Predicting Migraines with Machine Learning and Feature Selection

Rebecca Zhu[1], Yan Luo[2]

[1]Nashua High School South, 36 Riverside Street, Nashua NH, 03062, USA
[2]University of Massachusetts Lowell, 220 Pawtucket Street, Lowell MA, 01854, USA

***Abstract*—Often seen as severe headaches, migraines lower the quality of life for patients. To combat ineffective monitoring applications that can only provide information about previous attacks and not future occurrences, multiple types of machine learning models such as the Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) models were trained to predict migraines occurrences. The data used was augmented by employing Synthetic Minority Oversampling Technique (SMOTE). By using feature selection methods, more discriminative features were selected to train the models. The RF model outperformed existing models with a classification accuracy of 0.9924 on the testing data.**

I. Introduction

Migraines are the third most prevalent and sixth most disabling illness in the world [1]. Frequently, migraines are associated with sensitivity to light and loud sounds, nausea, and vomiting [1]. Currently, many migraine tracking diaries only allow patients to track retrospective information and provide no information to prevent future migraines [2]. However, machine learning models such as the Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) models can be utilized to predict future migraine occurrences, allowing patients to take preventative measures. These models were created using Python and the scikit-learn library [3].

II. Methods

The dataset used was called "Analysis of Trigger Factors in Episodic Migraineurs Using a Smartphone Headache Diary Applications," and was published on Plos One in 2016 [4]. There were numerous migraine triggers or features that could increase a chance of a migraine, such as "stress," "excess_sleep," "smoking," etc. The dataset contained 4,679 entries including both migraine and no-migraine days.

After preprocessing the data, several feature selection methods such as the Random Forest Feature Importance attribute (RFFI), Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and the novel Jenson-Shannon Divergence ranking values (JSD) were applied to the models to select more discriminative data to train on by dropping features with lower importances [5]. Additionally, the dataset was severely imbalanced, with 4,243 no-migraine occurrences and 336 migraine occurrences. To mitigate the bias in the dataset, Synthetic Minority Oversampling Technique (SMOTE) was utilized, where the number of under sampled instances of a class is increased using the k-nearest neighbors algorithm [6].

The models were evaluated on 20% of the dataset and the other 80% of the dataset was used to train the models. The metrics used to evaluate the results of the research and their formulae are shown below in Equation 1:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN},\ F1 = \frac{2*precision*recall}{precision+recall} \quad (1)$$

where *TP*=true positive, *TN*=true negative, *FP*=false positive, and *FN*=false negative migraine occurrences. F1 scores were utilized because accuracy by itself does not provide detailed insight about the performances especially if the dataset is imbalanced. Thus, F1 scores are calculated based on precision and recall metrics. Additionally, macro-average scores were used because they do not take class imbalance into consideration, which allowed the results to be compared before and after applying SMOTE. *Figure 1* depicts the mechanism used for all three ML models.

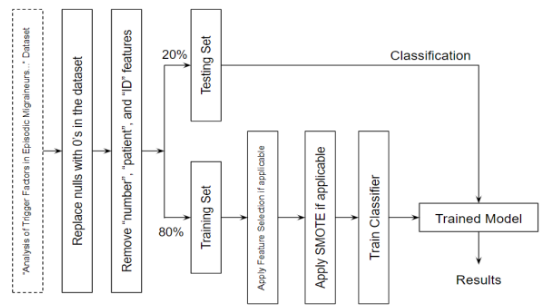

Figure 1. The proposed approach for the ML models to classify migraines.

III. Results and Discussion

The hyperparameters for each ML model were found through grid search on the training data. The classification metrics were found on the testing data. Specifically, the RF model contained 30 trees and the SVM model used a polynomial kernel function. In Table 1 without SMOTE, it was evident that the RF and SVM models outperformed the LR model, as the RF and SVM models both achieved an accuracy of 0.99017. However, the RF model achieved a higher macro-average F1 score than the SVM model because it predicted fewer false negatives.

TABLE I. Classification Metrics for ML Models

ML Model SMOTE?	*RF Model* No	Yes	*LR Model* No	Yes	*SVM Model* No	Yes
Accuracy:	0.9902	**0.9913**	0.9571	0.9673	0.9902	0.9913
Macro F1 Score:	0.9707	**0.9741**	0.8918	0.8884	0.9704	0.9740

While the RF and SVM models performed well before applying SMOTE, there was an inherent bias because the number of no-migraine occurrences was more than 12 times the number of migraine occurrences. To see if the imbalance in the classes caused false negatives, SMOTE was applied. Without SMOTE, it seemed obvious that the model could have kept predicting no-migraine and getting a high accuracy. Thus, other metrics such as F1 scores were considered in addition to accuracy. After applying SMOTE, there were 3,409 instances of the migraine and no-migraine class each. However, SMOTE was only applied on the training data so that the testing data is still the original data. As shown in Table 1 with SMOTE, all three models had an improvement in accuracy and F1-score. Because macro-average scores do not take class imbalance into consideration, the improvement in F1 scores means the RF and SVM classifiers improved in the ability to predict fewer false negatives. Additionally, the RF model performed the best out of the three models, achieving an accuracy and macro-average F1 score of 0.9913 and 0.9741 respectively with SMOTE.

