Subgrouping Factors influencing Migraine Intensity in Women: A Semi-automatic Methodology based on Machine Learning and Information Geometry

Abstract

Background: Migraine is a heterogeneous condition with multiple clinical manifestations. Machine-learning algorithms permit the identification of population groups providing analytical advantages over other modeling techniques. **Objective:** The aim of this study was to analyze critical features that permit to differentiate subgroups of patients with migraine according to the intensity and frequency of attacks by using machine-learning algorithms. **Methods**: Sixty-seven women with migraine participated. Clinical features of migraine, related-disability (MIDAS), anxiety/depressive levels (HADS), anxiety state/trait levels (STAI) and pressure pain thresholds (PPT) over the temporalis, neck, second metacarpal, and tibialis anterior were collected. Physical examination included the flexion-rotation test, cervical range of cervical motion, forward head position in sitting and standing, passive accessory intervertebral movements (PAIVMs) with headache reproduction, and joint positioning sense error. Subgrouping was based on machine-learning algorithms by using Nearest Neighbors algorithms, multisource variability assessment, and Random Forest. **Results**: For migraine intensity, group 2 (women with regular migraine headache intensity of 7) were younger, had lower joint positioning sense error in cervical rotation, greater cervical mobility in rotation and flexion, lower flexion-rotation test, positive PAIVMs reproducing migraine, normal PPTs over tibialis anterior, shorter migraine history, and lower cranio-vertebral angle in standing than the remaining migraine intensity subgroups. The most discriminative variable was the flexion-rotation test to the symptomatic side. For migraine frequency, no model was able to identify differences between groups, i.e. patients with episodic or chronic migraine. **Conclusions**: A subgroup of women with migraine with common migraine intensity was identify with machine-learning algorithms.

Keywords: Migraine, Random Forest, Machine Learning, Multisource variability

Subgrouping Factors influencing Migraine Intensity in Women: A Semi-automatic Methodology based on Machine-Learning and Information Geometry

Introduction

Migraine is a primary headache disorder with a worldwide prevalence of 11.6% within female: male ratio 2:1 (1). In the last Global Burden of Disease Study, headache (e.g., migraine and tension-type headache) was found to be the second most prevalent pain condition in the world (2). In fact, health care costs of primary headache in Europe (€13.8 billion) mainly account for migraine and tension-type headache (3).

Migraine attacks are characterized by recurrent episodes of severe headache with accompanying symptoms of autonomic nervous system dysfunction. It is accepted that the pathophysiology of migraine is associated to abnormal neuronal excitability leading to cortical spreading depression and to sensitization of trigemino-vascular pathways (4). In general, pain is a complex subjective experience that includes sensory-discriminative, affective, and cognitive aspects. In such a scenario, it is usually seen in clinical practice that migraine can be heterogeneous condition with multiple manifestations. Therefore, the identification of subgroups of patients can help to a better understanding of migraine and provides useful data to support developing clinical decision support systems.

Machine-learning algorithms trained to automatically classify patient populations can be used as classification methods since they provide distinct analytical advantages over other modeling techniques. For instance, supervised machine-learning techniques have the ability to assess all available covariates in every possible clinically meaningful combination and report the combinations in mutually exclusive groups capable of being easily incorporated into decision-support modeling (5). In fact, they can be combined with network methods for improving prediction and detecting potential correlations between variables (6,7).

Supervised machine-learning analyses have been able to identify groups of patients experiencing the highest rates of mortality post-interhospital transfer (8); however, its use is scarce in patients with headache. Garcia-Chimeno et al were able to distinguish with 93% accuracy between patients with sporadic migraine, patients with chronic migraine, and patients at risk of medication overuse via feature selection techniques and machine-learning analyses over diffusion tensor images (DTIs) and questionnaire answers related to emotion and cognition (9). An overview of how Machine Learning techniques have been used in the general context of pain research has been presented by Lötsch and Ultsch (10).

The intensity and frequency of headache attacks are two features that are clinically used in the differential diagnosis of headaches. For instance, migraine is characterized by headache attacks of moderate-severe intensity lasting 4-72 hours as opposite to headache attacks of mild-moderate intensity lasting from 30 min to 7 days as occurs in tension-type headache (11). The frequency of headache is mainly used for classification between episodic or chronic headache. The episodic form comprises headache attacks occurring less than 15 days per month, while the chronic comprises headaches occurring 15 or more days/month for more than 3 months and with migraine features on at least 8 days/month (11). Therefore, we aimed to identify differences in clinical features and the presence of musculoskeletal disorders that permit to subgrouping patients with migraine according to the intensity and frequency of the migraine attacks. We chose these clinical variables for subgrouping since migraine is characterized by moderate-severe intensity of headache and because headache frequency is considered the main outcome in clinical trials. Further, the variables used in this study to subgrouping included clinical features and questionnaires focusing on migraine-related items and also the presence of cervical musculoskeletal impairments, e.g. cervical range of motion, head position, joint position sense error, or reproduction of the headache on manual palpation, commonly associated with primary headaches (12). We hypothesized that patients with higher intensity and/or higher frequency of migraine would exhibit more severe musculoskeletal disorders, e.g. lower cervical range of motion, decrease pressure pain thresholds, higher joint position sense error, than those with lower intensity and/or frequency of migraine attacks.

Methods

Participants

Consecutive women with migraine recruited from a Headache Unit located in a tertiary university-based hospital were included. To be eligible, they had to meet the diagnostic criteria of migraine according to the International Classification of Headache Disorders, 3rd edition (11). Migraine features including location, years with disease, frequency and intensity of migraine attacks, family history, and medication intake were collected. All participants were screened by an experienced neurologist with more tan 20 years of experience in headaches. Participants were excluded if presented any of the following: 1, other primary or secondary headache; 2, history of cervical and/or head trauma; 3, pregnancy; 4, history of cervical herniated disk or cervical osteoarthritis on medical records; 5, underlying systematic medical disease, e.g., rheumatoid arthritis, lupus erythematous; 6, comorbid fibromyalgia syndrome; 7, had received treatment including anesthetic blocks, botulinum toxin or physical therapy within the previous 6 months; or, 8, male gender. All participants signed the informed consent form before their inclusion in the study. The local Ethics Committee of the [redacted]

All examinations were held when patients were headache-free and when at least one week had elapsed since the last migraine attack to avoid migraine related allodynia.

