Predicting Migraine Early from Fitbit Data with Deep Learning

Stephen Price
Evidation Health
Santa Barbara, CA, USA
stephenprice100@gmail.com

Raghu Kainkaryam
Evidation Health
Santa Barbara, CA, USA
rkainkaryam@evidation.com

Elizabeth Phillips
Evidation Health
Santa Barbara, CA, USA
ephillips@evidation.com

Arinbjörn Kolbeinsson
Evidation Health
Santa Barbara, CA, USA
arinbjorn@evidation.com

Luca Foschini
Evidation Health
Santa Barbara, CA, USA
lfoschini@evidation.com

Abstract

Migraine is common and debilitating. Early detection and surveillance of migraine can improve quality of life for patients with chronic migraine by providing a window of opportunity to prevent impending attacks, allow patients to adjust behaviors or their schedule to prepare for the incoming migraine, and give patients a greater sense of control. Here, we present a real-world dataset of 102 migraineurs followed for 60 days with daily migraine assessment surveys and continuous Fitbit data, describe a migraine classification task with considerations for production and patient outcomes, and show the results of a deep learning migraine prediction model. In predicting migraine on a held-out dataset, our model had a AUC ROC of 0.53 and we discuss possible explanations for the poor evaluation performance and future directions. This work aims to be a starting point for other researchers working with this dataset and migraine prediction from wearable data.

1 Introduction

Migraine is a prevalent and debilitating headache disorder characterized by recurring pulsing or throbbing pain in the head and a collection of possible symptoms including nausea, vomiting, or sensitivity to light and sound [1]. In the last decade, the 3-month prevalence of migraine in the United States has remained stable at 15% of the population and migraine is the fifth leading cause of emergency department visits [2]. Most migraineurs report severe impairment or need of bed rest during a migraine attack [3] and, globally, migraine is a major cause of disability [4].

Given how disruptive migraines can be, early prediction of migraine can improve patient quality of life through preemptive migraine treatment, better planning, and improvement in patient mental health by giving patients a sense of control [5]. Existing and new pharmaceutical treatments for acute migraine may provide relief if given prior to a migraine attack [6]. Additionally, with prediction, patients can block out time or adjust their plans to account for the increased disability that accompanies a migraine attack. Lastly, the uncertainty of when a migraine can strike can lead to anxiety symptoms

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

and unhealthy chronic alertness [7]. With migraine prediction, patients may experience greater self-efficacy and improvements in their mental health.

Several attempts have been made to predict migraine headache before onset. An early attempt made by Giffin et al. in 2003 examined the ability of patients to self-predict migraine headache from premonitory symptoms (i.e. non-headache symptoms which can occur before migraine headache) [8]. Migraines with premonitory symptoms were able to be correctly predicted 68% of the time six or more hours before the headache and 19% or more twenty-four or more hours before the headache with yawning and difficulties with speech/reading as the most predictive symptoms [8]. Houle et al. in 2017 developed a headache prediction model using daily stress and headache data which could predict next day headache with an AUROC of 0.65 [9]. More recent work has leveraged passively collected wearable data to predict migraine. Using Empatica E4 sleep time data and migraine data, a 2018 study on seven participants by Siirtola et al. was able to predict migraine from one night prior with a balanced accuracy of 84% [10].

In this work, we aim to improve migraine surveillance by predicting migraine early with real-world, longitudinal Fitbit data. Here, we apply deep learning to predict migraine onset from a dataset of n=102 migraineurs with 60 days per participant. Siirtola et al. [10] is the only prior work that attempts to use wrist actigraphy to predict migraine onset, but we expand on this work with a substantially larger dataset on a common consumer wearable device and a classification task that generalizes to predicting migraine in any 12 hour window. Additionally, in order to emulate real-world settings, we ensure a break between our input prediction data and the target window of interest to allow for an opportunity for the model to warn the user of an impending migraine.

