A decision support system for primary headache through machine learning

Fangfang Liu [1] , **Guanshui Bao** [Corresp., 1] , **Mengxia Yan** [1] , **Guiming Lin** [1]

[1] Shanghai Jiao Tong University School of Medicine, Shanghai Ninth People's Hospital, Shanghai, Huangpuqu, China

Corresponding Author: Guanshui Bao
Email address: baogs@163.com

Background: Primary headache is a disorder with a high incidence and low diagnostic accuracy; the incidence of migraine and tension-type headache ranks first among primary headaches. AI decision support systems have shown great potential in the medical field. Therefore, we attempt to use machine learning to build a clinical decision-making system for primary headaches. **Methods:** The demographic data and headache characteristics of 173 patients were collected by questionnaires. Decision tree, random forest, gradient boosting algorithm and SVM models were used to construct a discriminant model and a confusion matrix was used to calculate the evaluation indicators of the models. Furthermore, through univariate statistical analysis and machine learning, we finally identified the two most important characteristics for distinguishing migraines and tension-type headaches. **Results:** In the models, we give more importance to the F1 score than to the other confusion matrix-based metrics. The logistic regression model has the best discrimination effect, with the F1 score reaching 0.90 and the area under the ROC curve also being the largest at 0.95. Furthermore, we identified the most important factors of the two disorders through statistical analysis and machine learning: nausea/vomiting and photophobia/phonophobia may be potential independent factors for the identification of migraines and tension-type headaches. **Conclusions:** Applying machine learning to the decision-making system for primary headaches can achieve a high diagnostic accuracy. Among them, the discrimination effect obtained by the integrated algorithm is significantly better than that of a single learner. Second, nausea/vomiting, photophobia/phonophobia may be the most important factors for distinguishing migraine from tension-type headaches.

A decision support system for primary headache through machine learning

Fangfang Liu[1], Guanshui Bao[1*], Mengxia Yan[1]，Guiming Lin[1]

[1] Department of Neurology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
Corresponding Author: Guanshui Bao [1]
Street Address, Shanghai, Huangpuqu, 200001, China
Email address: baogs@163.com

Abstract

Background： Primary headache is a disorder with a high incidence and low diagnostic accuracy; the incidence of migraine and tension-type headache ranks first among primary headaches. AI decision support systems have shown great potential in the medical field. Therefore, we attempt to use machine learning to build a clinical decision-making system for primary headaches.

Methods： The demographic data and headache characteristics of 173 patients were collected by questionnaires. Decision tree, random forest, gradient boosting algorithm and SVM models were used to construct a discriminant model and a confusion matrix was used to calculate the evaluation indicators of the models. Furthermore, through univariate statistical analysis and machine learning, we finally identified the two most important characteristics for distinguishing migraines and tension-type headaches.

Results： In the models, we give more importance to the F1 score than to the other confusion matrix-based metrics. The logistic regression model has the best discrimination effect, with the F1 score reaching 0.90 and the area under the ROC curve also being the largest at 0.95. Furthermore, we identified the most important factors of the two disorders through statistical analysis and machine learning: nausea/vomiting and photophobia/phonophobia may be potential independent factors for the identification of migraines and tension-type headaches. **Conclusions：** Applying machine learning to the decision-making system for primary headaches can achieve a high diagnostic accuracy. Among them, the discrimination effect obtained by the integrated algorithm is significantly better than that of a single learner. Second, nausea/vomiting, photophobia/phonophobia may be the most important factors for distinguishing migraine from tension-type headaches.

Introduction

Headache is one of the most common symptoms in neurology clinics [1]. More than 50% of adults in European countries say they have suffered from headaches in the past year [2]. In China, the 1-year prevalence of primary headache is reported to be 23.8%. The prevalence of migraine was 9.3%, and that of tension-type headaches was 10.3% [3]. In North America, although headaches have a higher burden of disability than Parkinson's disease, multiple sclerosis, and epilepsy, the National Institutes of Health has the least amount of research funding dedicated to headaches. In China, due to the massive population base, patients spend 672.7 billion yuan each year because of primary headaches, accounting for 2.24% of China's GDP [4]. Although headaches do not seriously threaten the lives of patients, they can severely affect their work and quality of life, cause them to withdraw from society, and place heavy burdens on the patients' psychology, physiology and family as well as China's national economy.

Headaches are divided into primary headaches and secondary headaches. There are many causes of headaches. Due to the similarity of symptoms, it is easy for general practitioners to misdiagnose and miss these diagnoses. Second, the International Headache Society (IHS) released the latest headache classification in January 2018, which is the International Classification of Headache Disorders (ICHD-III) [5], listing more than 200 headache variants. This complicated classification is a very challenging task for general clinicians, especially for primary headache, which lacks clear laboratory and imaging examination assistance. In addition, there is no objective gold standard, so it is more difficult to diagnose and classify headaches. At last but not least, because the medical community has generally not paid enough attention to headaches in clinical practice for a long time, the proficiency level regarding the headache classification standards by clinicians is uneven. For example, "vascular headache" and “nervous headache” are still used to diagnose primary headache. Standardizing and improving the accuracy of the clinical diagnosis of headache has a long way to go.

According to reports, there are more primary headaches than secondary headaches, and the incidence of migraines and tension-type headaches ranks first among primary headaches [6]. Migraine includes migraine with aura and migraine without aura. They are typically unilateral, pulsating, and moderate to severe; daily physical activity can exacerbate these headaches, and they are often accompanied by nausea/vomiting and/or photophobia/phonophobia. Migraine without aura is a recurring headache lasting 4-72 hours. Migraine with aura is the gradual appearance of visual, sensory, or other central nervous system symptoms that can be fully recovered from on one side before the onset of headache. Tension-type headache is the most common type of primary headache; these headache attacks are not frequent, usually lasting several minutes to several days. These headaches are typically characterized by mild to moderate bilateral compression or bandlike headaches that are not aggravated by daily physical activity and are not often accompanied by nausea/vomiting, or photophobia/phonophobia. Although there is a large difference between typical migraines and tension-type headaches, the symptoms of most patients are not typical, especially tension-type headaches and migraine without aura. Thus, it is often difficult to distinguish between them. Due to the many differences in the treatment of the two disorders, misdiagnosis and missed diagnosis will inevitably delay the treatment of the patient.