Since some patients exhibit high frequency of migraine attacks, careful observation of this parameter was considered for examination. If not possible, those women with high frequency of attacks were evaluated at least 48 hours after the last attack. Participants were asked to avoid any analgesic or muscle relaxant 24 hours prior to the examination. No change was made on their prophylactic treatment.

Self-reported Outcomes

A 4-weeks headache diary was used to register clinical features of the migraine (13): 1, migraine intensity (the mean intensity of the days with migraine attack based on a 11-points Numerical Pain Rate Scale (NPRS); 0: no pain, 10: maximum pain); 2, migraine frequency (days/week); 3, migraine duration (hours/attack).

The Hospital Anxiety and Depression Scale (HADS) was used to evaluate anxiety (HADS-A, 7items) and depressive (HADS-D, 7items) levels (14). In headache patients, the HADS has shown good internal consistency (15). Higher scores indicate greater levels of anxiety or depressive levels.

The State-Trait Anxiety Inventory (STAI) was used to assess state (STAI-S) and trait (STAI-T) anxiety levels (16). The STAI-S assesses relatively enduring symptoms of anxiety at a moment and the STAI-T scale measures a stable propensity to experience anxiety and tendencies to perceive stressful situation as threatening. Both subscales had exhibited good internal consistency and high reliability (17). Higher scores are indicate of greater state or trait anxiety levels.

The Migraine Disability Assessment Scale (MIDAS) questionnaire was used to assess the degree of related-disability in daily activities (work or school, family and social) caused by migraine (18). The final score comes from the sum of the missed days regarding the 3 activities.

Widespread Pressure Pain Sensitivity

Pressure pain thresholds (PPTs), i.e., the minimal amount of pressure where a sensation of pressure first changes to pain, were bilaterally assessed with an electronic algometer (Somedic AB, Farsta, Sweden) over the temporalis muscle, the cervical spine, the second metacarpal and the tibialis anterior muscle following previous guidelines (19). All participants attended a session for familiarization with the pressure test procedure over the wrist extensors. The order of assessment was randomized. The mean of 3 trials on each point was calculated and used for the analysis. Since no side-to-side differences were observed, mean of both sides were used in the analysis. Participants were asked to avoid any analgesic or muscle relaxant 24 hours prior to the examination.

Physical Examination

Physical examination included the musculoskeletal impairments most commonly associated to patients with headache (12,20): cervical flexion-rotation test, active range of cervical motion, forward head posture, passive accessory intervertebral movements with head pain reproduction and joint position sense error (JPSE).

The cervical flexion-rotation test (FRT) and active cervical range of motion were assessed as previously described (21). Briefly, for the FRT, participants were positioned in supine and a CROM® device was placed at their head. The evaluator performed a maximum flexion of the cervical spine followed by rotation toward either side. The rotation limit was determined when the evaluator self-perceived tissue resistance or the patient reported the presence of pain at the upper cervical area. Active cervical range of motion was assessed with a CROM® device and participants seated in a relaxed position on a chair. The CROM® device was positioned on the subject's head and a familiarization session was performed. The mean of three repetitions was considered in the analysis. This procedure has shown excellent reliability in migraine patients (22).

Forward head position, passive accessory intervertebral movement with headache reproduction and Joint Position Sense Error (JPSE) were assessed following previous guidelines (23). The cranio-vertebral angle, i.e., the angle between the horizontal plane and a line from the tip of the C7 spinous process to the tragus of the ear, was calculated in sitting and standing positions for assessing forward head posture as previously described (24). A smaller angle reflects a greater forward head position. Passive accessory inter-vertebral motions were used to evaluate the presence of referred pain to the head elicited by a posterior to anterior (PA) pressure applied to C1-C2 segment in an attempt to provoke a pain response able to reproduce a migraine attack. This procedure has been able to differentiate 3 migraine subtypes: pain-free, local pain, and pain referral to the head (25). Finally, the JPSE was evaluated by assessing the subject ability to relocate the head to a natural head posture, whilst blindfolded, on active cervical extension, left and right rotations. The difference between the starting (zero) and the position on return was calculated in absolute degrees for each movement tested. Three trials were performed in each direction and the mean JPSE was used in the analysis (23).

All examinations were conducted by an experienced therapist with more than 15 years of experience in the management of headache patients and who was blinded to the migraine headache features (subgrouping classification as described below).

Data Analysis Methods

We considered a fully automated methodology that can be split into 4 steps. Firstly, we first input missing data using the Nearest Neighbors (NN) algorithm. Secondly, we assessed the multisource variability (26,27). According to the results, we sub-grouped the variables of migraine intensity and migraine frequency in order to ensure intergroup differences. Finally, random forests classifiers were used to determine physical factors influencing migraine headache intensity and frequency subgroups.

Nearest Neighbors (NN) algorithm

One of the most widely used algorithms to impute missing data is the NN algorithm. These algorithms are efficient methods to fill in missing data. Each missing value on a record is replaced by a value from related cases in the whole set of records that depends on the type of variable used: categorical missing values are replaced by the mode and quantitative ones are replaced by the mean (28). The number of neighbors was fixed to 10 before conducting experiments. Several papers including DNA microarray studies (29), forest inventory (30), or breast cancer (31) have shown benefits of NN as missing data imputer method.

Multisource Variability Assessment (MSV)

This MSV is based on Information Geometry (32,33), which provide a way for the comparison of dissimilarities between the probability distributions (Probability Density Functions, PDFs) of different data sources. In our case, we modeled headache intensity subgroup distributions using Kernel Density Estimation (KDE) (34). Due to KDE provides a non-parametric distribution, we used the non-parametric Jensen Shannon distance (JSD) to measure the distance between pairs of PDF's (35,36). A JSD is bounded between 0 and 1; where a value of 1 indicates that the compared distributions are disjoint. We constructed a simplex in which each point corresponds to a PDF and each edge joining two points measures the distance between the PDF's. Then, this can be reduced by applying projection methods, such as Principal Component Analysis (PCA) (37) or Multidimensional Scaling (MDS) (38,39), providing a graphical way to detect inter-group variability.

Case labelling

Before conducting the final machine-learning analyses, a preprocess analysis was carried out in the subgrouping variables. The original dataset was completed with two processed variables for grouping, headache intensity and headache frequency due to the low number of cases.

