Sakarya University Journal of Science

ISSN : 2147-835X
Publisher : Sakarya University

Vol. 28, No. 4, 816-823, 2024
DOI: https://doi.org/10.16984/saufenbilder.1332882

Research Article

Improved Automatic Migraine Classification Performance with Naive Bayes

Arzum Karataş

Bandırma Onyedi Eylül University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Balıkesir, Türkiye, akaratas@bandirma.edu.tr

ARTICLE INFO

Keywords:
Migraine type diagnosis
Naïve Bayes
Classification

Article History:
Received: 26.07.2023
Accepted: 03.06.2024
Online Available: 01.08.2024

ABSTRACT

The expertise of the physician and the patient's ongoing observation are the two primary contributing factors in diagnosing migraines. However, individuals who experience migraines in the early stages frequently visit emergency rooms or different outpatient clinics, such as internal medicine, ophthalmology, and family medicine. Additionally, the type of migraine is frequently misdiagnosed due to the severity of the symptoms being misjudged or because the five-to-ten-minute examination period is insufficient for achieving an accurate diagnosis. Incorrect treatment of this type can have adverse effects on the patient's health. The majority of research in this field has concentrated on the study of brainwaves, leading to the development of complex tests that are only available to a small proportion of the population. However, one study has made progress in automatic migraine classification. The study, which demonstrates 97% classification performance above that of previous studies and produces findings in a timely manner, provides a decision support mechanism that will assist clinicians in the proper classification of migraine type. Given that over 20% of Turkey's population suffers from migraines, our study concentrated on the same issue to enhance classification performance in terms of accuracy and training time. The Naive Bayes model was employed in the study to categorize the various types of migraines, and the performance of the model was evaluated using data from actual migraine sufferers. The classification model utilized exhibited superior classification performance compared to previous studies, with 99% accuracy and precision. Additionally, the model's training time in the same dataset was the shortest when compared to other benchmarked classifier models. The application of the Naive Bayes classifier to the classification of migraines represents a highly effective technique that can facilitate rapid, accurate clinical diagnoses, thereby enabling physicians to provide their patients with precise diagnoses.

1. Introduction

Migraine is a neurological pain illness that affects millions of people worldwide [1]. It is thought of as a long-term nerve system problem [1, 2]. The International Headache Society (IHS) describes a migraine as a recurrent headache with or without aura, lasting 4–72 hours in adults and 2–72 hours in children. It is frequently accompanied by nausea, vomiting, or sensitivity to light, sound, or movement [3-6]. The illness has a high rate of impairment and a significant financial cost [3,7], and is influenced by genetic, hormonal, environmental, dietary, sleep, and psychological factors [8-10]. According to a press release from the Turkish Neurological Society in Turkey, one in five women and one in ten males, or around 20% of the population, suffer from migraines in 2019 [11]. Globally, more than a billion people were impacted by migraine in 2019 [2]. Additionally, migraine is a significant contributor to disability and job loss [1]. A substantial body of evidence indicates that migraine has a profound impact on a country's economy and health [1, 7, 12-15].

The illness burden and the rate of disability can be reduced through the early, accurate

Cite as: A. Karataş (2024). Improved Automatic Migraine Classification Performance with Naive Bayes, *Sakarya University Journal of Science*, 28(4), 816-823. https://doi.org/10.16984/saufenbilder.1332882

identification and treatment of the condition [1, 3, 7, 16]. Decision support systems can facilitate the accurate diagnosis of each patient's migraine type during the official examination process, thereby improving their quality of life. Such a strategy benefits social well-being and indirectly aids in the restoration of the nation's workforce, which has been affected by migraines.

The revised second edition of the IHS’s International Classification of Headache Disorders [16] is divided into six basic categories: migraine without aura, migraine with aura, retinal migraine, complications of migraine, probable migraine, and frequently occurring childhood periodic syndromes [17]. However, the actual dataset employed for this investigation was divided into three subgroups, comprising patients with migraines with aura, migraines without aura, and other migraines.

A correct diagnosis of the type of migraine is of critical importance, given that 20% of Turkey's population suffers from this illness [11]. In general, individuals who experience migraines in the early stages of their illness frequently seek medical attention in emergency rooms or other outpatient clinics, such as those specializing in internal medicine, ophthalmology, and family medicine. Nevertheless, some symptoms are also associated with other illnesses. It is therefore evident that diagnosis requires expertise.