Since the RF model achieved the highest result, it was used to evaluate the four feature selection methods. Running feature selection methods on the RF classifier without SMOTE was not performed because the RF classifier with SMOTE already produced better accuracy and F1 metrics, which were evaluated before feature selection.

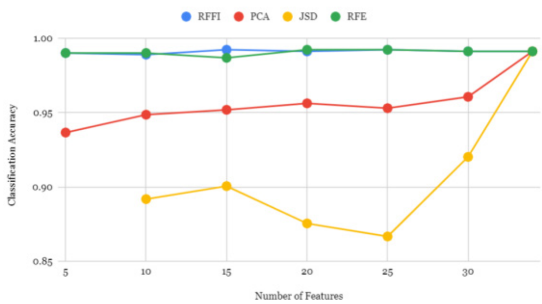

Figure 2. Comparing different feature selection methods using the RF model with SMOTE.

Figure 2 shows the classification accuracy of each method according to different numbers of features, namely 5, 10, 15, 20, 25, 30, and 34 or all features. It is observed that feature selection with PCA and JSD degrade in performance when fewer features are selected. The best performance with the RF model using SMOTE and feature selection was when the top 15 features were selected using the RFFI method, which achieved an accuracy of 0.9924, an increase from before applying feature selection. When 15 of the most discriminative features were selected by the RFFI method, the RF model achieves the best results. Therefore, around half of the features in the dataset are crucial in training the RF model for a more accurate migraine prediction.

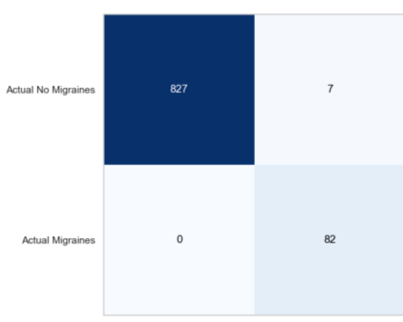

Figure 3. The confusion matrix for the RF model with SMOTE and top 15 features selected by RFFI.

Figure 3 shows the confusion matrix (CM) or the distribution of TN, TP, FP, and FN for migraine prediction. The top left and bottom right corner depict the number of correctly predicted migraine and no migraine occurrences. After applying SMOTE and the RFFI feature selection method, however, the model predicted zero migraines as no-migraines, which meant the false-negative rate fell to zero. Additionally, the model only inaccurately predicted seven no-migraines as migraines, which was similar to the metric before applying SMOTE and feature selection. Thus, SMOTE and RFFI were able to decrease the false-negative rate while maintaining the false-positive rate, which meant the overall classification performance increased.

Figure 4 shows the importance of the top 5 features selected by RFFI, which included "nausea_vomiting," "helping_factors," "sound_sensitivity," "light_sensitivity," and "rest." The most important factor or "nausea_vomiting" had an importance of 0.4 evaluated by the RFFI method, with the second most important feature with an importance of 0.19. These importances measure the contribution of the features to the classification result.

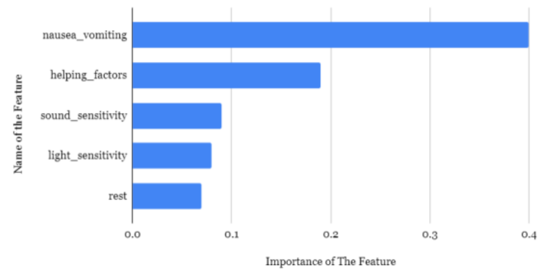

Figure 4. The top 5 features given by RFFI attribute feature selection.

IV. Conclusion

A demand exists for accurate migraine prediction through ML. All models achieved high results, but the best model was the RF model with SMOTE and the top 15 features from RFFI. This model achieved an accuracy of 0.9924 and a macro-average F1 score of 0.9774, which outperformed existing models on the standard dataset. The research showed that SMOTE improved the results of all classifiers and the RFFI feature selection method worked best, which decided the most important feature in predicting migraines was nausea or vomiting. In the future, an application where users can plug their symptoms, health factors, and biometrics could be created where the ML algorithm could predict in real-time whether or not a patient would get a migraine.

V. Acknowledgements

I would like to thank Professor Yan Luo from UMass Lowell's ECE Department for his valuable feedback and help.

References

[1] Migraine Research Foundation. (n.d.). Migraine Facts. Retrieved from https://migraineresearchfoundation.org/about-migraine/migraine-facts

[2] Migraine Buddy. (n.d.). Migraine Buddy. Retrieved from https://migrainebuddy.com/

[3] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

[4] Min Kyung Chu, J.-M. K. (n.d.). Analysis of Trigger Factors in Episodic Migraineurs Using a Smartphone Headache Diary Applications. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149577

[5] Karl Pearson F.R.S. . (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

[6] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Int. Res. 16, 1 (January 2002), 321–357.

[7] Google Research. (2017, April 6). Federated Learning: Collaborative Machine Learning Without Centralized Training Data. Retrieved from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html