2 Data

2.1 Population and Recruitment

The migraine dataset was collected from a subset of participants in a larger, one-year, case-control observational study of people with chronic pain (DiSCover). The DiSCover protocol was approved by the committee on research ethics at the Western Institutional Review Board and was conducted in accordance with the Declaration of the World Medical Association (www.wma.net). As a part of DiSCover, n=507 (n=253 case, n=254 control) provided informed consent and enrolled in the migraine substudy. For this substudy, the inclusion criteria for the case included self-reported migraine experiences, completion of the one month survey, and at least 50% compliance to their daily 1Click survey questions during the first two months of the chronic pain study asking if they have experienced recent breakthrough pain. After eligibility was determined, participants were given an option to opt into the migraine substudy.

2.2 Data Description

For this study, we only consider the case participants who were each followed for 60-days (the controls were followed for 30-days). During this 60-day period, the case participants wore their Fitbit and were asked a daily 1click email survey: "Did you experience a migraine headache in the last 24 hours?". If the participant answered affirmatively, they were directed to a follow-up survey which included questions about when and how severe the patient's migraine episode was.

Each participant used their own personal Fitbit device which may or may not have a heart rate sensor. In this analysis, we only consider data from the case participants with both available heart rate and survey data. There were a total of n=507 patients enrolled in the migraine substudy. Of the n=306 participants with Fitbit data, there were n=130 case participants (i.e. had self-reported migraine), and of these, n=102 had Fitbit with heart rate data. All participant data was de-identified prior to analysis.

The plot of survey and migraine responses across time for these n=102 participants are shown in Figure 1.

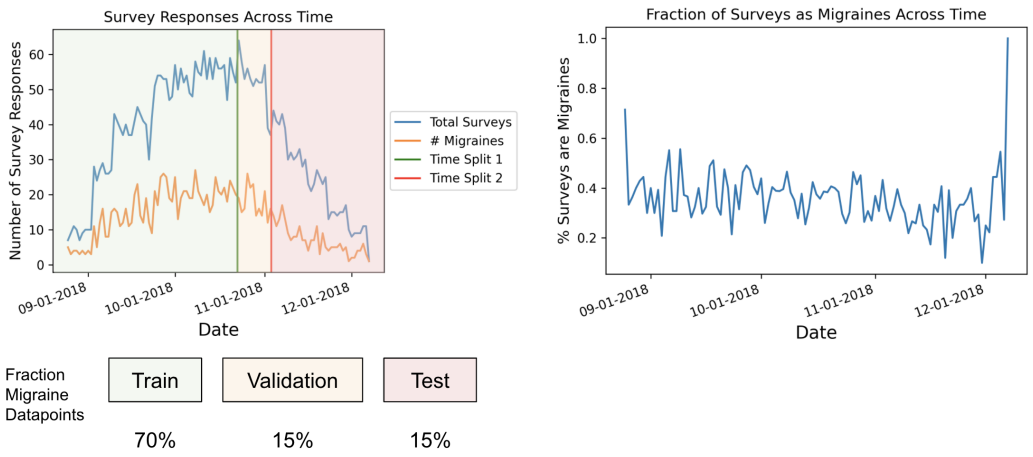

Figure 1: The left panel shows the distribution of survey responses and number of reported migraines across the study duration. Additionally, the left panel shows the time splits which were used to determine the train, validation, and test sets. The right panel shows the fraction of surveys on any day that report migraine.

3 Migraine Prediction

3.1 Datapoint Generation

We set up the classification task with three parameters: input window start, input window time end, and target window length. These define the input window, gap, and target window which together is called the sliding window. The sliding window slides along the participant time series with a step size of one hour, and for each position along the participant time series, we attempt to create a datapoint. A positive datapoint (i.e. migraine datapoint) is created whenever the target window overlaps with the start of a migraine period (assessed via the daily migraine surveys) and the input window contains Fitbit data that meets the adherence threshold of 80% (i.e. the Fitbit is worn at least 80% of the time). A negative datapoint (i.e. non-migraine datapoint) is created whenever the target window plus a 10 hour buffer on either side does not overlap with a migraine period as assessed by the daily migraine surveys and the input window contains Fitbit data that meets the adherence threshold of 80%. Once we create all our datapoints, we train a model on the input data to predict the binary class migraine or non-migraine for the target window. For this task, we set the parameters to be: input window start = -72, input window end = -6, and target window length = 12. This means we take 72 hours to 6 hours before the start of a 12 hour window as input to predict whether a migraine episode starts in the 12 hour window. This classification setup will provide migraineurs with a 6 hour warning prior to the start of a migraine attack.