In order to diagnose and detect migraines in a traditional manner, medical images obtained via machines such as CT (computed tomography), MRI (magnetic resonance imaging), and PET (positron emission tomography) are utilized in hospitals. Nevertheless, individuals may lack the financial resources to utilize these services in private hospitals or may not have timely access to them in state hospitals. For instance, in state hospitals in Turkey, it may take several months to obtain an appointment for the use of these machines. Furthermore, the interpretation of these medical images necessitates the expertise of highly trained medical professionals. However, the short examination periods of five to ten minutes in state hospitals frequently result in misdiagnosis of migraine types due to the inability to adequately assess the severity of the symptoms. The majority of research on the subject has been concentrated on the study of brainwaves, which has led to the development of complex tests that are only available to a very small portion of the population [18].

Consequently, an affordable, fast, accessible, accurate, and user-friendly approach is indispensable in the classification of migraine types. Only Sanchez-Sanchez's study [19] employed a machine learning-based methodology on symptoms rather than medical images for automatic migraine type classification, which achieved an accuracy rate of 97%. The aforementioned study employed traditional machine learning techniques, including k-nearest neighbor, decision trees (CART classification and regression tree), support vector machines (SVM), logistic regression, and artificial neural networks (ANN).

The objective of our study was to investigate whether the classification accuracy could be increased while the classification time was decreased. This would be beneficial in the context of migraine diagnosis, as an accurate diagnosis could help to reduce the potential consequences that patients may be susceptible to. To this end, we proposed a machine learning-based solution that was fast, accessible, accurate, and easy to use for clinicians who may wish to use it in their decision-making.

In this study, we used the Naive Bayes classifier, a machine learning approach, to analyze the data. In this study, we propose that Naive Bayes can enhance classification performance, particularly when the predictors are categorical. This is despite the naive assumption that the predictors are independent and identically distributed, as discussed in [20]. Furthermore, Naive Bayes can be applied to the solution of multi-class classification issues. Additionally, it is straightforward, rapid, and easy to implement. The research findings have enhanced the existing literature on the automatic migraine-type classification and have provided support for our initial intuition.

The remainder of this document will proceed as follows: The methodology employed in this research is outlined in the subsequent section, along with the workflow. Then, the dataset, classifier model, implementation environment, and performance assessment metrics utilized in the experimental investigation are described in detail. Later, the findings of the experiment will be presented. The results and contributions of the research are then discussed. The final section of the paper presents a conclusion to the research.

2. Methodology

In this study, we propose a Naive Bayes-based model to enhance the classification performance of earlier efforts. The methodology employed in this investigation is illustrated in Figure 1, which represents a straightforward supervised classification approach. First, the "Migraine dataset," as detailed in the 3.2 Dataset section, was divided into two subsets: a training set (representing 80% of the dataset) and a test set (representing 20% of the dataset). The training set was then used to train the classification model, Naive Bayes, while the test set was used to evaluate the classification performance of the trained classifier. Finally, the performance of the classifier is evaluated by means of performance metrics, such as accuracy and precision, as detailed in Section 3.3 Measures for Performance Evaluation.

This approach is designed for medical practitioners to assist them in their decision-making regarding the classification of migraine types in their patients. This method is designed to be user-friendly, requiring only the entry of migraine symptoms as specified in Section 3.2 of the dataset by the doctor or medical staff. The method then automatically predicts the migraine type.

3. Experimental Study

3.1. Naïve Bayes classifier

The Bayes theorem, a formula for calculating conditional probability, was developed by Thomas Bayes in 1812 for use in statistics. This theorem elucidates the relationship between conditional probabilities and marginal probabilities in the probability distribution for a random variable. The calculation is performed using Equation (1).

$$P(E_1|E_2) = \frac{P(E_1|E_2) \times P(E_1)}{P(E_2)} \tag{1}$$

where the components of the formula as is defined below.

