Migraine Attack Prediction Using Patient-Reported Data

Prabhat Mattaparthi (pkm5487@psu.edu) & Rohan Samuel Sampath (rss5845@psu.edu)

DS402: Machine Learning for Healthcare

Migraine is a complex neurological condition characterized by debilitating attacks often precipitated by time-delayed triggers such as stress accumulation, sleep irregularities, and dietary factors. Accurately predicting these attacks requires analyzing temporal dependencies in patient data, a task for which traditional static models are often ill-suited. In this project, we proposed a Deep Learning framework using **Bidirectional Long Short-Term Memory (Bi-LSTM)** networks to predict daily migraine risk. We utilized a high-fidelity simulated dataset of 1,800 daily logs, engineering trend-based features such as rolling averages and lag indicators to capture trigger accumulation. We benchmarked our proposed architecture against Logistic Regression and Random Forest classifiers. Experimental results demonstrated that traditional machine learning models significantly outperformed the deep learning approach on this structured dataset. **Logistic Regression achieved the highest accuracy of 88.52%**, followed by Random Forest at 85.79%, while the Bi-LSTM achieved only 76.60% with an ROC-AUC of 0.43. These findings suggest that for structured patient logs with rigorous feature engineering, interpretable linear models offer superior predictive power over complex neural architectures.

I. Introduction

Migraine attacks are more than just headaches; they are complex neurological events that can last for days. Managing migraines often relies on identifying and avoiding "triggers." However, these triggers vary wildly between patients and are often time-delayed (e.g., poor sleep on Monday causing a migraine on Wednesday). This temporal lag makes it difficult for patients to self-diagnose their triggers effectively.

The challenge this project addresses is the automated prediction of migraine attacks using longitudinal patient data. While traditional machine learning can analyze single-day events, it often fails to capture the cumulative effect of triggers over time without explicit feature engineering.

To address this, we proposed using Recurrent Neural Networks (RNNs), specifically **Bidirectional Long Short-Term Memory (Bi-LSTM)** models. LSTMs are designed to handle sequential data and can "remember" patterns from previous days. We aimed to demonstrate whether a deep learning approach could effectively model these temporal dependencies to provide accurate daily risk assessments compared to static baselines.

II. Methodology

A. Dataset Rationale & Generation

For this project, we utilized a simulated patient-reported dataset generated to model physiological correlations known in medical literature. We opted for a simulated approach because publicly available datasets were primarily cross-sectional rather than longitudinal. Since our primary research objective was to explore temporal dependencies (how past days affect future outcomes), we required a continuous timeline.

The dataset consists of approximately **1,800 daily logs** (representing 5 years of data) containing 7 key features:

- **Stress Level:** Rated 1-10.
- **Sleep Duration:** Hours per night.
- **Caffeine Intake:** Milligrams consumed.
- **Water Intake:** Liters consumed.
- **Screen Time:** Daily hours of exposure.
- **Physical Activity:** Minutes of exercise.
- **Weather Sensitivity:** Binary trigger (0 or 1).

B. Feature Engineering & Pre-Processing

Raw daily data is often noisy. To enable both our deep learning and baseline models to detect trends, we applied advanced feature engineering:

1. **Rolling Averages (Trends):** We calculated 3-day rolling means for Stress, Sleep, Caffeine, and Water. This captures the *accumulation* of triggers (e.g., chronic sleep deprivation vs. a single bad night).
2. **Lag Features:** We created 1-day lag features to explicitly model the immediate impact of the previous day's physiological state.
3. **Normalization:** We applied Min-Max Scaling to normalize all continuous features to the range [0, 1] to ensure stability during gradient descent.
4. **Time-Series Splitting:** To respect the temporal nature of the data, we split the dataset chronologically: the first 80% was used for training and the final 20% for testing. Random shuffling was strictly avoided to prevent data leakage.

C. Proposed Architecture: Bidirectional LSTM

We modeled the migraine prediction task as a binary classification problem on time-series data.

- **Input Layer:** Accepts sequences of 7 days of data (Lookback Window) to capture weekly cycles.
- **Bidirectional LSTM Layer:** We employed a layer with 64 units that processes the sequence in both forward and backward directions, allowing the network to learn context from the entire window.
- **Regularization:** A dropout rate of 0.3 and Batch Normalization were applied to prevent overfitting.
- **Output Layer:** A single Dense neuron with a Sigmoid activation function outputs a probability score (0 to 1).
- **Optimization:** We used the Adam optimizer ($lr=0.001$) and Binary Cross-Entropy loss. To handle class imbalance, we computed and applied **Class Weights** during training.

III. Experiments and Results

A. Baselines

To evaluate the effectiveness of our proposed Deep Learning model, we compared it against two standard machine learning baselines:

1. **Logistic Regression:** A simple linear classifier that treats each day as an independent event.
2. **Random Forest Classifier:** An ensemble method capable of capturing non-linear interactions between features.

B. Experimental Results

We evaluated all models on the held-out test set (the final 20% of the timeline). The accuracy scores are summarized below:

Model	Accuracy	Notes
Logistic Regression	**88.52%**	**Best Performer.** Modeled linear trends effectively.
Random Forest	85.79%	Strong performance, captured non-linearities.
Bi-LSTM (Proposed)	76.60%	Failed to generalize beyond majority class.

C. Analysis

1. **Superiority of Linear Models:** The Logistic Regression model achieved the highest accuracy (88.52%). This suggests that once "trend" features (like 3-day rolling averages) were explicitly engineered into the dataset, the relationship between triggers and migraines became linearly separable. The complex LSTM architecture was not required to extract these patterns.

2. **Deep Learning Limitations:** The LSTM achieved 76.60% accuracy, which matched the frequency of "No Migraine" days in the dataset. An ROC-AUC score of 0.43 confirmed that the model suffered from **mode collapse**, defaulting to predicting the majority class despite the use of class weights. This highlights a known limitation of deep learning on small, tabular datasets where signal-to-noise ratios are low.
3. **Feature Importance:** Analysis using the Random Forest model identified **Screen Time**, **Sleep Duration**, and **Water Intake** as the dominant predictors. This aligns with medical consensus that dehydration and digital eye strain are primary physiological triggers.

IV. Conclusion

In this project, we successfully developed a migraine prediction pipeline. While we hypothesized that a Bidirectional LSTM would outperform static models by learning latent temporal dependencies, our results proved otherwise. The **Logistic Regression** model outperformed the deep learning approach by nearly 12%.

This outcome validates a critical lesson in healthcare data science: **Feature Engineering is often more valuable than Model Complexity.** By explicitly calculating trends (rolling averages) in the pre-processing stage, we enabled a simple, interpretable model to achieve high performance, making it a more viable candidate for real-world deployment in patient monitoring apps.

References

Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. Neural Computation 9, 8 (1997), 1735-1780.

Dodick, D. W. 2018. Migraine. The Lancet 391, 10127 (2018), 1315-1330.