10.1097/01.wco.0000227040.16071.a9
5. *Evans RW*. Migraine mimics. Headache 2015;55:313-22. https://doi.org/10.1111/head.12518
6. *Applebee A*. The clinical overlap of multiple sclerosis and headache. Headache 2012;52 Suppl 2:111-6. https://doi.org/10.1111/j.1526-4610.2012.02243.x
7. *Lin GY, Wang CW, Chiang TT, Peng GS, Yang FC*. Multiple sclerosis presenting initially with a worsening of migraine symptoms. J Headache Pain 2013;14:70. https://doi.org/10.1186/1129-2377-14-70
8. *Wang Y, Wang XS*. Migraine-like headache from an infarction in the periaqueductal gray area of the midbrain. Pain Med 2013;14:948-9. https://doi.org/10.1111/pme.12096
9. *Otlivanchik O, Liberman AL*. Migraine as a stroke mimic and as a stroke chameleon. Curr Pain Headache Rep 2019;23:1-12. https://doi.org/10.1007/s11916-019-0801-1
10. *Kahriman A, Zhu S*. Migraine and tension-type headache. Semin Neurol 2018;38:608-18. https://doi.org/10.1055/s-0038-1673683
11. *Burch R*. Migraine and tension-type headache: Diagnosis and treatment. Med Clin North Am 2019;103:215-33. https://doi.org/10.1016/j.mcna.2018.10.003
12. *Bianchin MM, Londero RG, Lima JE, Bigal ME*. Migraine and epilepsy: A focus on overlapping clinical, pathophysiological, molecular, and therapeutic aspects. Curr Pain Headache Rep 2010;14:276-83. https://doi.org/10.1007/s11916-010-0121-y
13. *Evans RW, Burch RC, Frishberg BM, Grosberg BM, Solomon S, Wöber-Bingöl Ç, et al*. Neuroimaging for migraine: The American headache society systematic review and evidence-based guideline. Headache 2020;60:318-36. https://doi.org/10.1111/head.13720
14. *Linnman C, Moulton EA, Barmettler G, Becerra L, Borsook D.* Neuroimaging of the periaqueductal gray: state of the field. Neuroimage 2012;60:505-22. https://doi.org/10.1016/j.neuroimage.2011.11.095
15. *Mathew NT*. Pathophysiology of chronic migraine and mode of action of preventive medications. Headache 2011;51:84-92. https://doi.org/10.1111/j.1526-4610.2011.01955.x
16. *Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö*. Radiomics with artificial intelligence: A practical guide for beginners. Diagnostic Interv Radiol 2019;25:485-95. https://doi.org/10.5152/dir.2019.19321
17. *Gillies RJ, Kinahan PE, Hricak H.* Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-77. https://doi.org/10.1148/radiol.2015151169
18. *Wang Q, Li Q, Mi R, Wu Q, Yuan Q, Zhang X, et al*. Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: A cohort study. J Magn Reson Imaging 2019;49:825-33. https://doi.org/10.1002/jmri.26265
19. *Peng J, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L*. A radiomics nomo-

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.