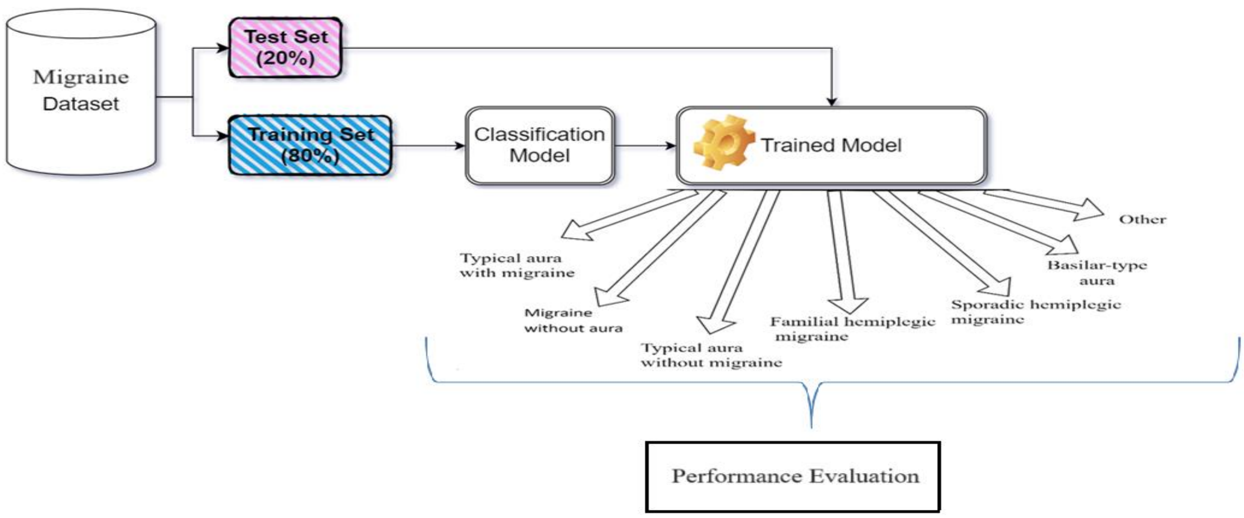

Figure 1. Methodology used in this work

- E_1 and E_2 are events.
- The probabilities of events E_1 and E_2 , respectively, are represented by $P(E_1)$ and $P(E_2)$.
- $P(E_1|E_2)$ is the probability of event E_1 occurring when event E_2 occurs. $P(E_2|E_1)$ is the probability of event E_2 occurring when event E_1 occurs.

The Naive Bayes classifier is based on the application of Bayes' theorem with the naive assumption that each pair of features will have a conditional dependency given the value of the class variable. In other words, the classifier first creates a Naive Bayes probability model and then combines it with a typically maximum posterior decision rule. In conclusion, a Bayes classifier, represented by Equation (2), can be conceptualized as a function that assigns class labels to data item properties.

$$class = \arg max_{class} \times P(class) \prod_{i=1}^{n} P(feature_i | class) \qquad (2)$$

The probability function P, the class label to be assigned, and the i^{th} att attribute of the data point are all shown in Equation (2).

3.2 Dataset

The "Migraine" dataset created by Sanchez-Sanchez is publicly available on CodeOcean [21]. The data was collected during the course of a master's thesis research project. In 2013, a medical professional maintained records at the Hospital Materno Infantil de Soledad. The dataset comprises 400 individuals' medical records and contains no personal data that could be used to identify the patients. The dataset comprises 24 variables, including both migraine symptoms and diagnoses. The variables are listed in Table 1 and have natural number values.

Furthermore, the dataset is free of missing values. The diagnosis of the patient, as indicated by the variable "Type," is based on the patient's symptoms and medical history. One of the seven forms of migraines is indicated, including "1. Typical aura with migraine", "2. Migraine without aura", "3. Typical aura without migraine", "4. Familial hemiplegic migraine", "5. Sporadic hemiplegic migraine", "6. Basilar-type aura" and "7. Other". Figure 2 presents the percentage of cases in the dataset by type of migraine. It should be noted that migraine types that can develop without headache, such as acephalgic migraine or silent migraine, are represented under other categories. Consequently, this particular type is not fully represented in the study.

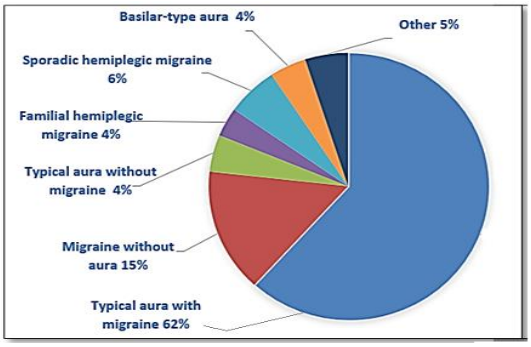

Figure 2. A pie chart illustrating the relative prevalence of different types of migraines within the dataset

3.3 Measures for performance evaluation

In this study, precision and accuracy metrics are employed to assess the performance of machine learning models. To facilitate comprehension of these measurements, it is essential to present a number of factors, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These parameters are employed to assess precision, recall, and accuracy. As illustrated in Figure 3, a confusion matrix, which is a table for a binary classifier, is a valuable tool in the classification process, which allows for the comparison of actual values with predicted values by a machine learning model.

