

Migraine Classification and Trigger Analysis Using Machine Learning Algorithms

Jai Singh (student)^{1*}

1. Plano West Senior High School, 5601 W Parker Rd, Plano, TX, 75093

* Corresponding author email: jai.singh3705@gmail.com

Abstract

Headaches and migraines are debilitating forms of pain that many people frequently suffer from. According to Migraine Buddy, 15% of the population in the US and Canada suffer from migraines. While most headaches can be treated with over-the-counter medication, migraines, which can last hours to days, sometimes require trips to the doctor and different prescription medications to attempt to combat them. According to the International Classification of Headache Disorders (ICHD-3), there are four primary types of migraines, each associated with different symptoms. While the causes of migraines are not fully understood, several triggers have been found to predict migraines with high certainty. However, any relationship between a migraine type and its corresponding trigger is unknown. This research leverages machine learning algorithms on a Korean headache dataset to first group migraines and headaches using clustering algorithms and then answer whether specific triggers are associated with a particular migraine type. A multinomial logistic regression model predicts migraine types from the triggers. Specific triggers, such as stress, fatigue, and menstruation, are significant predictors of all types of migraines and headaches, whereas odor, oversleeping, and cheese and chocolate are only important predictors of a subset of migraine types.

Keywords

Machine Learning; Migraine Trigger Analysis; Multinomial Logistic Regression Model; K-Means Clustering.

Introduction

Headaches and migraines are incredibly prevalent illnesses that millions of people across the world face. In the US and Canada alone, 15% of the population, or 42 million people, suffer from migraines.¹ They often struggle to get by daily. Migraines are debilitating and chronic conditions that cause throbbing or pulsing pain, light, and sound sensitivity, and, sometimes, vomiting.²

While the terms headache and migraine are often used interchangeably (in this paper, I use migraine to refer to both migraine and headache), there are many distinctions in the types each person faces. People can often have multiple types of one or both.³ Identifying what kind of headache or migraine a person has can help determine how to treat it properly. Symptoms help classify migraines, while triggers are a precursor to headaches and migraines and can help predict migraines. Triggers vary widely, and one of the goals of this research is to see if individual triggers are specific to a type of migraine and whether these triggers are a better predictor of these particular migraines.

Prevention & Treatment

Migraines and headaches can be dealt with in two ways. The first is through preventative measures and medication to stop getting the migraine or halt its progression once it begins. The second is treatment, which is done by mitigating the suffering through other medication classes like NSAIDs once it has

already started. Both medication classes come with side effects, and their chronic use can result in patients getting another type of headache known as a "medication overuse headache."⁴

Triptans

Triptans are a type of medication that is commonly used to prevent migraines. They often cause side effects such as nausea, dizziness, and coronary vasoconstriction. The most common side effects that come with the use of triptans are called "triptan sensations," which include chest tightness, paresthesia, flushing of the face, tingling, and neck pain.⁵ While the benefits of migraine prevention have been demonstrated to be powerful, many people prefer alternatives because taking the medication in large doses has been shown to have less-than-ideal outcomes.

NSAIDs

NSAIDs such as Ibuprofen and Advil and over-the-counter medications used to treat migraine pain often work to alleviate it. However, with the frequency of migraines, the side effects of daily use, according to the Cleveland Clinic, can include ulcers, liver toxicity, nausea, stroke, and heart attack.⁶ Many times, these NSAIDs only offer temporary relief. While migraines can last for days, patients either take medication for multiple days in a row or endure the pain, both of which are undesirable outcomes.

Prevention without drugs

The diagnosis and treatment of migraines start with a visit to a doctor, who classifies the ailment after studying the patient's symptoms. Apart from prescribing pain relief medications as part of a treatment plan, patients are asked to keep a migraine diary to track their daily activities to determine triggers. There are many online and phone applications to help patients record their daily activities to help them identify their triggers. Still, since daily activities could vary widely among people, the likelihood of successfully determining a trigger from this could be small and time-consuming at the very least. These apps, such as Migraine Buddy, have made keeping a journal easy; however, maintaining an up-to-date migraine journal can still be arduous.