At present, the development of Artificial Intelligence (AI) is in full swing. Applying expert systems and machine learning methods to each field will set off a new wave. Among these fields, machine learning has begun to be applied in medicine. Currently, the use of classifier medical decision support systems is gradually increasing and shows great potential in medical diagnosis. Automatic classifiers, which are faster than clinicians, can minimize errors in disease recognition and improve diagnostic accuracy. In addition, support vector machine (SVM) models, random forests, etc. have been used in the diagnosis of heart disease [7], breast cancer [8], prostate cancer [9], Alzheimer's disease [10], and many other diseases.

As early as 2013, Bartosz et al proposed the automatic diagnosis of primary headaches through machine learning. By comparing the diagnosis difference between advanced machine learning technology and clinicians, the computer decision support system achieved a higher diagnostic accuracy [11]. In recent years, Gilles et al proposed an end-to-end decision support system to improve the efficiency of the diagnosis, treatment and follow-up stage in the treatment of primary headaches. The decision support system includes three large components and a shared backend: a mobile application for patients, a web application for doctors to visualize the collected data, and an automatic diagnosis module. In the automatic diagnosis module, a decision tree is used for modeling [12]. Yin et al proposed a primary headache decision-making system based on international headache diagnostic criteria and conducted a four-month clinical evaluation at the International Headache Center of a tertiary hospital in Beijing. Good results have been obtained in terms of the sensitivity and specificity of this system for diagnosing headaches [13]. Considering the incomplete language rules when human experts express their knowledge, Monire et al improved the algorithm and used the Learning-From-Examples (LEF) algorithm to train the diagnostic fuzzy system, and the correct recognition rate reached 85%. They further proposed SVM- and multilayer perceptron (MLP)--based decision support systems, achieving 90% and 88% accuracy rates, respectively [14]. Although various types of research have been devoted to computer decision support systems, there are still major obstacles to their widespread use in clinical practice. Further exploration of decision support systems and the transformation of AI are our research directions.

To achieve a higher headache diagnostic accuracy, we first collected information and related characteristics of primary headache patients in neurology clinics through questionnaires and then entered them into the system. We try to find the best model by comparing various machine learning algorithms. Furthermore, through feature selection, we identified the most important factors that distinguish migraine from tension-type headaches, which will provide a basis for clinicians to quickly diagnose headaches.

Materials & Methods

This is a cross-sectional study designed to obtain a diagnostic discriminant model for migraines and tension-type headaches and to screen out the most important factors for distinguishing the two. The study was approved by the Ethics Committee of the Ninth People’s Hospital affiliated to Shanghai Jiao Tong University Medicine (approval no.SH9H-2021-T72-1), and met the requirements of the Declaration of Helsinki. Eligible patients were diagnosed with headaches between October 2019 and September 2020 at the Department of Neurology, Shanghai Ninth People's Hospital. After magnetic resonance imaging (MRI) examination of the brain excluding organic brain disease, the patients were diagnosed with primary headaches. All the patients are residents of China. We collected demographic data and headache characteristics of the patients by completing the questionnaire. Before the study, we obtained signed informed consent from the participating patients. After the questionnaire was collected, we followed up on the patient's headache improvement two weeks later over the phone to further verify the diagnosis. Finally, we included 173 patients with a definite diagnosis of primary headache (84 patients with migraine headaches and 89 patients with tension-type headaches) for research.

Discriminant model establishment

Firstly, we designed a paper questionnaire for outpatient to fill in. The questionnaire included a total of 19 questions to collect the demographic data (age, sex, occupation, height, and weight) of the patients and the headache characteristics (course, duration, nature, location, severe intensity, accompanying symptoms, triggers, alleviative way, and whether activity aggravates the headache). After analysis and modification by three experienced neurologists, the questionnaire can effectively collect patients' related information and the data obtained are reliable to a certain extent.

Furthermore, related examinations and MRI were used to rule out the patient's secondary factors. For the questionnaire information we collected, three different neurologists are invited to make a diagnose for each patient. Combined with the diagnosis and follow-up results, each patient was finally diagnosed accurately. Due to the low proportion of primary headaches such as neuralgia and cluster headaches in the collected samples, to reduce the problems caused by sample imbalance, we excluded these relatively rare types of headaches, finally including 173 patients (84 patients with migraines and 89 patients with tension-type headaches) in the study (Fig1). Each patient's headache may have multiple natures or be accompanied by multiple symptoms. Therefore, we performed a binary classification of the collected data and obtained a total of 48 variables. Considering that the incidence of many of the variables is extremely low, we first identified 10 statistically significant indicators between the two through the chi-square test: the course of the disease, whether the headache is accompanied by throbbing, whether it is located in the occiput, the severity, whether it is accompanied by nausea/vomiting, whether it is accompanied by photophobia/phonophobia, how does the headache change after activity, and how the headache is relieved. Using the above 10 feature variables, we randomly divided the entire dataset into a training set and a test set at a ratio of 8:2, which are used to build and evaluate the primary headache models. Data analysis was performed in Python (version 3.6.1). We used the decision tree [15], random forest [16], gradient boosting [17], logistic regression, and SVM [18] algorithms to construct discriminant models. The decision tree is a nonparametric supervised learning method [15]. It can summarize decision rules from a series of data with features and labels, then present these rules in a tree structure to solve classification problems. Random forest and gradient boosting are integrated algorithms that complete the learning task by constructing and combining multiple learners. Integrated learning by combining multiple learners, often achieves a significantly superior generalization performance than a single learner. Random forest, as the name suggests, builds a forest composed of many decision trees in a random manner; but there is no obvious dependency between each decision tree [16]. In contrast, gradient boosting requires a strong dependence between individual learners through the continuous decline in the loss function so that each model is built in the direction of the gradient descent of the model loss function [17]. An SVM is an algorithm that has been rapidly developed in recent years and is increasingly used in the field of biological information [18]. The purpose of this method is to find an optimal decision boundary in a multidimensional space, which can divide all the sample units into two categories and maximize the distance between two closest points in different categories. The edge point between the two closest points is called an SVM. From an academic point of view, the SVM may be the closest machine learning algorithm to deep learning. Furthermore, we calculated the F1 score, accuracy, sensitivity, and specificity as the evaluation indicators of the model through the common confusion matrix, and then measured the prediction result (receiver operating characteristic, ROC) curve and the area under the ROC curve [19]. The F1 score is the harmonic mean of the precision and recall. It is used in statistics to measure the accuracy of two classifications and assume that recall and precision are equally important. Therefore, we primarily focus on the F1 score rather than the other confusion matrix indicators and rank the results based on this indicator to identify the best discriminant model.