Patients were grouped according to their migraine headache intensity as follows: group 1, patients with migraine pain intensity ranging from 4 to 6; group 2, patients with migraine pain intensity equal to 7 (regular migraine attack pain intensity); group 3, patients with migraine intensity equal to 8; and, groups 4 and 5, patients who suffered headache attacks intensities of 9 and 10, respectively. A second subgrouping according to the frequency of migraine was also identified: group 0, patients with 1 to 8 days per month with migraine (episodic); group 1, patients with 9 to 16 days migraine attacks per month (episodic to chronic); group 2, patients with more than 16 days per month with migraine (chronic).

Random Forest Classifier

One of the current trends in machine learning research concerns ensemble methods that combine their results, as the case of Random Forest (RF), which constructs many decision trees that are used to classify by the majority vote (40,41). RF classifiers also allow to measure the variables that best explain intra-groups variance. Several authors proved that RF classification outperforms other conventional machine learning algorithms, such as back propagation neural networks and support vector machines and has the advantages of dealing with unbalanced or multiclass classification problems. These reasons have motivated the use of RF in the current study (42-44).

The parameters were fixed to 512 decision trees composing the forest, the maximum number of decision variables in each tree equal to the log_2N where N is the number of model inputs and the rest of parameters were fixed to the default proposed by the python implementation of scikit-learn (45).

Due to the number of samples in our database is short, we have used an ensemble of Random Forest to obtain more robust results. Besides, each Random Forest of the ensemble was cross-validated using 8 random stratified folds. This concept consists of creating 8 folds where the proportions of predictor labels are similar to original dataset (46). A visual description of the ensemble is presented in **Figure 1**. Finally, to assess the performance of the models, the recall and the F1-score were computed (47), according with the equations (1). Here, TP_c (True Positive) is the number of patients of a given group *c* hat are correctly classified, FP_c (False Positive) is the number of patients of other groups that are wrongly classified in the given group *c*, TN_c (True Negative) is the number of patients of other groups that are not classified in group *c*, and finally FN_c (False Negative) is the number of patients of a given group classified in other groups. The F1-score ranges between [0, 1], being 1 the perfect classification.

$$Recall = \frac{TP_c}{TP_c + FN_c}$$

$$Precision = \frac{TP_c}{TP_c + FP_c}$$

$$F1 - score = 2\frac{Recall \cdot Precision}{Precision + Recall} = \frac{2 \cdot TP_c}{2 \cdot TP_c + FP_c + FN_c} \quad (1)$$

Results

Participants

Ninety (n=90) consecutive women presenting with headache were screened for eligibility criteria. Twenty-three (25%) were excluded for the following reasons: co-morbid headaches (n=10); previous head or neck trauma (n=6); receiving anesthetic block in the past 3 months (n=5) or pregnancy (n=2). Finally, 67 women migraine (20% chronic, mean age: 42±12 years) satisfied all criteria and signed the informed consent. Participants were headache-free at the moment of examination with a mean of 7.5 ± 3.0 days without a migraine attack. Seventy (70%) of the patients self-reported the presence of neck pain mainly during their migraine attacks. Only 4 (6%) self-reported neck pain in interictal phases. **Table 1** shows clinical, psychological and psychophysical data of the sample.

Accuracy of the subgrouping models

After imputing missing data and checking the interclass difference distributions with MSV for migraine intensity (**Fig. 2A**) and frequency (**Fig. 2B**), the dataset was 200 times randomly stratified 8-fold cross-validated. This overcomes the limitation of the low number of individuals. Each of the 200 stratifications produced 8 different folds which contained similar proportions to the original dataset. As can be seen in Table 2, the group, to which more patients belong to, has a total of 21 women. Each fold is composed of 2 individuals of this class, and then the number of possible combinations is 210. We chose 200 RF because each of them will be cross-validated using 8 random stratified folds. This gives us a totally of 1600 different splits, which makes almost impossible not to consider the whole set of combinations.

For migraine intensity, the 8-fold cross-validation averaged recall and frequency of each group are presented in **Table 2.** The averaged F1-score for the 200 models is shown within **Figure 3A**. Looking at the F1-score, random forest models outperform random classification in a 50% on average. This shows that the variables enclosed in the current study have a certain discriminatory power for determining migraine intensity. The weighted sensitivity mean was 30.86%. It is worth to mention that groups with low density were the worst estimated, because of the low number of cases used to train and to validate the model. Additionally, group 1 contained patients with different headache intensities, which may probably hinder the estimation accuracy.

For migraine frequency, the mean accuracy of the 200 implemented models was 0.41, which implied a modest, but not despicable, improvement respect to randomness (**Fig. 3B**). According to the results showed in **Table 3**, none of the random forests was able to find group 2 individuals (a 0 score of sensitivity implies no true positives). This indicates that there was no evidence in the current data which facilitates to discriminate group 2. In this situation, the major possible accuracy score was near to 0.8.

An explanation to this fact can be found looking at how random forests models are generated, since they are not robust to unbalanced data and they usually tend to be biased towards the groups with the majority of elements. Even though the 8-fold cross-validation of the 200 models obtained an F1-score of 0.41 on average, that is a slightly higher than the expected F1-score associated to a random classification, not finding group 2 individuals makes impossible to interpret correctly which variables are influencing the estimation of the migraine frequency.

Variables importance

Random Forests also provide a quantification of the importance of the features within the subgrouping discrimination. The 10 most influential features of each of the 200 models were extracted only for migraine intensity. As it can be seen in **Figure 4**, 20 variables were chosen as the most important from the 200 generated models.

For migraine intensity, 6 variables were selected by all the models and other 3 by more than the 50% of the models. Therefore, the results can be considered to be robust. The 10 more frequent variables for identifying subgroup 2 were: age, JPSE in cervical rotation, active cervical range of motion in rotation and flexion, FRT to both symptomatic and non-symptomatic sides, positive PAIVMs, PPT on the tibialis anterior, years with migraine, and cranio-vertebral angle in standing. In such a scenario, group 2 (women with migraine headache intensity of 7) were younger, had lower JPSE in cervical rotation, greater active cervical range of motion in rotation and flexion, lower FRT to both sides, positive PAIVMs reproducing their migraine headache, normal PPT on tibialis anterior, shorter history with migraine and lower cranio-vertebral angle (i.e., higher forward head posture) in standing position than the remaining groups.

Once these clinical features were selected, we quantify their importance in the discriminative power of the models. In this sense, the histograms of the averaged 8-folds corresponding to each of the 200 models were computed just for migraine intensity (**Figure 5**). The descriptive statistics can be found in **Table 4**. The most discriminative variable in mean over the 200 models after a stratified 8-fold cross-validation was FRT to the symptomatic side (averaged influence of 3.02%).