gram for preoperative prediction of microvascular invasion risk in hepatitis b virus-related hepatocellular carcinoma. Diagnostic Interv Radiol 2018;24:121-7.
https://doi.org/10.5152/dir.2018.17467
20. *Castellano G, Bonilha L, Li LM, Cendes F.* Texture analysis of medical images. Clin Radiol 2004;59:1061-9.
https://doi.org/10.1016/j.crad.2004.07.008
21. *Chen Z, Chen X, Liu M, Liu S, Ma L, Yu S*. Nonspecific periaqueductal gray lesions on T2WI in episodic migraine. J Headache Pain 2016;17:3-7.
https://doi.org/10.1186/s10194-016-0695-9
22. *Headache Classification Committee of the International Headache Society (IHS).* The International Classification of Headache Disorders, 3rd edition. Cephalalgia 2018;38:1-211.
https://doi.org/10.1177/0333102417738202
23. *Kocak B, Kus EA, Kilickesmez O.* How to read and review papers on machine learning and artificial intelligence in radiology: a survival guide to key methodological concepts. Eur Radiol 2021;31:1819-30.
https://doi.org/10.1007/s00330-020-07324-4
24. *Kocak B, Durmaz ES, Erdim C, Ates E, Kaya OK, Kilickesmez O*. Radiomics of renal masses: Systematic review of reproducibility and validation strategies. Am J Roentgenol 2020;214:129-36.
https://doi.org/10.2214/AJR.19.21709
25. *Park JE, Park SY, Kim HJ, Kim HS*. Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives. Korean J Radiol 2019;20:1124-37.
https://doi.org/10.3348/kjr.2018.0070
26. *Li S, Zhou B*. A review of radiomics and genomics applications in cancers: the way towards precision medicine. Radiat Oncol 2022;17:217.
https://doi.org/10.1186/s13014-022-02192-2
27. *Kikinis R, Pieper SD, Vosburgh KG*. 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. In: *Jolesz F*, ed. Intraoperative imaging and image-guided therapy. 1st ed. New York: Springer; 2014. p. 277-89.
https://doi.org/10.1007/978-1-4614-7657-3_19
28. *Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, et al.* Orange: Data mining toolbox in python. J Mach Learn Res 2013;14:2349-53
29. *Janssens ACJW, Martens FK*. Reflection on modern methods: Revisiting the area under the ROC Curve. Int J Epidemiol 2020;49:1397-403.
https://doi.org/10.1093/ije/dyz274.
30. *Baykan B, Ertas M, Karl N, Zarifoglu M, Siva A, Saip S, et al*. Migraine incidence in 5 years: a population-based prospective longitudinal study in Turkey. J Headache Pain 2015;16:103.
https://doi.org/10.1186/s10194-015-0589-2
31. *Rasmussen BK, Glostrup T, Studies P*. Epidemiology of headache. Cephalalgia 1995;15:45-68.
https://doi.org/10.1046/j.1468-2982.1995.1501045.x
32. *Buse DC, Silberstein SD, Manack AN, Papapetropoulos S, Lipton RB*. Psychiatric comorbidities of episodic and chronic migraine. J Neurol 2013;260:1960-9.
https://doi.org/10.1007/s00415-012-6725-x
33. *Bigal ME, Lipton RB, Winner P, Reed ML, Diamond S, Stewart WF; AMPP advisory group.* Migraine in adolescents: association with socioeconomic status and family history. Neurology 2007;69:16-25.
https://doi.org/10.1212/01.wnl.0000265212.90735.64
34. *Chen Z, Chen X, Liu M, Liu S, Ma L, Yu S*. Disrupted functional connectivity of periaqueductal gray subregions in episodic migraine. J Headache Pain 2017;18:36.
https://doi.org/10.1186/s10194-017-0747-9
35. *Rocca MA, Ceccarelli A, Falini A, Filippi M.* Brain gray matter changes in migraine patients with T2-visible lesions: A 3-T MRI Study. Stroke 2006;37:1765-70.
https://doi.org/10.1161/01.STR.0000226589.00599.4d
36. *Tepper SJ, Lowe MJ, Beall E, Dodick DW, Phillips H, Schwedt TJ, et al*. Iron deposition in pain-regulatory nuclei in episodic migraine and chronic daily headache by MRI. Headache 2012;52:236-43.
https://doi.org/10.1111/j.1526-4610.2011.02056.x
37. *Domínguez C, López A, Ramos-Cabrer P, Garcia-Ruiz-Espiga P, Solís García del Pozo J, et al*. Iron deposition in periaqueductal gray matter as a potential biomarker for chronic migraine. Neurology 2019;92:E1076-85.
https://doi.org/10.1212/WNL.0000000000007047
38. *Remeseiro B, Bolon-Canedo V*. A review of feature selection methods in medical applications. Comput Biol Med 2019;112:103375.
https://doi.org/10.1016/j.compbiomed.2019.103375
39. *Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al*. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2:36.
https://doi.org/10.1186/s41747-018-0068-z
40. *Fan R, Zhong AM, Wang S, Cao J, Shen C, Zou Y, et al*. Entropy-based information gain approaches to detect and to characterize gene-gene and gene-environment interactions/correlations of complex diseases. Genet Epidemiol 2011;35:706-21.
https://doi.org/10.1002/gepi.20621
41. *Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al*. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med 2022;5:66.
https://doi.org/10.1038/s41746-022-00611-y
42. *Wagner MW, Namdar K, Biswas A, Monah S, Khalvati F, Ertl-Wagner BB*. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. Neuroradiology 2021;63:1957-67.
https://doi.org/10.1007/s00234-021-02813-9
43. *Adugna T, Xu W, Fan J*. Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images. Remote Sensing 2022;14:574.
https://doi.org/10.3390/rs14030574
44. *Couronné R, Probst P, Boulesteix A*. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 2018;19:270-84.
https://doi.org/10.1186/s12859-018-2264-5
45. *Menze BH, Kelm BM, Masuch R, Franke J, Hamprecht FA, et al*. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics 2009;10:213-29.
https://doi.org/10.1186/1471-2105-10-213

Az alábbi dokumentumot magáncélra töltötték le az eLitMed.hu webportálról. A dokumentum felhasználása a szerzői jog szabályozása alá esik.