Feature selecting

For clinicians to quickly distinguish whether a headache is a migraine or tension-type headache, the 10 variables still have redundancies. Therefore, we identify the two variables that are most meaningful for diagnosing migraines and tension-type headaches through feature ranking. First, we adopted traditional univariate biometric analysis and then performed machine learning analysis. For the univariate test, we used the Pearson correlation coefficient (PCC) [20], and the chi-square test to compare each feature across the two groups [21]. The PCC represents the linear correlation between the elements of the two lists [20]. If the elements in the two lists are linearly correlated, the absolute value of the PCC will produce a high value close to 1; otherwise, it will be close to 0. The chi-square test checks the two features to observe the probability of the distribution occurring by chance [21]. Each feature tested will produce a p-value. Although the P-value does not represent the strength of the relationship between the two variables, it gives us a hint: the lower the p-value is, the greater certainty that the two variables are related. Furthermore, we ranked the feature importance with the random forest method [22]. The random forest model is a nonlinear decision tree combination model. It is easy to implement and has superior performance. It was once known as "the method that represents the level of integrated learning technology". Using the random forest algorithm for feature selection is superior to linear discriminant analysis and mean squared error methods for eliminating redundant features. Its main idea is to judge how much each feature has contributed to each tree in the random forest and then take the average value and compare the contribution of each feature separately. Usually, we use the Gini index as an evaluation indicator [22]. Compared with the PCC, the random forest is more capable of mining the deep correlation of data features.

Results

Patient baseline characteristics

In our study, we enrolled 242 patients with primary headache. In total, 45 patients were excluded according to the exclusion criteria. In addition, 24 patients were not followed up within 2 weeks (Fig 1). Finally, we included 173 patients (84 patients with migraines and 89 patients with tension-type headaches). For these 173 patients, we randomly divided them into a training set and test set at an 8:2 ratio. Our questionnaire collected 48 patient characteristics through 19 questions. First, we used the chi-square test to identify 10 meaningful characteristics and included them in the study (Table 1).

Model building

For the above 10 feature variables, we used the decision tree, random forest, gradient boosting, logistic regression, and SVM algorithms to construct the discriminant models. The F1 score, accuracy, sensitivity, and specificity were calculated through the confusion matrix (Table 2), and then the discrimination result curve (ROC curve) was constructed, and the area under the ROC curve were measured (Fig 2). The F1 score of the decision tree is 0.69, which is significantly lower than that of the integrated learning algorithm and SVM models. The random forest, gradient boosting algorithm, and SVM models have similar discrimination effects. The F1 scores are 0.86, 0.87, and 0.87, and the areas under the ROC curves are 0.90, 0.91, and 0.84, respectively. Logistic regression had the best discrimination effect, with the F1 score reaching 0.90 and the area under the ROC curve also being the largest at 0.95. The discrimination effect achieved by the integrated algorithm is better than that of a single learner, and the discrimination effect of logistic regression is the best.

Feature selection

For feature selection, we chose two methods: univariate statistical analysis and machine learning. For the univariate test, we used the PCC (Fig 3) and the chi-square test (Table 3) to compare each feature in the two groups and rank them according to the p-value. Through the univariate chi-square tests, we determined that the p-values of the variables indicating whether the headache was accompanied by nausea/vomiting and whether the headache was accompanied by photophobia/phonophobia were the smallest. These two variables have the greatest power in distinguishing the two disorders. The PCC further confirms the strong correlation between elements of two list. The odds ratios (ORs) for nausea/vomiting and photophobia/phonophobia were 0.4, which were higher than those of the other characteristic variables. Through a simple correlation analysis, we observe that patients with nausea/vomiting or photophobia/phonophobia are more likely to be diagnosed with migraine headaches than tension-type headaches. To confirm and explore the deeper relationship between the two disorders, we further ranked the feature importance through the random forest model (Table 4). Among them, the importance of nausea/vomiting and photophobia/phonophobia were 0.1897 and 0.1573, respectively, ranking as the top two variables.

In actual clinical work, migraine patients have more severe headaches and longer disease courses, usually accompanied by nausea/vomiting and photophobia/phonophobia. However, tension-type headaches are generally mild, and are not accompanied by nausea/vomiting and photophobia/phonophobia. Our results are consistent with clinical experience. Therefore, we further compared the headache severity and nausea/vomiting and photophobia/phonophobia (Fig 4). Compared with tension-type headaches, migraine patients are more likely to experience nausea/ vomiting and photophobia/phonophobia. Migraines were more severe and were mainly concentrated among the moderate to severe cases, while tension-type headaches were mainly concentrated among the mild to moderate cases.