Discussion

A group of women with migraine with common migraine intensity was identified with machine-learning algorithms. Random forest models identified the following most frequent variables in individual trees: age, JPSE in rotation, cervical mobility in rotation and flexion, positive flexion-rotation test, positive PAIVMs reproducing migraine, PPTs over tibialis anterior, migraine history, and cranio-vertebral angle in standing. The most discriminative variable in the model was the flexion-rotation test to the symptomatic side. The random forest model was not able to identify any subgroup depending on the frequency of migraine attacks (episodic, frequent episodic or chronic migraine). These results did not support the a priori hypothesis of this study since individuals with higher intensity or frequency of migraine attacks did not exhibit more severe musculoskeletal disorders.

It is important to note that features were selected in the current study to carry out a clinical classification when differentiating groups of women with migraine according to their intensity or frequency of migraine attacks. From a full set comprising clinical, psychological, and psychophysical outcomes and also physical examination a subgroup of women with migraine suffering from pain intensity of 7 (moderate-intense) during their attacks was identified. It is important to note that migraine pain is characterized by headache attacks of moderate-severe intensity lasting 4-72 hours accordingly to the International Classification of Headache Disorders (11). Since the results were robust, it seems that the random forest classifier model offered an efficient method for classifying this subgroup of migraine sufferers, as it has solid foundations in terms of statistical learning, enabling to optimize the decision function in the process.

The subgroup of migraine sufferers identified within the random forest model were younger, lower JPSE in cervical rotation, greater cervical mobility in rotation and flexion, lower flexion-rotation test (positive), positive PAIVMs reproducing migraine symptoms, normal PPTs over the tibialis anterior, shorter migraine history, and lower cranio-vertebral angle in standing as compared to other migraine intensity subgroups. The association of these variables with migraine is not new since some previous studies have investigated the presence of cervical musculoskeletal disorders in this population (20-25); although its association is still questioned. In fact, a recent meta-analysis has concluded that, among several cervical spine musculoskeletal impairments, individuals with migraine exhibit minimally reduced cervical range of motion with no differences in head posture or JPSE as compared to headache-free people (12). The current study identified a subgroup of women with migraine with some musculoskeletal disorders of the neck, e.g., positive flexion-rotation test, manual examination (PAIVMs) of the upper cervical able to reproduce their migraine symptoms, and greater forward head posture in standing, when compared to other subgroups of women with migraine. Current results agree with some previous studies suggesting a relevant role of the flexion-rotation test (21,23), the ability of reproducing migraine symptoms with manual examination of the upper cervical spine joints (25) or a forward head position (24) in migraine. In fact, it is interesting to note that other variables identified by the random forest model, such as cervical range of motion or PPTs over tibialis anterior muscle, should not be considered as impaired, since their values were normal. Similarly, shorter migraine history could be also related to the younger age of this group of patients. Therefore, our study identified that subclassification of individuals with migraine is a highly complex process needing sophisticated analysis such as machine-learning algorithms. Additionally, it is probably that musculoskeletal impairments of the cervical spine have different roles, not only, in promoting or precipitating migraine attacks but also in the intensity of the attacks. From a clinical viewpoint, the variables identified in our study would suggest that the upper cervical spine could be more relevant for this subgroup of patients with migraine than in others. This assumption is supported by the fact that this subgroup of patients exhibited normal cervical range of motion but a positive flexion-rotation test, which supports the presence of upper cervical spine impairment. Therefore, examination of musculoskeletal impairments of the cervical spine should focus on specific groups of migraine patients.

We should also discuss that our sample of women with migraine was explored in a headache-free situation for avoiding migraine-related allodynia and other concomitant symptoms. For instance, this situation also permitted the absence of neck pain during our exploration, a common symptom experienced by patients with migraine during their attacks and associated with a poor clinical presentation (48). It is possible that patients experienced concomitant neck pain during migraine attacks could also exhibit different musculoskeletal impairments of the cervical spine representing another subgroup.

We were not able to identify by using random forest models a cluster of variables associated with a group of women with migraine according to the frequency of attacks. We used a clinical subgrouping for headache frequency, mostly based on identification of infrequent episodic, frequent episodic, or chronic migraine. The lack of classification based on the frequency of migraine attacks may be related to the fact that some of the outcomes included in our study, e.g., PPTs, (19), active cervical range of motion (22), JPSE (23) or migraine pain reproduction with passive accessory inter-vertebral motions (25), have not been found to be significantly different between individuals with episodic or chronic migraine, whereas the differences in others, e.g., flexion-rotation test (21) are small. It is also possible the small number of patients within the chronic migraine group, as previously reported in the results section, would lead to an unpowered subgrouping.

Future studies should investigate variables associated to frequency of migraine attacks with other outcomes, i.e., migraine-related disability, or kinesiophobia.

Finally, although this is the first study using machine-learning algorithms for the identification of groups of patients with migraine, we should recognize some technical limitations. First, we should highlight that the short number of cases in some subgroups, having fewer than 20 subjects/group. This situation could have led to poor classification accuracy due to the dispersion of the decision space, e.g., in the classification according to migraine frequency. Future studies should include larger dataset of patients to avoid this problem and the main goal should bet the percentage of accuracy of the classifier. Second, future studies could include the use of algorithms for feature selection, such as sequential forward/backward floating selection (49), where the dimension of decision spaces would be reduced and therefore the points sparsity. Further, we only included a sample of women with migraine; therefore, current results should not be extrapolated to men with this condition. In addition, the current subclassification was based on clinical findings observed in a headache-free (interictal phase) status; hence, it is possible that examination during an active phase of a migraine attack could lead to different findings.

Conclusion

A subgroup of women with migraine with common migraine intensity (moderate to intensity, 7/10) was identify by using machine-learning algorithms. The random forest models identified age, JPSE in rotation, cervical mobility in rotation and flexion, positive flexion-rotation test, positive PAIVMs reproducing migraine, PPTs over tibialis anterior, migraine history, and cranio-vertebral angle in standing as main variables associated with the group of patients. No cluster of variables was identified accordingly the frequency of migraine.

Legend of Figures

Figure 1: Ensemble of Random Forest. Each Random Forest is composed of 512 decision trees. Each random forest is cross-validated using 8 random stratified folds.