3.2 Classification Task

For the classification task, we take all users and use two time splits to split the dataset into train, validation, and test set such that the number of migraines in train:val:test is 70:15:15. As shown in Figure 1, the time before the first time split is the training set, the time between the first and second time split is the validation set, and the time following the second time split is the held-out test set.

3.3 Minute-level Features

The Fitbit data consists of a sleep, steps, and heart rate value for each minute. Sometimes the Fitbit is unable to produce a value for one or all of the three channels and the values are missing. To allow our model to clearly identify missing values, for each channel, we create an additional missingness channel that has a binary value of 1 if its corresponding sensor channel is missing and 0 otherwise. In the primary sensor channel, we fill in any missing values with 0. Additionally, we create time

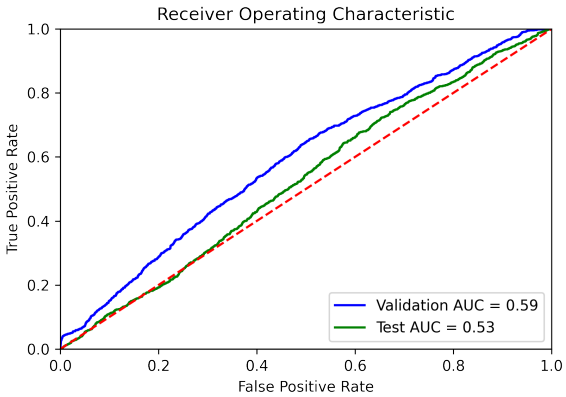

Figure 2: The receiver operating characteristic (ROC) curves of the validation and test set.

feature channels to allow the model to identify the minute of the day and the day of week. To do this, we create a sine and cosine transform channel for each of the minute of the day and the day of week feature for a total of four additional time feature channels. The sine/cosine transforms give the model a cyclical pattern so that the end and start of each period (i.e. day or week) are adjacent values in each of the sine/cosine transform while also ensuring that each value in the period has a unique sine/cosine coordinate. To reduce dimensionality, we then aggregate our data into hour blocks. We do this by taking the sum of sleep, sleep missingness, heart rate, heart rate missingness, steps, and steps missingness. We additionally take the standard deviation of steps and heart rate for the hour period. Lastly, we take the first minute of the hour's value of all the time features. This yields twelve features per hour.

3.4 Modeling

For our task, we use the InceptionTime model presented by Fawaz et al. in 2020 [11]. This model is an ensemble of deep convolutional neural networks and has strong performance on classifying time series data while also being scalable to large inputs. To train, we leveraged the tsai library's implementation of InceptionTime (i.e. InceptionTimePlus) with a fully connected dropout parameter set to 0.4 [12].

4 Experiment and Results

To ensure our model is sufficiently sensitive to migraine instances, we balanced the training set by undersampling the majority class, fine tune our model selection on the untouched, imbalanced validation set, and report our results on our untouched test set. Our balanced training set has 6,907 negative datapoints and 6,907 positive datapoints, and our validation set has 11,029 negative datapoints and 1,085 positive datapoints.

The InceptionTime model was trained on ten epochs with a learning rate of 4E-5, and the model with the best AUC ROC was chosen. This model on the validation set had a precision = 0.11, recall = 0.60, and an AUC ROC of 0.59. When evaluated on the test set, this model had an AUC ROC of 0.53. Both ROC curves are shown in Figure 2.

5 Discussion

This paper introduces a novel and to-be-made public real-world migraine and Fitbit dataset. We present a classification task and model to predict the presence of a migraine in any 12 hour window six hours early to enable the migraineur to act on an impending migraine attack.