Knowing the type of migraine or migraines one gets is useful, as one can be prescribed treatment based on the type of migraine they are suffering from. The IHS divides migraine-type headaches into four main categories: 1.) Migraine, 2.) Tension-Type Headache, 3.) Trigeminal Autonomic Cephalalgias, and then 4.) Other primary headache disorders.³ Some symptoms for this classification of a migraine include hyper- or hypoactivity, depression, food cravings, excessive yawning, neck stiffness, and fatigue.

The causes of migraines are not fully understood; however, genetics and environmental factors play a role.² While migraines seem to come out of nowhere, doctors and patients agree that there are usually some signs that a migraine attack is on its way. These indicators can reveal a pattern in your symptoms and even provide you with migraine prevention tools.⁴ Some common triggers recorded are stress, sleep pattern, hormonal changes, caffeine, alcohol, weather change, diet, dehydration, light, noise, smell, and medication overuse.

Without knowing the triggers, little non-medication prevention can be done. Knowing one's triggers is crucial in reducing migraine episodes and reliance on medications. For example, if a patient identifies a trigger, such as a particular food, they could remove it from their diet. Because the average person suffers from a migraine 8.8 times per month,¹ taking medication to prevent or treat it can have a long-term effect on the body; thus, identifying one's migraine triggers is a top priority for both patient and doctor.

Machine learning algorithms can be classified into two categories: unsupervised, which includes clustering algorithms, and supervised, which provides for regression models. The k-means clustering algorithm is an unsupervised algorithm aiming to classify x points into k clusters. This is done by repeatedly taking the centroids of a cluster and attempting to assign points to the cluster by minimizing the distance from every point to a centroid. The logistic regression algorithm models the odds of events

based on the logistic function. Predicting the odds between an event happening and not happening uses binary logistic regression, which takes the output of the sigmoid function and produces a result between 0 and 1, giving a probability. Multinomial logistic regression models allow for the classification of more than two outcomes (occurring and not occurring).

While migraines and headaches are classified based on symptoms, triggers cause the onset of different symptoms and thus can be used as predictors for these classifications.³ In past research on migraine prediction using machine learning algorithms, triggers have been shown to be good predictors of migraines.⁷ In this study, I propose using machine learning algorithms to group migraines based on their symptoms and then running a predictive algorithm on these groups to determine the importance of different triggers for each group. The data in this research contains information on symptoms and triggers for each entry. Every entry is labeled as either migraine, headache, or neither. I start by fitting a prediction model to the three predefined outcomes in this research. Next, I divided the dataset into more subgroups using a clustering algorithm. The entire data set is relabeled based on which group an entry belongs to. Finally, I fit a prediction model to these new labels. The results of prediction models are analyzed to understand the importance of each trigger.

The organization of the paper is as follows: in the next section, I analyze the migraine dataset, then I discuss the K-Means clustering algorithm and the multinomial logistic regression model and analyze the results of the model fitted to the dataset, which is followed by the conclusion.

Methods

Data Manipulation

The data was downloaded from a Korean study and first translated into English. The only two categories needed for this research were symptoms and triggers, so irrelevant and duplicate columns were removed. Singularly unavailable data points, i.e., in a log for a patient's migraine day with a missing entry on "smoking," were assumed to be left blank as they did not occur and were thus replaced with 0s.

Clustering

The k-means algorithm was used to cluster the data, using five clusters, including the "no-symptoms" cluster, in which every point would be identical. After running the clustering algorithm on the data, the results were looked at by hand to check for a proper fit. An additional cluster was added to one of the migraine sub-categories as the other two were exclusive in their symptoms. Some migraines had nausea and noise sensitivity that were not adequately accounted for.

Trigger Analysis

The multinomial logistic regression model was obtained from the stats-models package. This model was run against the original pre-processed data, which grouped the entries into three categories: "headache," "migraine," and "no symptoms." The p-values were then analyzed to determine if there was a broader relationship between the triggers and classes of ailment. Using the clustered data, the model was rerun, checking the p-values to assess if the migraine category was related to the type of relevant triggers.