Figure 2: (A) The MSV-Plot for the different intensity classes (B) The MSV-Plot for the different frequency classes. Source Probabilistic Outlyingness (SPO) measures the Jensen Shannon distance to the central probabilistic tendency of the whole dataset probability. This metric also ranges between [0, 1]. It is worth to mention that distances in B are very small and may not provide enough dissimilarity to be discriminative.

Figure 3: The histogram of the mean F1-score obtained in the 8-fold cross validation of the 200 Random Forest models for migraine intensity (A) and frequency (B) models.

Figure 4: Counting of the variables selected by the RF models. Age, JPSE rotation, FRT symptomatic side, FRT non-symptomatic side and positive PAIVMs were selected as one of the 10 most influential variables by all the models.

Figure 5: The histograms of the importance of the 10 most important variables of the 200 RF models for migraine intensity

References

1. Woldeamanuel YW, Cowan RP. Migraine affects 1 in 10 people worldwide featuring recent rise: A systematic review and meta-analysis of community-based studies involving 6 million participants. J Neurol Sci 2017; 372: 307-315.
2. Vos T, Abajobir AA, Abbafati C, Abbas KM, Abate KH, Abd-Allah F et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the global burden of disease study 2016. Lancet 2017; 390: 1211-1259.
3. Raggi A, Leonardi M. Burden and cost of neurological diseases: a European North-South comparison. Acta Neurol Scand 2015; 132: 16-22.
4. de Tommaso M, Sciruicchio V. Migraine and central sensitization: Clinical features, main co-morbidities and therapeutic perspectives. Curr Rheumatol Rev 2016; 12: 113-26.
5. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019; 20: e262-e273.
6. Pérez-Benito FJ, Villacampa-Fernández P, Conejero JA, García-Gómez JM, Navarro-Pardo E. A happiness degree predictor using the conceptual data structure for deep learning architectures. Comput Methods Programs Biomed 2019; 168: 59-68.
7. Navarro-Pardo E, González-Pozo L, Villacampa-Fernández P, Conejero, JA. Benefits of a dance group intervention on institutionalized elder people: a Bayesian network approach. Applied Mathematics and Nonlinear Sciences 2018; 3: 503-512.

8. Reimer AP, Schiltz N, Ho V, Madigan EA, Koroukian SM. Applying supervised machine learning to identify which patient characteristics identify the highest rates of mortality post-interhospital transfer. Biomed Inform Insights 2019; 11: 1178222619835548.

9. Garcia-Chimeno Y, Garcia-Zapirain B, Gomez-Beldarrain M, Fernandez-Ruanova B, Garcia-Monco JC. Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data. BMC Med Inform Decis Mak 2017; 17: 38.

10. Lötsch J, Ultsch A. Machine learning in pain research. Pain 2018; 159: 623-630.

11. ICHD-III Headache Classification Subcommittee of the International Headache Society: The International Classification of Headache Disorders, 3 edition. Cephalalgia 2018; 38: 1-211.

12. Liang Z, Galea O, Thomas L, Jull G, Treleaven J. Cervical musculoskeletal impairments in migraine and tension type headache: A systematic review and meta-analysis. Musculoskelet Sci Pract 2019; 42: 67-83.

13. Phillip D, Lyngberg AC, Jensen R. Assessment of headache diagnosis: A comparative population study of a clinical interview with a diagnostic headache diary. Cephalalgia 2007; 27: 1-8.

14. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand1983; 67: 361-370.

15. Juang KD, Wang SJ, Lin CH, Fuh JL. Use of the Hospital Anxiety and Depression Scale as a screening tool for patients with headache. Zhonghua Yi Xue Za Zhi (Taipei) 1999; 62: 749-55.

16. Spielberger CD. State-Trait Anxiety Inventory: A comprehensive bibliography. Palo Alto, CA: Consulting Psychologists Press; 1989.

17. Barnes LLB, Harp D, Jung WS. Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. Educ Psychol Meas 2002; 62: 603-618.

18. Stewart WF, Lipton RB, Kolodner K, Sawyer J, Lee C, Liberman JN. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. Pain 2000; 88: 41-52.

19. Palacios-Ceña M, Lima Florencio L, Natália Ferracini G, Barón J, Guerrero ÁL, Ordás-Bandera C, Arendt-Nielsen L, Fernández-de-las-Peñas C. Women with chronic and episodic migraine exhibit similar widespread pressure pain sensitivity. Pain Med 2016; 17: 2127-2133.

20. Luedtke K, Boissonnault W, Caspersen N et al. International consensus on the most useful physical examination tests used by physiotherapists for patients with headache: A Delphi study. Man Ther 2016; 23: 17-24.

21. Oliveira-Souza AIS, Florencio LL, Carvalho GF, Fernández-de-las-Peñas C, Dach F, Bevilaqua-Grossi D. Reduced flexion rotation test in women with chronic and episodic migraine. Braz J Phys Ther 2019; 23: 387-394.

22. Bevilaqua-Grossi D, Pegoretti KS, Goncalves MC, Speciali JG, Bordini CA, Bigal ME. Cervical mobility in women with migraine. Headache 2009; 49: 726-31.

23. Ferracini GN, Florencio LL, Dach F et al Musculoskeletal disorders of the upper cervical spine in women with episodic or chronic migraine. Eur J Phys Rehabil Med 2017; 53: 342-350.

24. Fernández-de-las Peñas C, Cuadrado ML, Pareja JA. Myofascial trigger points, neck mobility and forward head posture in unilateral migraine. Cephalalgia

2006; 26: 1061-1070.

25. Luedtke K, May A. Stratifying migraine patients based on dynamic pain provocation over the upper cervical spine. J Headache Pain 2017; 18: 97.

26. Sáez C., Zurriaga O, Pérez-Panadés J, Melchor I, Robles M, García-Gómez JM. Applying probabilistic temporal and multisite data quality control methods to a public health mortality registry in Spain: a systematic approach to quality control of repositories. J Am Med Infor Assoc. 2016; 23: 1085-1095.

27. Pérez-Benito FJ, Sáez C, Conejero JA, Tortajada S, Valdivieso B, García-Gómez JM. Temporal variability analysis reveals biases in electronic health records due to hospital process reengineering interventions over seven years. PLoS One 2019; 14: e0220369

28. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: A critical evaluation, BMC Med. Inform. Decis. Mak 2016; 16: 74.