We find low precision and AUC ROC with the InceptionTime model on our classification task. Additionally, the model's AUC ROC drops substantially from 0.59 to 0.53 when moving from the validation to test set. We hypothesize this performance drop may be due to the distribution of survey

responses as shown in Figure 1. Specifically, given the staggering of participants over this four month period, there are many participants who were trained and validated on, but were not represented in the test set. Future work may consider splitting train, validation, and test within user rather than across all users.

Our hope is that this work will act as a starting point for future work on this dataset and the migraine prediction task. Future work should try other models, personalized approaches, and a more thorough search of the parameter space for model fine-tuning. Additionally, while this work looked at predicting migraine with only passively collected data, there is an opportunity for future work to look at including demographic features present in the dataset to enhance the migraine prediction performance. Migraine prediction done poorly can lead to false positives, greater anxiety, and overtreatment of migraine patients. As a result, improving these results are important for giving models that can be deployed to positively impact patients and improve their quality of life.

References

[1] Headache Classification Committee of the International Headache Society (IHS). (2013). The International Classification of Headache Disorders, 3rd edition (beta version). Cephalalgia : an international journal of headache, 33(9), 629–808. https://doi.org/10.1177/0333102413485658

[2] Burch, R., Rizzoli, P., & Loder, E. (2021). The prevalence and impact of migraine and severe headache in the United States: Updated age, sex, and socioeconomic-specific estimates from government health surveys. Headache: The Journal of Head and Face Pain, 61(1), 60-68.

[3] Lipton, R. B., Bigal, M. E., Diamond, M., Freitag, F., Reed, M. L., & Stewart, W. F. (2007). Migraine prevalence, disease burden, and the need for preventive therapy. Neurology, 68(5), 343-349.

[4] Stovner, L. J., Nichols, E., Steiner, T. J., Abd-Allah, F., Abdelalim, A., Al-Raddadi, R. M., ... & Murray, C. J. (2018). Global, regional, and national burden of migraine and tension-type headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology, 17(11), 954-976.

[5] Buse, D. C., McGinley, J. S., & Lipton, R. B. (2020). Predicting the Future of Migraine Attack Prediction.

[6] Silberstein, S. D. (2015). Preventive migraine treatment. Continuum: Lifelong Learning in Neurology, 21(4 Headache), 973.

[7] Rutberg, S., & Öhrling, K. (2012). Migraine–more than a headache: Women's experiences of living with migraine. Disability and rehabilitation, 34(4), 329-336.

[8] Giffin, N. J., Ruggiero, L., Lipton, R. B., Silberstein, S. D., Tvedskov, J. F., Olesen, J., ... & Macrae, A. (2003). Premonitory symptoms in migraine: an electronic diary study. Neurology, 60(6), 935-940.

[9] Houle, T. T., Turner, D. P., Golding, A. N., Porter, J. A., Martin, V. T., Penzien, D. B., & Tegeler, C. H. (2017). Forecasting individual headache attacks using perceived stress: development of a multivariable prediction model for persons with episodic migraine. Headache: The Journal of Head and Face Pain, 57(7), 1041-1050.

[10] Siirtola, P., Koskimäki, H., Mönttinen, H., & Röning, J. (2018). Using sleep time data from wearable sensors for early detection of migraine attacks. Sensors, 18(5), 1374.

[11] Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., ... & Petitjean, F. (2020). Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery, 34(6), 1936-1962.

[12] Oguiza, I. (2020). tsai - A state-of-the-art deep learning library for time series and sequential data. Github.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Discussion.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Discussion.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]

(b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] The data and code are currently proprietary.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section Data and Section Experiments and Results
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] The paper's focus is on the dataset and classification task setup. The model's results are presented as the first pass.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [Yes] Specified that these will eventually be made public.
   (c) Did you include any new assets either in the supplemental material or as a URL? [No] These will be made public in the future.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See Population and Recruitment.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Data Description.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No] IRB name was provided.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No]