Results and Discussion

The data set consists of symptoms per patient that would be used to classify each patient's migraine into different clusters. The data was pulled from a Korean study, "Analysis of Trigger Factors in Episodic Migraineurs Using a Smartphone Headache Diary Application." This dataset contains the daily diaries of 62 patients with a total of 4,579 entries.⁸

Each entry is already labeled as either migraines, headaches, or neither. Of the 4,579 entries, 336 were migraines, and 762 were headaches. The rest were no-symptom days. Some inferences can be made by looking at the data and analyzing the frequency of symptoms for both migraines and headaches [Table 1]. For example, throwing up is a symptom in 98.15% of migraine cases, making it the most prevalent migraine symptom. Other migraine-leaning symptoms include food sensitivity and nausea & vomiting, having about 90.5% of their occurrence for migraines. These fit with the common symptoms of migraines. Meanwhile, no headache symptom occurs more than 75% of the time, the most common being a dull, heavy headache (74.37%).

P-values for all symptoms except two are significant. Ache on the back of the neck and headache on both sides have a P-value of 0.97 and 0.05, respectively, which signifies their independence from migraine and headaches. A P-value of 0.01 or less is considered significant with 99% confidence.

We can determine the relevance of these triggers for each group by analyzing the percentages of migraines and headaches for each trigger [Table 2]. For both, stress is a widespread trigger, with 36.01% of migraines and 24% headaches occurring during stress. Other relevant factors between the two include fatigue, at around 24% for migraines and 19% for headaches, and sleep deprivation, at around 22% for migraines and 20% for headaches.

While most triggers have a very close spread between the two types of ailments, some triggers have greater relevance to one over the other. For example, 18.45% of migraines occurred during menstruation, an almost triple percentage over the 6.82% that caused headaches. Strong odors show a similar story, with 9.23% of migraines triggered by them, which is around double the 3.94% of headaches they began. Irregular meals also started approximately double the percentage of migraines as they did headaches.

Another notable trigger to look at is cheese and chocolate, which in the dataset are reported as a possible trigger in 1.49% of migraines and 0% of headaches. Whether it is due to the limited size of the data or some other reason, cheese and chocolate are relatively uncommon triggers. The fact that there is a considerable discrepancy between the triggers of migraines and headaches, along with some triggers having a much larger percent frequency than others, shows a correlation between the triggers for the ailment and the classification of it.

Table 1: Symptom Prevalence

	Migraine	Headache	P-value*
headache on left side	36.63 %	63.37 %	0.00
headache on right side	38.96 %	61.04 %	0.00
headache in middle	28.63 %	71.37 %	0.50
headache on both sides	34.75 %	65.25 %	0.05
headache on around eyes	41.28 %	58.72 %	0.00
ache on back of neck	30.49 %	69.51 %	0.97
unilateral headache	35.34 %	64.66 %	0.00
throbbing 1 headache	44.92 %	55.08 %	0.00
tight headache	48.15 %	51.85 %	0.00
throbbing 2 headache	36.54 %	63.46 %	0.00
dull heavy headache	25.63 %	74.37 %	0.02
headache worse with movement	48.84 %	51.16 %	0.00
food sensitivity	90.44 %	9.56 %	0.00
noise sensitivity	53.74 %	46.26 %	0.00
throw up	98.15 %	1.85 %	0.00
nausea & vomiting	90.91 %	9.09 %	0.00
sensitivity to smell	55.56 %	44.44 %	0.00
sensitivity to light	65.64 %	34.36 %	0.00
light and noise sensitivity	84.50 %	15.50 %	0.00

*chi squared analysis

Table 2: Underlying Triggers

	Migraine	Headache	P-value*
stress	36.01 %	24.02 %	0.00
oversleeping	2.38 %	2.23 %	0.10
sleep deprivation	24.40 %	18.64 %	0.00
exercise	1.49 %	1.31 %	0.00
no exercise	1.19 %	1.44 %	0.00
fatigue	22.32 %	19.95 %	0.00
menstruation	18.45 %	6.82 %	0.00
ovulation	2.08 %	0.92 %	0.29
emotional change	7.44 %	5.12 %	0.00
weather temp change	8.04 %	10.76 %	0.00
excessive sunlight	1.79 %	2.10 %	0.00
noise	8.04 %	2.62 %	0.00
inadequate lighting	1.19 %	0.26 %	0.00
odors	9.23 %	3.94 %	0.00
drinking	6.25 %	2.23 %	0.00
irregular meals	8.93 %	4.46 %	0.00
overeating	4.46 %	1.31 %	0.00
caffeine	2.38 %	2.23 %	0.00
smoking	0.60 %	1.05 %	0.74
cheese_chocolate	1.49 %	0.00 %	0.01
travel	3.27 %	0.79 %	0.00
other triggers	16.37 %	12.20 %	0.00