Discussion

Model building

It is the current trend to apply AI to all walks of life and the application of AI to the medical field is a way for us to follow this trend. We used machine learning to identify primary headaches, which providing a starting point for advancing the transformation of AI. In this study, we established a discriminant model for the two types of primary headaches (migraines and tension-type headaches) by machine learning algorithms based on 10 indicators. Diagnosis of primary headache, which is a functional disorder without an objective gold standard for diagnosis, is very difficult. Especially for the intermediate state of these two diseases, the ICHD-III diagnostic criteria is only suitable of the diagnosis of a typical headache. For atypical headache and the intermediate headache state, many clinicians can rely only on their own clinical experience, and this subjective method inevitably has a great impact on the accuracy of disease diagnosis. In other words, the clinical diagnosis made by different clinicians is highly subjective, different and unstable. Second, some scholars believe that there may be an overlap between many primary headaches [23], where multiple headache symptoms exist simultaneously. Such overlapping headaches are common with migraines and tension-type headaches. There are differences in the treatment of each type of headache. Only a clear diagnosis can improve these treatments. This intermediate state of existence and overlapping conditions make it difficult for clinicians to diagnose primary headaches. Computer aided decision-making system is developing rapidly in various fields, naturally, the application in clinical medicine is no exception. Computer-aided diagnosis by clinicians will be the direction of future development. Previous studies have mainly focused on expert decision-making systems based on international diagnostic standards [24-26]. It is difficult to make a diagnosis based on the ICHD-III criteria for the intermediate state and the overlap of clinical diseases. Perhaps it is more efficient and accurate to diagnose diseases based on individualized learning and reasoning based on samples than a pure expert decision-making system. We build this model through machine learning and explore the differences between samples through the learning and processing of sample data. In addition, for related headache data, we can further perform cluster analysis, improve headache classification. Because of the subjective nature of the diagnosis, previous clinicians performed their own work, and often came to different conclusions for the same case. After the promotion and application of the decision-making system and through continuous learning and revision, the diagnostic criteria used by clinicians can develop in the same direction.

Feature screening

In the process of diagnosing primary headache, more than 10 variables are still redundant. In addition, various characteristic variables may contradict each other; some of the characteristics may be consistent with migraine, and other characteristics may be consistent with tension-type headache. To help clinicians quickly grasp the key points of the disease, this study further screened the 10 variables through univariate statistical analysis and machine learning to find the most important factors for distinguishing migraines and tension-type headaches: whether the headache is accompanied by nausea/vomiting and whether the headache is accompanied by photophobia/phonophobia may be potential independent predictors. In previous studies on simplified headache diagnostic criteria [27], the univariate migraine model including nausea can achieve a positive likelihood ratio of 4.8 and a negative likelihood ratio of 0.23. By including three variables for nausea, photophobia, and throbbing headache, the migraine model achieves a positive likelihood ratio of 6.7 and a negative likelihood ratio of 0.23. The ID Migraine™ screening instrument has been found to be an effective and reliable migraine screening instrument, among which disability, nausea, and photophobia provide the best performance [28]. In our research, although we did not separately screen for nausea, vomiting, photophobia, and phonophobia, in general, our results obtained through statistical analysis and machine learning are consistent with those of previous studies. Nausea/vomiting, photophobia/phonophobia, and phonophobia play a vital role in distinguishing migraines from tension-type headaches.

Inevitably, our study also has certain flaws. 1. Our discriminant model includes only the two types of headaches with the highest incidence: migraines and tension-type headaches. Although it can solve most of the problems related to clinical diagnosis of headaches, other primary headaches and secondary headaches are not included. So, adding other headache categories will be the future direction of our system. 2. The diagnosis of headache is easily affected by the clinical experience of each expert. Although we followed up with the patient's headache improvement after 2 weeks to verify the diagnosis, changes in the patient's living habits and many other factors had a certain impact on the follow-up results. 3. We mainly included headache patients who visited a doctor, leading to selection bias. Patients with milder headaches who did not go to the doctor were not included in the study. 4. Our sample size is still small, and we need to further increase the sample size to verify and test the model.

Conclusions

Primary headache is a disorder with high incidence and low diagnostic accuracy. This study used machine learning to construct a discriminant model for migraines and tension-type headaches. The discriminant effect achieved by the integrated algorithms such as the random forest and gradient boosting algorithms is better than that of a single learner, and the discrimination effect of the logistic regression model is the best. Furthermore, we identified the most important factors for the identification of the two diseases through statistical analysis and machine learning; whether the headache is accompanied by nausea/vomiting and whether the headache is accompanied by photophobia/phonophobia may be potential independent factors for identifying migraines and tension-type headaches. However, due to our small sample size, we need to increase the sample size to further verify and improve the model.

Acknowledgements

X-MY and M-GL collected the data, and contributed to the data management. F-FL conceived the study, compared the results of the biostatistics feature rankings and the machine learning feature rankings, wrote the manuscript. G-SB revised the final version of the manuscript. All authors have read and approved the final manuscript.

References

[1] Dong Zhao,Yu Shengyuan. Neuropathic pain and headache.J Chinese Journal of Contemporary Neurology and Neurosurgery,2013; 13(9) : 752-754.