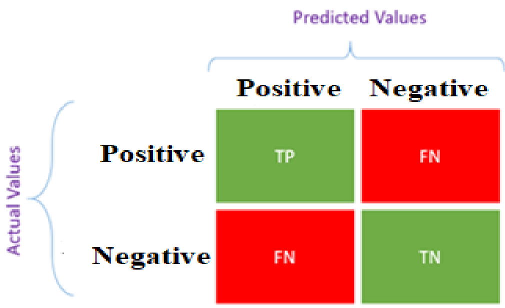

Figure 3. A confusion matrix for assessing classification performance

Table 1. Description of the Migraine dataset

#	Identified Variable	Value Range	Description
1	Age	15-77	Age of the patient
2	Duration	1-3	Duration of symptoms last episode in days
3	Frequency	1-8	Frequency of episodes per month
4	Location	0-2	Unilateral/bilateral pain location *{0: None, 1: Unilateral, 2: Bilateral}*
5	Character	0-2	Throbbing or constant pain {0: None, 1: Throbbing, 2: Constant}
6	Intensity	0-3	Pain intensity {0: None, 1: Mild, 2: Medium, 3: Severe}
7	Nausea	0-1	Nauseous feeling {0: Not, 1: Yes}
8	Vomit	0-1	Existence of vomiting {0: Not, 1: Yes}
9	Phonophobia	0-1	Noise sensitivity {0: Not, 1: Yes}
10	Photophobia	0-1	Light sensitivity {0: Not, 1: Yes}
11	Visual	0-4	Reversible visual symptoms
12	Sensory	0-2	Reversible sensory symptoms
13	Dysphasia	0-1	Lack of speech coordination {0: Not, 1: Yes}
14	Dysarthria	0-1	Disarticulated sounds and words {0: Not, 1: Yes}
15	Vertigo	0-1	Dizziness {0: Not, 1: Yes}
16	Tinnitus	0-1	Ringing in the ears {0: Not, 1: Yes}
17	Hypoacusis	0-1	Hearing loss {0: Not, 1: Yes}
18	Diplopia	0-1	Double vision {0: Not, 1: Yes}
19	Visual_defect	0-1	Simultaneous frontal eye field and nasal field defect and in both eyes {0: Not, 1: Yes}
20	Ataxia	0-1	Lack of muscle control {0: Not, 1: Yes}
21	Conscience	0-1	Jeopardized conscience {0: Not, 1: Yes}
22	Paresthesia	0-1	Simultaneous bilateral paresthesia {0: Not, 1: Yes}
23	DPF	0-1	Family background {0: Not, 1: Yes}
24	Type	1-7	Diagnosis of migraine type

True positives (TP) represent the number of outcomes that were correctly identified as positive by the machine learning model. True negatives (TN), on the other hand, refer to the number of outcomes that were correctly identified as negative. False positives (FP) are instances where a negative result was erroneously classified as a positive. Similarly, false negatives (FN) are instances where a favorable outcome was erroneously projected as a negative one. The metrics employed in this study for evaluating the performance of the model are outlined below.

Precision is defined as the degree of agreement between the predicted and actual values. The precision of the learning model's positive predictions is evaluated by measuring its accuracy. Equation (3) specifies that this metric is calculated as the ratio of true positives to all positive predictions. This is calculated as the sum of the model's true positives and false positives.

$$Precison = \frac{TP}{TP + FP} \tag{3}$$

Accuracy is a metric used to evaluate the quality of predictions. It determines the percentage of accurate predictions made by the machine learning model. For instance, if 95% of the predictions are correct, the model's accuracy is 95%. The accuracy of the model is calculated using the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{Number\ of\ correct\ predictions}{Number\ of\ total\ predictions} \tag{4}$$

4. Results and Discussion

In this study, we utilize the recommended methodology to categorize the migraine dataset. The dataset encompasses information on adults who have been diagnosed with migraines. The classification process makes use of every detected variable. To provide a fair comparison of Sanchez-Sanchez's work [19], a

Python 3 script utilizing the "Scikit-learn" [22] and "Pandas" [23] libraries was employed, as these were used in the implementation of Sanchez-Sanchez's work [19].

The precision and accuracy figures that Sanchez-Sanchez published for their studies, those of their opponents, and those that we empirically discovered for our work are presented in Table 2.