*chi squared analysis

K-Means Clustering

Clustering, or cluster analysis, is the process of identifying similar groups within a dataset. K-means is a type of clustering that minimizes the sum of the squared distances of cluster members and their means. [Equation 1]

$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{x \in S_i} ||x - \mu_i||^2$$

Although most symptoms are categorical, several symptoms, like headache duration and intensity index, were not. Thus, the K-Means algorithm was decided to be used on a normalized dataset. The symptoms used to cluster the 4,579 headaches were: left side, right side, middle, both sides, around eyes, back of the neck, unilateral, throbbing 1, throbbing 2, tight head, dull heavy, worsened move, food, nausea vomit, throw up, noise, smell, light, light noise, duration min, intensity, and intensity index.

The K-Means algorithm gave consistent results with five clusters. The optimal number of clusters was identified using the K-means elbow method (Table 3). The first cluster consisted of all unilateral headaches; the second cluster consisted of all headaches that were not unilateral; the third cluster consisted of all migraines with noise sensitivity; the fourth cluster was all migraines with nausea and vomiting; and the final group was the one without any symptoms.

On careful analysis of the results, a particular group stands out, which is not revealed in the K-Means analysis above. The cluster associated with migraines has some points with both nausea and vomiting

and noise sensitivity as symptoms. Therefore, another cluster was manually created with both symptoms of nausea and vomiting and noise sensitivity, while the other two clusters had either one or the other.

Table 3: **K-Means Results**

Cluster	0	1	2	3	4
No symptoms	3,483	217	119	374	386
Migraine	0	217	119	0	0
Headache	2	0	0	374	386
Symptoms (per observation)					
left side	0.00	0.40	0.43	0.63	0.00
right side	0.00	0.21	0.43	0.40	0.00
middle	0.00	0.14	0.33	0.08	0.37
both sides	0.00	0.32	0.45	0.03	0.57
around eyes	0.00	0.14	0.34	0.08	0.18
back of neck	0.00	0.10	0.45	0.15	0.30
unilateral	0.00	0.60	0.63	1.00	0.00
throbbing 1	0.00	0.52	0.47	0.35	0.19
throbbing 2	0.00	0.46	0.76	0.45	0.42
tight head	0.00	0.29	0.56	0.18	0.19
dull heavy	0.00	0.16	0.47	0.24	0.46
worsen move	0.00	0.71	0.82	0.40	0.30
food sensitivity	0.00	0.95	0.49	0.01	0.06
nausea vomit	0.00	1.00	0.53	0.01	0.06
throw up	0.00	0.15	0.18	0.00	0.00
noise sensitivity	0.00	0.15	1.00	0.13	0.21
smell sensitivity	0.00	0.16	0.55	0.08	0.13
light sensitivity	0.00	0.09	0.92	0.06	0.11
light noise	0.00	0.00	0.92	0.00	0.05
duration(min)	0.45	534.91	678.82	438.70	412.06
intensity	0.00	0.74	0.76	0.33	0.37
intensity index	0.00	4.51	6.12	3.40	3.62

Multinomial Logistic Regression

A logistic regression models the probability of an outcome by fitting the log odds for an event to a linear combination of one or more independent features [Equation 2].

$$z = \beta_0 + \sum_{i=1}^{i=N} \beta_i x_i$$

$$P(Y = 1) = \frac{1}{1 + e^{-z}}$$

The logistic regression in Equation 2 can be generalized to get multinomial logistic regression for multiple outcomes. For K possible outcomes Y_i , $i \in [1..K]$ and N predictors X, a multinomial logistic model can be written as [Equation 3]:

$$\begin{aligned}
Pr(Y_i = 1) &= \frac{e^{\beta_1} \times X_j}{1 + \sum_{k=1}^{K-1} e^{\beta_k} \times X_j} \\
Pr(Y_i = 2) &= \frac{e^{\beta_2} \times X_j}{1 + \sum_{k=1}^{K-1} e^{\beta_k} \times X_j} \\
&\dots \\
Pr(Y_i = K - 1) &= \frac{e^{\beta_{K-1}} \times X_j}{1 + \sum_{k=1}^{K-1} e^{\beta_k} \times X_j} \\
Pr(Y_i = K) &= \frac{1}{1 + \sum_{k=1}^{K-1} e^{\beta_k} \times X_j}
\end{aligned}$$