[2] Stovner LJ, Andree C. Prevalence of headache in Europe: a review for the Eurolight project. J Headache Pain 2010; 11:289-299. doi: 10.1007/s10194-010-0217-0
[3] Yu S, Liu R, Zhao G, Yang X, Qiao X et al.The prevalence and burden of primary headaches in China: a population-based door-to-door survey. Headache. 2012 Apr;52(4):582-91. doi: 10.1111/j.1526-4610.2011.02061.x.
[4] Yu SY, Cao XT, Zhao G, Yang XS, Qiao XY, Fang YN, et al The burden of headache in China: validation of diagnostic questionnaire for a population-based survey. J Headache Pain. 2011 Apr;12(2):141-6. doi: 10.1007/s10194-011-0336-2.
[5] Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition. Cephalalgia. 2018 Jan;38(1):1-211. doi: 10.1177/0333102417738202.
[6] Guerrero ÁL, Rojo E, Herrero S, Neri MJ, Bautista L, Peñas ML et al. Characteristics of the first 1000 headaches in an outpatient headache clinic registry. Headache. 2011 Feb;51(2):226-31. doi: 10.1111/j.1526-4610.2010.01828.x.
[7] Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. 2020 Sep 29;10(1):16057. doi: 10.1038/s41598-020-72685-1.
[8] Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM Ensembles in Breast Cancer Prediction. PLoS One. 2017 Jan 6;12(1): e0161501. doi: 10.1371/journal.pone.0161501.
[9] Li J, Weng Z, Xu H, Zhang Z, Miao H, Chen W, et al . Support Vector Machnes (SVM) classification of prostate cancer Gleason score in central gland using multiparametric magnetic resonance images: A cross-validated study. Eur J Radiol. 2018 Jan;98:61-67. doi: 10.1016/j.ejrad.2017.11.001.
[10] Shen T, Jiang J, Li Y, Wu P, Zuo C, Yan Z. Decision Supporting Model for One-year Conversion Probability from MCI to AD using CNN and SVM. Annu Int Conf IEEE Eng Med Biol Soc. 2018 Jul;2018:738-741. doi: 10.1109/EMBC.2018.8512398.
[11] Krawczyk B, Simić D, Simić S, Woźniak M. Automatic diagnosis of primary headaches by machine learning methods. Open Medicine 2013; 8:157-165.doi: 10.2478/s11536-012-0098-5
[12] Vandewiele G, De Backere F, Lannoye K, Vanden Berghe M, Janssens O, Van Hoecke S,et al. A decision support system to follow up and diagnose primary headache patients using semantically enriched data. BMC Med Inform Decis Mak. 2018 Nov 13;18(1):98. doi: 10.1186/s12911-018-0679-6.
[13]Yin ZM, Dong Z, Kong YY.Assistant decision-making system based on international diagnostic criteria for primary headache disorders. Application Research of Computers 2019; 36:2.
[14] Khayamnia M, Yazdchi M, Heidari A, Foroughipour M. Diagnosis of Common Headaches Using Hybrid Expert-Based Systems. J Med Signals Sens 2019; 9:174-180. doi: 10.4103/jmss.JMSS_47_18.
[15] Loh WY. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2011; 1(1):14-23.
[16] Breiman L. Random Forests. Machine Learning 2001; 45:5-32.
[17] Chen T,Guestrin C. XgBoost: a scalable tree boosting system. Proceedings of KDD 2016 – the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

2016; Association for Computing Machinery (ACM); 2016: 785–794.
[18] Wu SA. Improving support vector machine classifiers by modifying kernel functions. Neural Networks 1999; 12(6):783-789. doi: 10.1016/s0893-6080(99)00032-5.
[19] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015; 10(3):0118432. doi: 10.1371/journal.pone.0118432.
[20] Benesty J, Chen J, Huang Y. On the Importance of the Pearson Correlation Coefficient in Noise Reduction. IEEE Transactions on Audio Speech and Language Processing 2008; 16:757-765.
[21] McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013; 23(2):143-149. doi: 10.11613/bm.2013.018. PMID: 23894860.
[22] Chicco D, Rovelli C, Vellido A. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. PLoS ONE 2019; 14(1):0208737.
[23] Kaniecki RG. Migraine and tension-type headache: an assessment of challenges in diagnosis. Neurology 2002; 58:S15-20. doi: 10.1371/journal.pone.0208737.
[24] De Simone R, Coppola G, Ranieri A, Bussone G, Cortelli P, D'Amico D, et al. Validation of AIDA Cefalee, a computer-assisted diagnosis database for the management of headache patients. Neurol Sci. 2007 May;28 Suppl 2:S213-6. doi: 10.1007/s10072-007-0779-z.
[25] Andrew ME, Penzien DB, Rains JC, Knowlton GE, McAnulty RD. Development of a computer application for headache diagnosis: the Headache Diagnostic System. Int J Biomed Comput. 1992 Jul;31(1):17-24. doi: 10.1016/0020-7101(92)90050-3.
[26] Roesch A, Dahlem MA, Neeb L, Kurth T. Validation of an algorithm for automated classification of migraine and tension-type headache attacks in an electronic headache diary. J Headache Pain. 2020 Jun 12;21(1):75. doi: 10.1186/s10194-020-01139-w.
[27] Martin VT, Penzien DB, Houle TT, Andrew ME, Lofland KR. The predictive value of abbreviated migraine diagnostic criteria. Headache. 2005 Oct;45(9):1102-12. doi: 10.1111/j.1526-4610.2005.00234.x.
[28] Lipton RB, Dodick D, Sadovsky R, Kolodner K, Endicott J, Hettiarachchi J, et al; ID Migraine validation study. A self-administered screener for migraine in primary care: The ID Migraine validation study. Neurology. 2003 Aug 12;61(3):375-82. doi: 10.1212/01.wnl.0000078940.53438.83.

Table 1(on next page)