Table 2. Comparative performance outcomes in terms of accuracy and precision

Study	Classifier Model	Precision (%)	Accuracy (%)
Sanchez-Sanchez [20]	kNN	87.19	78.75
	CART	81.00	81.25
	SVM	95.31	86.25
	Logistic Regression	95.63	87.50
	ANN	97.00	97.50
Our Study	Naive Bayes	99.00	98.75

Table 2 presents the performance values for the same dataset when employing various machine learning classifier models. The weighted average of all migraine types was derived from precision calculations that were conducted independently for each form of migraine.

One of the objectives of this study is to develop a rapid and automated classification system for migraine types. To demonstrate the efficiency of the Naïve Bayes classifier in comparison to other classifiers, we quantified the training times of the classifiers. This is because the majority of the time consumed by machine learning algorithms is spent on the training stage. Table 3 presents the average training time, in seconds, required by machine learning classifiers. Training times are expressed in seconds and represent the average of ten separate executions. For the purposes of the experimental analysis, a laptop with the following configuration was utilised: The processor is an Intel Core i7 CPU @ 2.30 GHz. The processor is 64-bit and runs on the Windows 10 operating system, which has 16 GB of main memory.

Table 3. Comparative outcomes for the classification models' training times

Works	ML Classifier	Training Time (in seconds)
Sanchez-Sanchez [20]	kNN	0.041
	CART	0.035
	SVM	0.026
	Logistic Regression	29.856
	ANN	5.216
Our Study	Naive Bayes	0.002

Furthermore, this study does not encompass the full spectrum of migraine types, including vestibular migraine, menstrual migraine, abdominal migraine, and acephalgic migraine (or silent migraine), which are all classified as "others" in the dataset. Consequently, the dataset utilized in this study must be augmented in order to facilitate the identification of these migraine subtypes and other subtypes of migraines like retinal migraine.

5. Conclusion

It is a priori assumed that Naive Bayes may enhance the prediction accuracy on classification of the migraine dataset. This is based on the observation that Naive Bayes can perform well on relatively less data, can be trained in a short amount of time, and can provide highly accurate results on categorical variables, regardless of whether the variables are conditionally dependent or not. The objective of this study is to examine the potential of the Naive Bayes machine learning model to enhance the precision and accuracy of diagnosing the various types of migraine in migraine sufferers. The outcomes demonstrate that, when compared to numerous models employed for the classification of migraine types in the literature, the Naive Bayes model outperforms all other reference works in terms of precision, accuracy, and model training time.

Future studies may expand the dataset to include additional migraine types that are currently underrepresented, such as vestibular migraine, menstrual migraine, abdominal migraine, and acephalgic migraine (or silent migraine). These migraine types are all categorized as "others" in the dataset. Additionally, the dataset may be extanded by other migraine subtypes like retinal migraine.

Article Information Form

Funding
The author has no received any financial support for the research, authorship or publication of this study.

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Conflict of Interest/ Common Interest
No conflict of interest or common interest has been declared by the author.

The Declaration of Research and Publication Ethics
The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

References

[1] M. Ashina, Z. Katsarava, T. P. Do, D. C. Buse, P. Pozo-Rosich, A Özge, A. V Krymchantowski, E. R. Lebedeva, K. Ravishankar, S. Yu, S. Sacco, S. Ashina, S. Younis, T. J Steiner, R. B Lipton, “Migraine: epidemiology and systems of care,” The Lancet, vol. 397, no. 10283, pp. 1485-1495, 2021.

[2] J. Gupta, S. S. Gaurkar, “Migraine: An Underestimated Neurological Condition Affecting Billions,” Cureus, vol. 14, no.8, Art no. e28347, 2022.

[3] J. Berg, L. J. Stovner, L. J., “Cost of migraine and other headaches in Europe,” European Journal of Neurology, vol. 12, no. June 2005, pp. 59-62, 2005.

[4] P. J. Goadsby, “Recent advances in understanding migraine mechanisms, molecules and therapeutics,” Trends in Molecular Medicine, vol. 13, no. 1, pp. 39-44, 2007.

[5] P. Parisi, A. Verrotti, M. C. Paolino, A. Ferretti, F. D. Sabatino, F.D., “Obesity and Migraine in Children,” in Omega-3 Fatty Acids in Brain and Neurological Health, R.R. Watson and F. D. Meester, Eds. Cambridge, UK: Cambridge Academic Press, 2014, pp. 277-286.