The multinomial logistic model was fitted to the dataset with three outcomes: migraine, headaches, and neither. Since the dataset already came with these labels, a regression was fitted, and the P-value of the weights is shown in Table 4. Synthetic Minority Over-Sampling Technique (SMOTE) was applied to the dataset to reduce imbalance and bias.⁹ Of the 4579 entries only 336 were labeled migraines and 762 were labeled headaches implying a bias from an overrepresentation of no symptom entries. The importance of each trigger is analyzed using their P-values.

A trigger weight with a P-value of 0.01 or less is important with 99% confidence. The results show that several triggers are important for both migraines and headaches. These triggers include stress, lack of sleep (or sleep deprivation), exercise, fatigue, menstruation, ovulation, emotional change, odors, drinking, overeating, and caffeine. Triggers like cheese and chocolate are primarily responsible for migraines while oversleeping, weather and temperature changes and irregular meals are important for headaches. The results of this analysis are encouraging due to the importance of some triggers in one group over the other. This means that if we can reclassify the headache diary data into more groups, we will get a clearer picture of which triggers are relevant for each subgroup. Next, the multinomial model is fitted to six newly labeled datasets. The criteria for new labels are shown in Table 5.

The results of the multinomial regression are shown in Table 6. It is clear from the data that stress, sleep deprivation, menstruation, and weather and temperature changes are important triggers for most migraine and headache types. The triggers particular to each migraine and headache type are:

- Weather & temp change is important for migraine type 1.
- Excessive eating is an important trigger for migraine type 2.
- Oversleeping, odors, cheese, and chocolate are important for migraine type 3.
- Irregular meals are an important trigger for headache type 4.
- Oversleeping, drinking, and smoking are important for headache type 5.

Table 4: Multinomial Regression Results (P-value)

Cluster	Migraine	Headache
stress	0.00	0.00
oversleeping	0.19	0.00
lack of sleep	0.00	0.00
exercise	0.00	0.00
no exercise	0.00	0.00
fatigue	0.00	0.00
menstruation	0.00	0.00
ovulation	0.00	0.00
emotional change	0.00	0.00
weather temp change	0.36	0.00
excessive sunlight	0.69	0.47
noise	0.17	0.92
inadequate lighting	0.19	0.26
odors	0.00	0.00
drinking	0.00	0.00
irregular meals	0.38	0.00
surfeit	0.00	0.00
caffeine	0.00	0.00
smoking	0.22	0.11
cheese chocolate	0.00	0.36
travel	0.75	0.03
other triggers	0.00	0.00

Table 5: Observation labels criteria

Label	Symptoms		
	Required		Exclude
0			headache migraine
1	migraine	noise	
2	migraine	noise	nausea
3	migraine	nausea	noise
4	headache	unilateral	
5	headache		unilateral

Table 6: Multinomial Regression Results (P-value)

Cluster	1	2	3	4	5
	Migraine			Headache	
stress	0.00	0.00	0.00	0.00	0.00
oversleeping	0.15	0.67	0.00	0.39	0.00
lack of sleep	0.00	0.40	0.00	0.00	0.00
exercise	0.11	0.00	0.00	0.00	0.00
no exercise	0.00	0.00	0.07	0.00	0.00
fatigue	0.00	0.00	0.00	0.00	0.00
menstruation	0.00	0.00	0.00	0.00	0.00
ovulation	0.12	0.00	0.00	0.00	0.00
emotional change	0.00	0.05	0.00	0.00	0.00
weather temp change	0.00	0.00	0.14	0.08	0.00
excessive sunlight	0.16	0.00	0.01	0.56	0.03
noise	0.04	0.03	0.32	0.66	0.86
inadequate lighting	0.13	0.16	0.56	0.24	0.58
odors	0.21	0.11	0.00	0.00	0.00
drinking	0.21	0.00	0.00	0.21	0.00
irregular meals	0.02	0.00	0.00	0.00	0.04
surfeit	0.08	0.00	0.78	0.00	0.00
caffeine	0.00	0.05	0.00	0.00	0.00
smoking	0.31	0.69	0.43	0.03	0.01
cheese chocolate	0.45	0.72	0.00	0.35	0.35
travel	0.09	0.85	0.94	0.17	0.10
other triggers	0.00	0.00	0.00	0.00	0.00