29. Troyanskaya O, Cantor M, Sherlock G et al. Altman, missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.

30. Eskelson BNI, Temesgen H, Lemay V, Barrett TM, Crookston NL, Hudak AT. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scand J For Res 2009; 24: 235-246.

31. Jerez JM, Molina I, García-Laencina PJ et al. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artif Intell Med 2010; 50: 105-115.

32. Amari S, Nagaoka A. Methods of Information Geometry, American Mathematical Society. Japan: Oxford University Press; 2000.

33. Csiszár I, Shields PC. Information theory and statistics: A Tutorial. Found Trends Commun Inf Theory 2004; 1: 417-528.

34. Heidenreich NB, Schindler A, Sperlich S. Bandwidth selection for kernel density estimation: A review of fully automatic selectors. AStA Adv Statistic Anal. 2013; 97: 403-433.

35. Lin J. Divergence measures based on the shannon entropy. IEEE Trans Inf Theory 1991; 37: 145-151.

36. Cover TM, Thomas JA. Elements of information theory, 2nd edition. New Yersey: Wiley-Interscience; 2006.

37. Jolliffe I. Principal component analysis and factor analysis. In: Principal component analysis. Springer Series in Statistics. New York: Springer; 1986. Pp. 115-128.

38. Young FW, Hamer RM. Multidimensional scaling: History, theory, and applications. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc; 1987.

39. Brandes U, Pich C. Eigensolver methods for progressive multidimensional scaling of large Data. In: Kaufmann M., Wagner D. (eds). Graph Drawing. GD 2006. Lecture Notes in Computer Science, vol 4372. Berlin: Springer Heidelberg; 2007.

40. Breiman L. Random forests. Mach. Learn 2001; 45: 5-32.

41. Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random Forest? In: Perner P (editor). Machine learning and data mining in pattern recognition. 8th International Conference, MLDM 2012. Berlin: Springer Heidelberg; 2012. Pp. 154-168.

42. Wu B, Abbott T, Fishman D et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics

2003; 19: 1636-1643.

43. Liu M, Wang M, Wang J, Li D. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. Sensors Actuators B Chem 2013; 177: 970-980.

44. PanduRanga Vital T, Murali Krishna M, Narayana GVL, Suneel P, Ramarao P. Empirical analysis on cancer dataset with machine learning algorithms. Adv Intelligent Systems Computing 2019; 758: 789-801.

45. Pedregosa F, Varoquaux G, Gramfort A et al Duchesnay, scikit-learn: Machine learning in Python. J Mach Learn Res 2011; 12: 2825-2830.

46. Chou J, Chien-Kuo S, Chiu PE, Farfoura M, Al-Taharwa I. Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. J Computing Civil Engineering 2011; 25: 242-253.

47. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-Score, with implication for evaluation. Lecture Notes in Computer Science 2005; 3408: 345-359.

48. Bragatto MM, Bevilaqua-Grossi D, Benatto MT, Lodovichi SS, Pinheiro CF, Carvalho GF, Dach F, Fernández-de-Las-Peñas C, Florencio LL. Is the presence of neck pain associated with more severe clinical presentation in patients with migraine? A cross-sectional study. Cephalalgia 2019; 39: 1500-1508

49. Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. Pattern Recognit Lett 1994; 15: 1119-1125.

Table 1: Clinical and demographic features of women with migraine

		Mean (95%CI)
Demographic Features	Age (years)	42 (38-46)
	History of migraine (years)	19.8 (16.5-23.1)
Clinical Features	Migraine intensity (NPRS, 0-10)	8.3 (7.8-8.8)
	Migraine duration (hours/attack)	24.3 (19.5-29.1)
	Migraine frequency (days/month)	13.0 (4.0-21.0)
	Related-disability (MIDAS)	45.0 (27.5-62.5)
Psychological variables	HADS-A (0-21)	12.5 (11.5-13.5)
	HADS-D (0-21)	10.5 (10.0-11.0)
	STAI-trait (0-60)	25.7 (24.0-27.4)
	STAI-state (0-60)	21.7 (20.6-22.8)
PPT (kPa)	Temporalis muscle	155.0 (132.0-178.0)
	C5-C6 zygapophyseal joint	131.5 (120.0-143.0)
	Second metacarpal	190.0 (170.0-210.0)
	Tibialis anterior muscle	315.0 (287.0-343.0)
Physical Examination	JPSE Extension (degree)	4.8 (4.2-5.4)
	JPSE Cervical Rotation (degree)	6.0 (5.4-6.6)
	FHP Sitting (CVA, angle)	35.5 (34.0-37.0)
	FHP Standing (CVA, angle)	24.0 (22.5-25.5)
	CROM Flexion (degree)	51.0 (47.0-55.0)
	CROM Extension (degree)	60.0 (56.0-64.0)
	CROM Latero-Flexion (degree)	39.0 (37.0-41.0)
	CROM Rotation (degree)	63.0 (60.0-66.0)

NPRS: Numerical Pain Rate Scale; MIDAS: Migraine Disability Assessment Scale; HADS-A: Hospital Anxiety and Depression Scale - Anxiety Subscale; HADS-D: Hospital Anxiety and Depression Scale - Depression Subscale; STAI: State-Trait Anxiety Inventory; PPT: Pressure Pain Threshold; JPSE: Joint Positioning Sense Error; FHP: Forward Head Posture; CVA: Cranio-vertebral Angle; CROM: Cervical Range of Motion

Table 2: First row shows the frequency of each group based on migraine intensity subgrouping. Second row shows a typical frequency of each stratified fold, and finally, last row presents the averaged sensitivity for each group.