[6] Headache Classification Committee of the International Headache Society, “The International Classification of Headache Disorders, 3rd edition,” Cephalalgia, vol. 38, no.1, pp.1-211, 2018

[7] T. Shimizu, F. Sakai, H. Miyake, T. Sone, M. Sato, S. Tanabe, Y. Azuma, D. W. Dodick,“ Disability, quality of life, productivity impairment and employer costs of migraine in the workplace,” The Journal of Headache and Pain, vol. 22, no. 29., Art no. 29, 2021.

[8] C. Wöber, W. Brannath, K. Schmidt, M. Kapitan, E. Rude, P. Wessely, Ç. Wöber-Bingöl, the PAMINA Study Group, “Prospective Analysis of Factors Related to Migraine Attacks: The PAMINA Study,” Cephalalgia, vol. 17, no. 4, pp. 304-314, 2007.

[9] P. T. Fukui, T. R. T. Goncalves, C. G. Strabelli, N. M. F. Lucchino, F. C. Matos, C. Fernanda, J. P. M. D. Santos, E. Zukerman, V. Zukerman-Guendler, J. P. Mercante, M. R. Masruha, “Trigger factors in migraine patients,” Arquivos de Neuro-Psiquiatria, vol. 66, no. September 2008, pp. 494-499, 2008 .

[10] J. M. Pavlovic, D. C. Buse, C.M. Sollars, S. Haut, R.B. Lipton, “Trigger Factors

and Premonitory Features of Migraine Attacks: Summary of Studies," Headache The Journal of Head and Face Pain, vol. 54, no. 10, pp. 1670-1679, 2014.

[11] Turkish Neurological Society. (2019, Jun. 23). Türk Nöroloji Derneği Basin Bülteni 22 Temmuz "Dünya Beyin Günü - Migren"[Online]. Available: https://www.noroloji.org.tr/haber/586/turk-noroloji-dernegi-basin-bulteni-22-temmuz-dunya-beyin-gunu-migren.

[12] R. Agosti, "Migraine burden of disease: From the patient's experience to a socio-economic view," The Journal of Headache and Pain, vol. 58, pp. 17-32, 2018.

[13] M. Bonafede, S. Sapra, N. Shah, S. Tepper, K. Cappell, P. Desai, "Direct and indirect healthcare resource utilization and costs among migraine patients in the United States," The Journal of Headache and Pain, vol. 58, no.5, pp. 700-714, 2018.

[14] T. Takeshima, Q. Wan, Y. Zhang, M. Komori, S. Stretton, N. Rajan, T. Treuer, K. Ueda, "Prevalence, burden, and clinical management of migraine in China, Japan, and South Korea: a comprehensive review of the literature," The Journal of Headache and Pain, vol. 20, Art no. 111, 2019.

[15] L. P. Wong, H. Alias, N. Bhoo-Pathy, I. Chung, Y. C. Chong, S. Kalra, Z. U. B. S. Shah, "Impact of migraine on workplace productivity and monetary loss: a study of employees in banking sector in Malaysia," The Journal of Headache and Pain, vol. 21, no. 68, 2020.

[16] Headache Classification Subcommittee of the International Headache Society, "The International Classification of Headache Disorders: 2nd edition," Cephalalgia, vol. 24 (Suppl 1), pp. 9-160, 2004.

[17] S. Tarantino, A. Capuano, R. Torriero, M. Citti, C. Vollono, S. Gentile, F. Vigevano, M. Valeriani, "Migraine Equivalents as Part of Migraine Syndrome in Childhood," Pediatric Neurology, vol. 51, no. 5, pp. 645-649, 2014.

[18] S. B. Akben, D. Tuncel, A. Alkan, "Classification of multi-channel EEG signals for migraine detection," Biomedical Research., vol. 27, no.3, pp. 743-748, 2016.

[19] P. A. Sanchez-Sanchez, J. R. García-González, J. M. R. Ascar, "Automatic migraine classification using artificial neural networks", F1000 Research, vol. 16, no. 9, Art no. 618, 2020.

[20] P. Domingos, M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," Machine Learning, vol. 29, pp. 103-130, 1997.

[21] P. A. Sánchez-Sánchez, J. R. García-González, J. M. R. Ascar. (2020). Migraine Classification Model [Online]. Available: https://codeocean.com/capsule/1269964/tree/v1.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, no.85, pp. 2825-2830, 2011.

[23] Zenodo.(2020). Pandas [computer software library] [Online]. Available: https://pandas.pydata.org/.