Table 1 Patient baseline characteristics

Table 1 Patient baseline characteristics

Characteristics	Entire cohort				Training set				Validation set			
	Migraine	Tension-type headache	Total	P-value	Migraine	Tension-type headache	Total	P-value	Migraine	Tension-type headache	Total	P-value
Sex												
Female	20 (23.8%)	39 (43.8%)	59 (34.1%)	P=0.01	17 (24.3%)	31 (45.6%)	48 (34.8%)	P=0.01	3 (21.4%)	8 (38.1%)	11 (31.4%)	P=0.46
Male	64 (76.2%)	50 (56.2%)	114 (65.9%)		53 (75.7%)	37 (54.4%)	90 (65.2%)		11 (78.6%)	13 (61.9%)	24 (68.6%)	
Course												
Year	11 (13.1%)	38(42.7%)	49 (28.3%)		11 (15.7%)	28 (41.2%)	39 (28.3%)	P=0.00	0	10 (47.6%)	10 (28.6%)	P=0.00
Month	73 (86.9%)	51(57.3%)	124 (71.7%)	P < 0.001	59 (84.3%)	40 (58.8%)	99 (71.7%)		14 (100%)	11 (52.4%)	25 (71.4%)	
Throbbing												
Yes	17 (20.2%)	6 (6.7%)	23 (13.3%)	P=0.01	14 (20.0%)	5 (7.4%)	19 (13.8%)	P=0.03	3 (21.4%)	1 (4.8%)	4 (11.4%)	P=0.28
No	67 (79.8%)	83 (93.3%)	150 (86.7%)		56 (80.0%)	63 (92.6%)	119(86.2%)		11 (78.6%)	20 (95.2%)	25 (88.6%)	
Occiput												
Yes	22 (26.2%)	43 (48.3%)	65 (37.6%)	P=0.00	18 (25.7%)	28 (41.2%)	46 (33.3%)	P=0.05	4 (28.6%)	15 (71.4%)	19 (54.3%)	P=0.02
No	62 (73.8%)	46 (51.7%)	108 (62.4%)		52 (74.3%)	40 (58.8%)	92 (66.7%)		10 (71.4%)	6 (28.6%)	16 (45.7%)	
Severe intensity	13 (15.5%)	30 (33.7%)	43 (24.9%)		12 (17.1%)	24 (35.3%)	36 (26.1%)		1 (7.1%)	6 (28.6%)	7 (20.0%)	P=0.31
Light	44 (52.4%)		95 (54.9%)	P < 0.001	35 (50.0%)	39 (57.4%)	74 (53.6%)	P < 0.001	9 (64.3%)	12 (57.1%)	21 (60.0%)	
Medium	27 (32.1%)	51 (57.3%)	35 (20.2%)		23 (31.9%)	5 (7.4%))	28 (20.3%)		4 (28.6%)	3 (14..3%)	7 (20.0%)	
Heavy		8 (9.0%)										
Nausea/ vomiting	44 (52.4%)	16 (18.0%)	60 (34.7%)		36 (51.4%)	14 (20.6%)	50 (36.2%)		8 (57.1%)	2 (9.5%)	10 (28.6%)	P=0.01
Yes	40 (47.6%)	73 (82.0%)	113 (65.3%)	P < 0.001	34 (48.6%)	54 (79.4%)	88 (63.8%)	P < 0.001	6 (42.9%)	19 (90.5%)	25 (71.4%)	
No												
Photophobia/ phonophobia												
Yes	27 (32.1%)	4 (4.5%)	31 (17.9%)	P < 0.001	24 (34.3%)	4 (5.9%)	28 (20.3%)	P < 0.001	3 (21.4%)	0	3 (8.6%)	P=0.06
No	57 (67.9%)	85 (95.5%)	142 (82.1%)		46 (65.7%)	64 (94.1%)	110 (79.7%)		11 (78.6%)	21(100%)	32 (91.4%)	
Spark												
Yes	11 (13.1%)	3 (3.4%)	14 (8.1%)	P<0.02	8(11.4%)	2 (2.9%)	10 (7.2%)	P=0.11	3 (21.4%)	1 (4.8%)	4 (11.4%)	P=0.28

No	73 (86.9%)	86 (96.6%)	159 (91.9%)		62 (88.6%)	66 (97.1%)	128 (92.8%)		11 (78.8%))	20 (95.2%)	31 (88.6%)	

Characteristics	Entire cohort			P-value	Training set			P-value	Validation set			P-value
	Migraine	Tension-type headache	Total		Migraine	Tension-type headache	Total		Migraine	Tension-type headache	Total	
Change after activities												
Aggravate	41 (48.8%)	18(20.2%)	59 (34.1%)	P < 0.001	34 (48.6%)	15 (22.1%)	49 (35.5%)	P=0.01	7 (50.0%)	3 (14.3%)	10 (28.6%)	P=0.07
Unchanged	38 (45.2%)	62 (69.7%)	100 (57.8%)		32 (45.7%)	46 (67.6%)	78 (56.5%)		6 (42.9%)	16 (76.2%)	22 (62.9%)	
Relieve	5 (6.0%)	9 (10.1%)	14 (8.1%)		4 (5.7%)	7 (10.3%))	11 (8.0)		1 (7,1%)	2 (9.5%)	3 (8.5%)	
Alleviative way												
Persistence	9 (10.7%)	14 (15.7%)	23 (13.3%)	P=0.00	8 (11.4%)	10 (14.7%)	18 (13.0%)	P=0.00	1 (7.1%)	4 (19%)	5 (14.3%)	P=0.62
Rest	25 (29.8%)	45 (50.6%)	70 (40.5%)		18 (25.7%)	35 (51.5%)	53 (38.4%)		7 (50.0%)	10 (47.6%)	17 (48.6%)	
Drug	48 (57.1%)	25 (28.1%)	73 (42.2%)		42 (60.0%)	19 (27.9%)	61 (44.2%)		6 (42.9%)	6 (28.6%)	12 (34.3%)	
Else	2 (2.4%)	5 (5.6%)	7 (4.0%)		2 (2.9%)	4 (5.9%)	6 (4.4%)		0	1 (4.8%)	1 (2.8%)	

Table 2(on next page)

Table 2 Evaluation of the discriminant effect of various models

Table 2 Evaluation of the discriminant effect of various models

	F1-score	Accuracy	TP rate	TN rate	ROC-AUC
Decision tree	0.69	0.74	0.76	0.71	0.74
Random Forests	0.86	0.89	0.90	0.86	0.90
Gradient boosting	0.87	0.89	0.86	0.93	0.91
Logistic regression	0.90	0.91	0.90	0.93	0.95
SVM-linear	0.87	0.89	0.86	0.93	0.84

Table 3 (on next page)