Conclusion

As migraines are an incredibly prevalent illness that millions face worldwide, a visit to a doctor for treatment and prevention often results in the patient keeping a daily diary of their activities to identify possible triggers for their ailment. This research attempts to answer whether there is a relationship between different migraine and headache types and triggers. Machine learning algorithms were used to analyze the database of daily headache diaries from 62 patients with 4,579 entries. When the dataset was divided into three groups labeled "migraine," "headache," and "no symptoms," it was found that stress, lack of sleep, exercise, fatigue, menstruation, ovulation, emotional change, odors, drinking, excessive eating, and caffeine are responsible for both migraines and headaches. However, oversleeping, weather and temperature changes, and irregular meals are usually only accountable for headaches, while cheese and chocolate are linked only to migraines. A clearer picture emerges when the data is further divided into six groups. The dataset is grouped using the K-Means algorithm, with symptoms like nausea and vomiting, noise sensitivity, and unilateral headache being the main clustering symptoms. The results of predicting these groups indicate the following: while stress, fatigue, and menstruation are important triggers for all migraine and headache types, weather and temperature changes are responsible for type 1 migraines, excessive eating is responsible for type 2 migraines, oversleeping, odors, and cheese and chocolate are responsible for type 3 migraines, irregular meals are responsible for type 4 headaches, and oversleeping, drinking, and smoking are responsible for type 5 migraines.

The results indicate that each type of migraine should have triggers that are particular to themselves. Since migraines are classified into four major groups with several subgroups³ a more in-depth study on a bigger dataset is necessary for a complete understanding of triggers associated with different migraine types. This research is the first step in showing the existence of a unique mapping between migraines

and their triggers. Furthermore, this will help patients suffering from migraines focus on the triggers relevant to their migraine type, saving them time and not prolonging their suffering.

Acknowledgments

I want to thank Ram Kandimalla, a graduate student at the University of Texas at Dallas, for his advice and valuable input throughout the research, and Professor Jeyakesavan Veerasamy at the University of Texas at Dallas, for his support, encouragement, and feedback.

References

1. Migrainebuddy.com. Migraine statistics in America & Canada. <https://migrainebuddy.com/infographic-migraine-statistics-in-america-and-canada/>, 2015
2. Migraine. <https://www.mayoclinic.org/diseases-conditions/migraine-headache/symptoms-causes/syc-20360201>, 2022
3. The international classification of headache disorders 3rd edition. <https://ichd-3.org>, 3, 2021.
4. Medication overuse headache. <https://americanmigrainefoundation.org/resource-library/medication-overuse/>, 2016
5. Samar Nicolas; Diala Nicolas. Triptans. <https://www.ncbi.nlm.nih.gov/books/NBK554507/>, 2022
6. Migraines in children and adolescents. <https://my.clevelandclinic.org/health/diseases/9637-migraines-in-children-and-adolescents>.
7. Rucha Dave Rebecca S. Zhu. Mygraine: Predicting migraines through various machine learning models utilizing user-inputted data. International Journal of High School Research.
8. Jeong-Wook Park; Min Kyung Chu; Jae-Moon Kim; Sang-Gue Park; Soo-Jin Cho. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. PLoS ONE 11(2): e0149577. doi:10.1371/journal.pone.0149577, 2016.
9. Smote: synthetic minority over-sampling technique. Journal of Artificial Intelligence, Vol: 16, 1, 2002.

Authors

Jai Singh is a Plano West Senior High School senior and hopes to major in Computer Engineering, focusing on efficient hardware design using ML.