	Group 1	Group 2	Group 3	Group 4	Group 5
Frequency total	10	8	17	11	21
Frequency fold	1	1	2	1	2
Sensitivity (%)	0.38	0.56	37.28	1.38	67.18

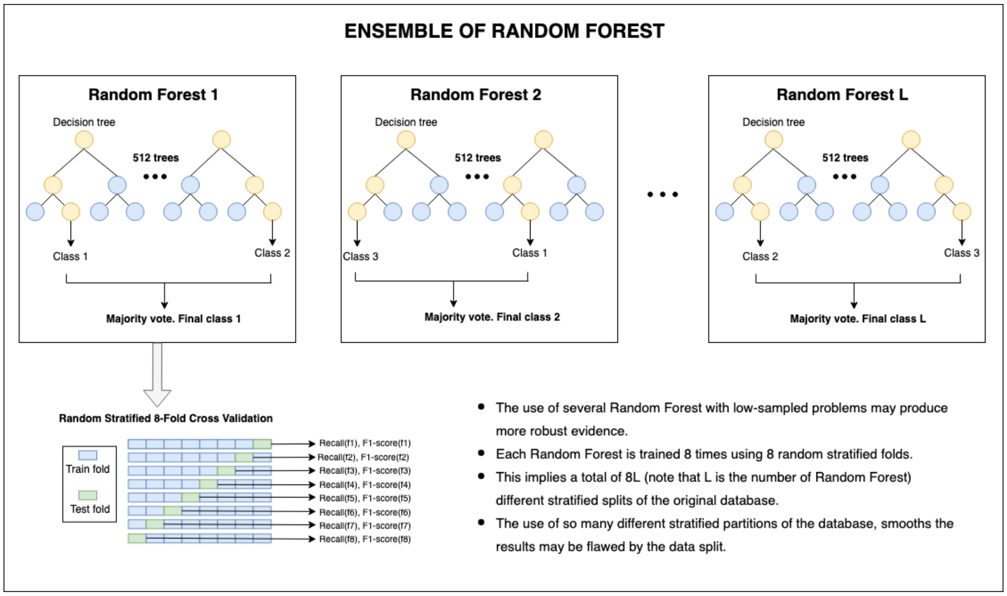
Table 3: First row shows the frequency of each group based on migraine frequency subgrouping. Second row shows a typical frequency of each stratified fold, and finally, last row presents the averaged sensitivity for each group. It is worth to mention that the Random Forest based models are not capable to discriminate patients from group 2. It is probably due to the unbalanced samples per class.

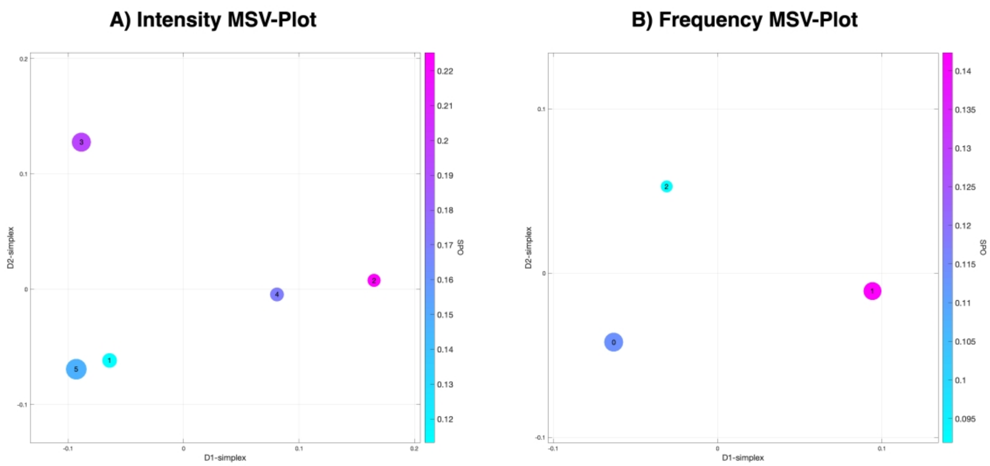
	Group 0	Group 1	Group 2
Frequency total	30	27	11
Frequency fold	4	4	2
Sensitivity (%)	61.51	34.60	0.00

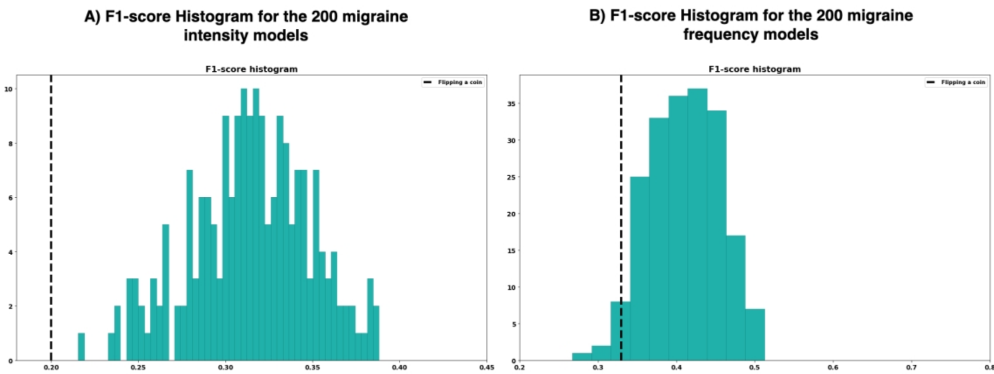
Table 4: Descriptive statistics (the percentage of relevance) of the 10 most discriminative variables for migraine intensity.

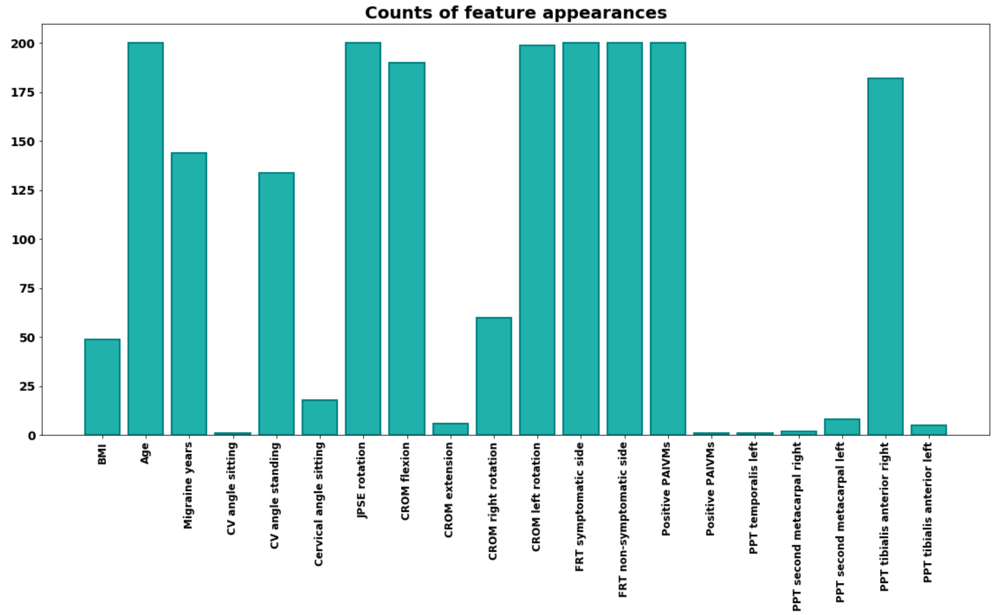
Variable	Mean (%)	Standard Deviation (%)
Age (years)	2.59	0.09
JPSE in cervical rotation (degrees)	2.53	0.08
Cervical Range of Motion in rotation (degrees)	2.30	0.07
FRT to the non-symptomatic side (degrees)	2.44	0.08
FRT to the symptomatic side (degrees)*	3.02	0.09
Positive PAIVMs	2.44	0.08
Cervical Range of Motion in flexion (degrees)	2.20	0.07
PPT Tibialis Anterior (kPa)	2.20	0.08
Years with Migraine	2.13	0.08
Cranio-Vertebral Angle Standing (degrees)	2.12	0.07