Table 3 chi-square test

Table 3 chi-square test

Characteristic variable	P-value
Photophobia/phonophobia	$P < 0.001$
Nausea/vomiting	$P < 0.001$
Course	$P < 0.001$
Change after activities	$P < 0.001$
Severe intensity	$P < 0.001$
Alleviative way	$P=0.00$
Occiput	$P=0.00$
Throbbing	$P=0.01$
Spark	$P=0.02$

Table 4(on next page)

Table 4 Random forest importance ranking

Table 4 Random forest importance ranking

Characteristic variable	importance
Nausea/vomiting	0.1897
Photophobia/phonophobia	0.1573
Change after activities	0.1144
Course	0.1124
Severe intensity	0.1083
Alleviative way	0.0837
Occiput	0.0754
Spark	0.0604
Throbbing	0.0444

Figure 1

Study flow chart

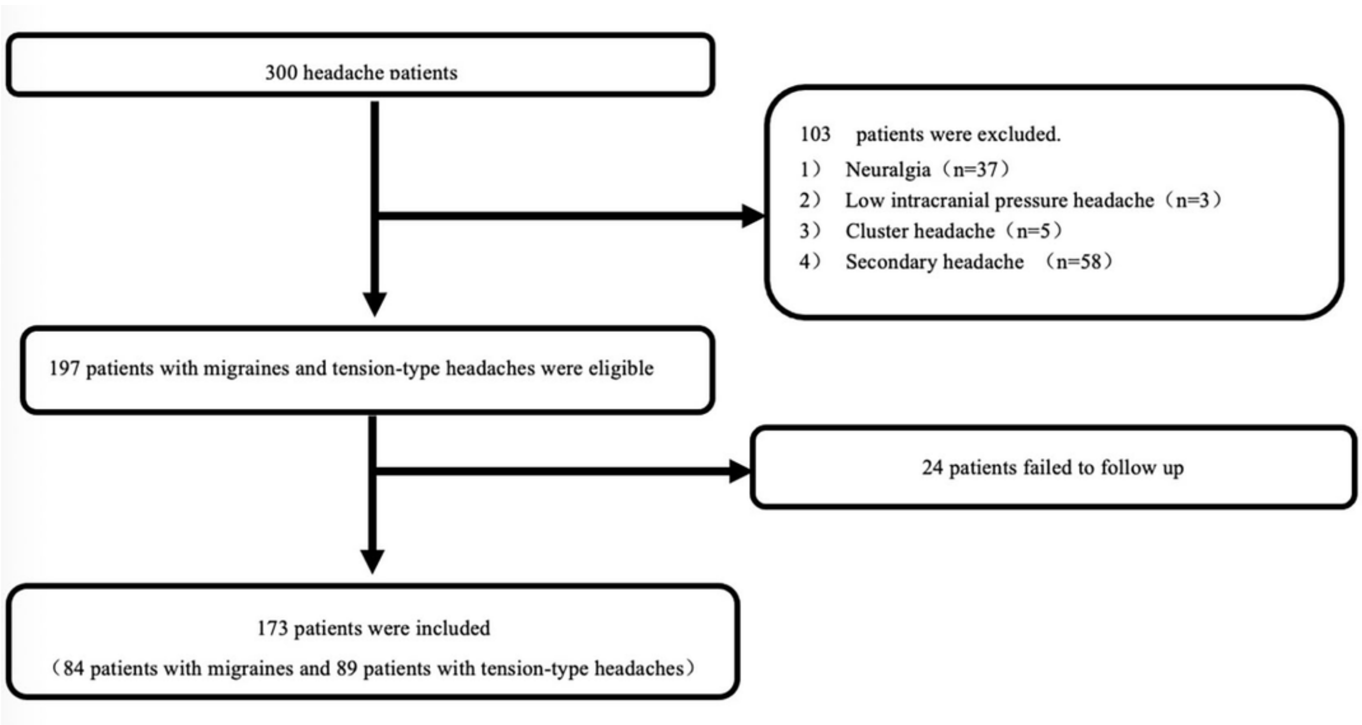

Figure 2

ROC curves of various discriminant models

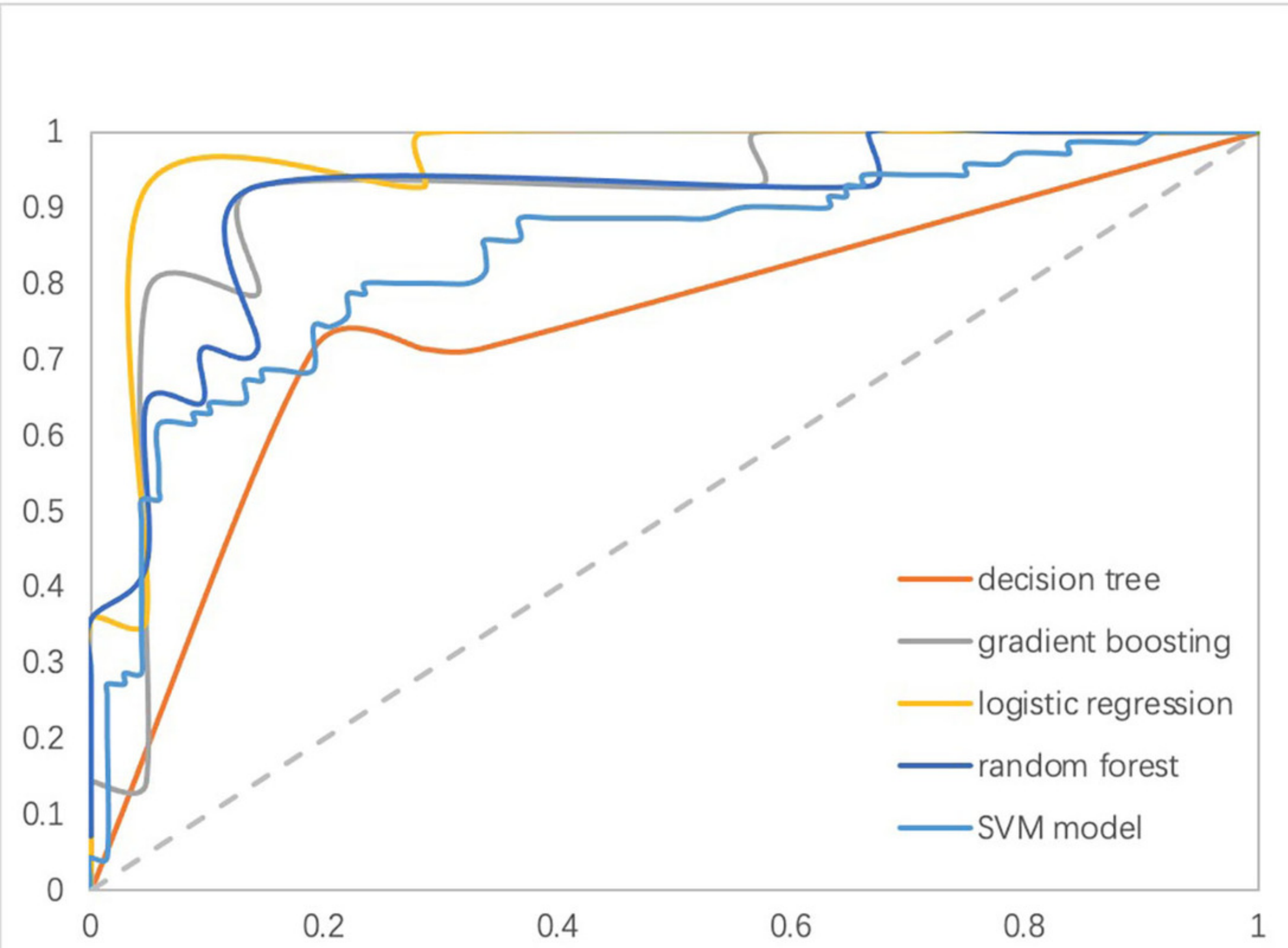

Figure 3

Pearson correlation coefficient

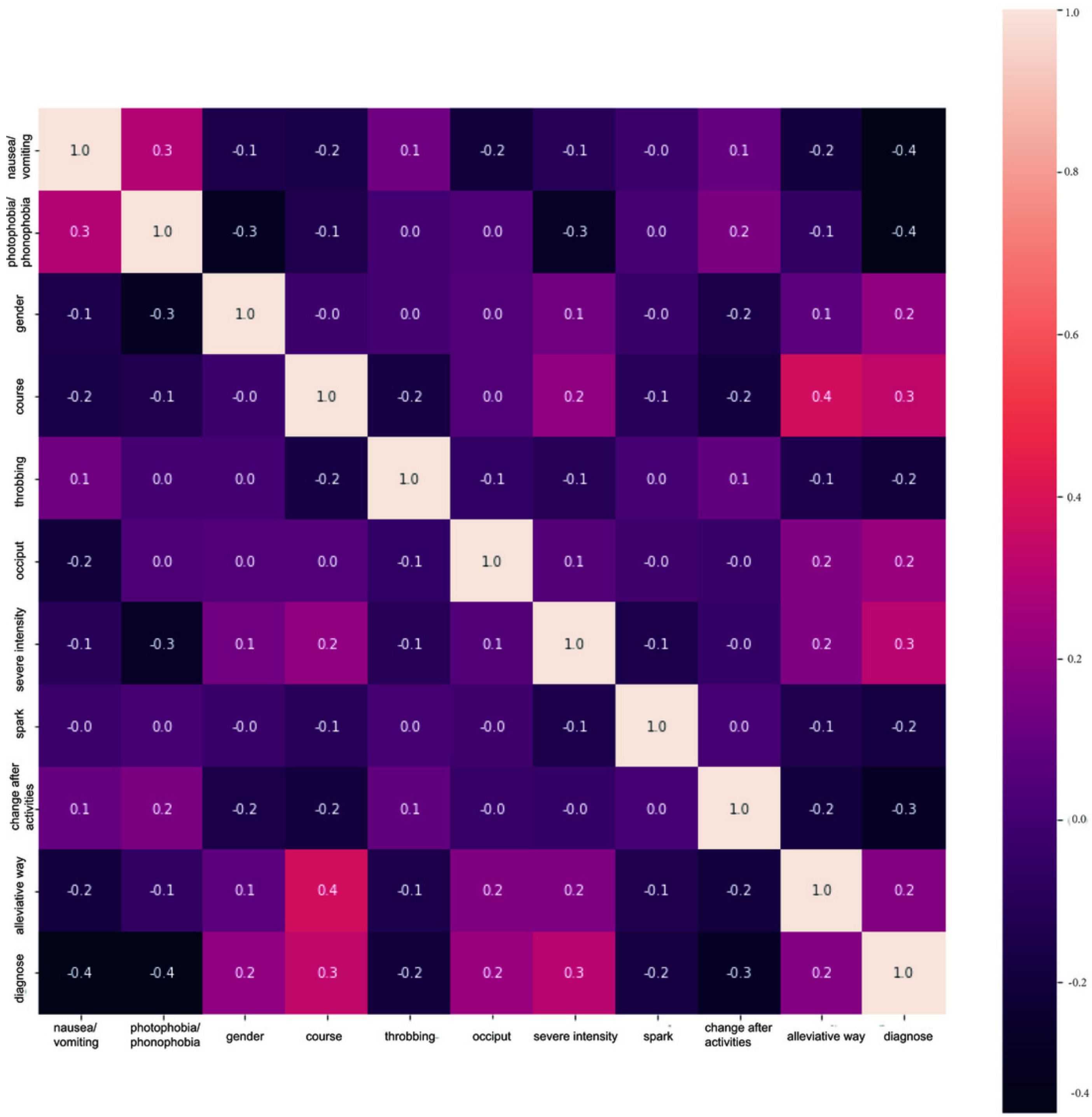

Figure 4

The correlation between headache severe intensity, nausea/vomiting, and photophobia/phonophobia