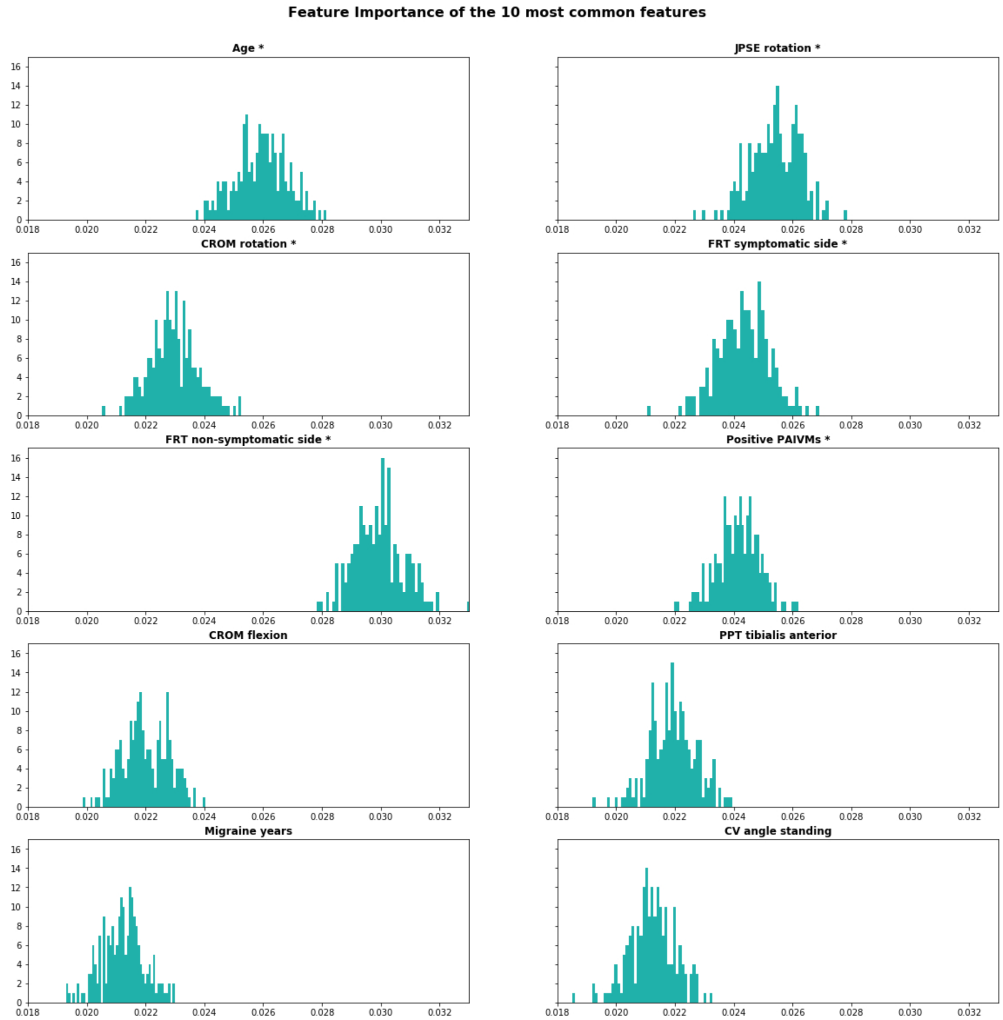
* The most discriminative variable for migraine intensity


ENSEMBLE OF RANDOM FOREST
Random Forest 1
Decision tree
512 trees
Class 1
Class 2
Majority vote. Final class 1
Random Forest 2
Decision tree
512 trees
Class 3
Class 1
Majority vote. Final class 2
Random Forest L
Decision tree
512 trees
Class 2
Class 3
Majority vote. Final class L
Random Stratified 8-Fold Cross Validation
Train fold
Test fold
Recall(f1), F1-score(f1)
Recall(f2), F1-score(f2)
Recall(f3), F1-score(f3)
Recall(f4), F1-score(f4)
Recall(f5), F1-score(f5)
Recall(f6), F1-score(f6)
Recall(f7), F1-score(f7)
Recall(f8), F1-score(f8)
The use of several Random Forest with low-sampled problems may produce more robust evidence.
Each Random Forest is trained 8 times using 8 random stratified folds.
This implies a total of 8L (note that L is the number of Random Forest) different stratified splits of the original database.
The use of so many different stratified partitions of the database, smooths the results may be flawed by the data split.


A) Intensity MSV-Plot
B) Frequency MSV-Plot
D1-simplex
D2-simplex
D1-simplex
D2-simplex


A) F1-score Histogram for the 200 migraine intensity models
F1-score histogram
Flipping a coin
B) F1-score Histogram for the 200 migraine frequency models
F1-score histogram
Flipping a coin


Counts of feature appearances
200
175
150
125
100
75
50
25
0
BMI
Age
Migraine years
CV angle sitting
CV angle standing
Cervical angle sitting
JPSE rotation
CROM flexion
CROM extension
CROM right rotation
CROM left rotation
FRT symptomatic side
FRT non-symptomatic side
Positive PAIVMs
Positive PAIVMs
PPT temporalis left
PPT second metacarpal right
PPT second metacarpal left
PPT tibialis anterior right
PPT tibialis anterior left


Feature Importance of the 10 most common features
Age *
JPSE rotation *
CROM rotation *
FRT symptomatic side *
FRT non-symptomatic side *
Positive PAIVMs *
CROM flexion
PPT tibialis anterior
Migraine years
CV angle standing