Anticipating menstrual migraine using deep learning

Sai Sharanya Y
B. Tech. (Hons.) – CSE - DS
SCSE – Jain (deemed-to-be) University
Bengaluru, Karnataka, India 562112
sharanyachowdary99@gmail.com

Pranav Polvarapu
B. Tech. (Hons.) – CSE - DS
SCSE – Jain (deemed-to-be) University
Bengaluru, Karnataka, India 562112
19btrcr008@jainuniversity.ac.in

Sai Keerthi Chelluri
B. Tech. (Hons.) – CSE - DS
SCSE – Jain (deemed-to-be) University
Bengaluru, Karnataka, India 562112
19btrcr036@jainuniversity.ac.in

Sushil Bokade
B. Tech. (Hons.) – CSE - DS
SCSE – Jain (deemed-to-be) University
Bengaluru, Karnataka, India 562112
19btrcr017@jainuniversity.ac.in

Dr. Mohammad Zabeeulla A N
Assistant Professor – CSE - DS
SCSE – Jain (deemed-to-be) University
Bengaluru, Karnataka, India 562112
zabee225@gmail.com

***Abstract*—Menstrual Migraines are headaches that occur without any typical aura or sensory disturbances. They can be labelled as 'Migraine without Aura'. A woman could undergo menstrual migraine just before or during her period starts. This is majorly due to the drop in estrogen levels in the body. Menstrual migraines can be intense, and they must be identified accurately and treated separately from regular ones. Anticipating Menstrual Migraines using Deep Learning methodology can help women in taking precautions beforehand and aid in improving their health. This paper provides an efficient approach by checking various feature correlations and implementing ANN model with tuned hyperparameters providing stronger accuracy.**

Keywords—Deep Learning, Adam Optimizer, Sequential Neural Network, Categorical Cross-Entropy, Estimator, keras Classifier, Correlation, Menstrual Migraine, Gradio User Interface

I. INTRODUCTION

Women often experience menstrual migraines, which are a specific type of migraine headache that happens during this time. Menstrual migraines are considered to impact 60% of women who get migraines. Hormonal changes, notably the decline in oestrogen levels that take place before and during menstruation, are the main culprits for menstrual migraines. Menstrual migraines share a number of similar symptoms as ordinary migraines, such as a throbbing or pulsating headache, nausea, sensitivity to light and sound, and occasionally aura or visual problems. Menstrual migraines, however, can be more challenging to manage because they frequently stay longer and are more severe than ordinary headaches. In order to create a treatment strategy that addresses both the migraine symptoms and the hormonal changes that cause them, women who get menstrual migraines may find it helpful to keep track of their menstrual cycles.

Without any sensory abnormalities or aura, menstrual migraine without aura is a specific type of migraine headache that affects women around the time of their period. About 75% of migraine cases are migraines without aura, its most frequent kind of migraine. Menstrual migraine without aura is defined by a headache that is habitually accompanied by other signs and symptoms such nausea, vomiting, light and sound sensitivity, and lethargy. When compared to typical migraines, these sensations are frequently more intense and persistent.

Although the precise origin of menstrual migraine without aura is unknown, it is thought to be a result of hormonal changes, particularly the decline in oestrogen levels that takes place before and during menstruation. Changes in oestrogen levels can cause migraine attacks in people who are vulnerable to them because oestrogen is known to impact how the brain interprets pain signals. Stress, alterations in sleep habits, and food triggers like caffeine and alcohol are among other variables that may trigger menstrual migraine without aura in susceptible individuals.

Menstrual migraine without aura may be treated with a blend of prescription drugs, dietary modifications, and hormonal therapy. Non-steroidal anti-inflammatory drugs (NSAIDs), triptans, and preventative therapies including beta-blockers and antidepressants are all commonly prescribed for menstrual migraines.

Migraines can be classified into a variety of groups, including menstrual, non-menstrual, and sub-groups for aura and without aura. Many models, including Custom Sequential ANN, RNN, and RBFN, can be used to accomplish this.

To create models that can segregate data into different categories, Keras Sequential can be utilized. A high-level neural network API called Keras Sequential enables programmers to quickly create, train, and test deep learning models. It is constructed on top of TensorFlow, a well-liked open-source machine learning model development and training framework. Users of Keras Sequential can define a model as a series of layers, where each layer applies a certain functionality to the input information. Users of the KerasClassifier estimator can additionally create unique loss functions, optimizers, and metrics, which might be helpful for optimising the accuracy of the model. The KerasClassifier estimator is a flexible tool for classification issues in a variety of fields because it can be employed with multiple forms of input data, including image data, text data, and numerical data.

II. LITERATURE SURVEY

Ahmet, Alkan, and Batuhan Akben Selahaddin [1] covers many strategies for detecting migraines that have been put forth in earlier studies. However, the majority of these earlier studies were founded on frequency domain analysis of EEG data. The frequency domain transformation of the EEG data required the use of sophisticated mathematical algorithms. The FIR filter is used to filter the beta band of each flash stimulated and non-stimulated EEG data for both migraine patients and healthy people. Comparing the figures, it is clear that the beta band of EEG signals has smaller amplitude. The time domain-based preprocessing approach known as the histogram is then used for all of the filtered EEG data. The outcome demonstrates how flash stimulation affected patients with and without migraines' beta band EEG data. In this analysis, characteristics were extracted from flash stimulated and non-stimulated EEG data using a time domain-based histogram approach, which was then clustered to represent healthy people and migraine patients. It was shown that, for migraine sufferers, the amplitude frequency histograms (Cluster 2) grouped very strongly with silhouette values greater than or equal to 0.63. Similar results are also attained with healthy people. The accurate clustering accuracy for all the data may be computed to be 86.6%.

Akben, Deniz Tuncel bdulhamit Subasi, and Selahaddin Batuhan [2], here, the frequency of flash stimulation that is most beneficial is identified, as well as how quickly a migraine can be diagnosed. This study's main objective was to determine the best flash stimulation frequency for migraine identification. As a result, it was discovered that 4 Hz was the most efficient flash stimulation frequency for identifying migraines. Finding the shortest window of time to detect a migraine is the second goal. It was discovered that 8 seconds was the absolute minimum needed to detect a migraine. The amplitude increase at the beta band of the EEG signal in the T5-T3 channel is applied to select the optimal flash stimulation frequency and shortest time duration in EEG recordings for migraine identification.

Akben, Selahaddin Batuhan, Abdülhamit Subaşı, and Mahmut Kemal Kıymık [3] reported that 23% of people suffer from the severe and excruciating brain ailment known as migraine. There is no documented or widely accepted automatic approach for diagnosing migraines using biomedical equipment. EEG signal changes caused by flash stimulation are typically utilized as a tool to gather information on migraines. Utilizing EEG signals from migraine patients who have been induced with a flash light, this study aims to autonomously diagnose migraines by completing a performance assessment of classification techniques. Then, EEG signals from both migraine patients and healthy individuals are converted into the frequency domain using the (AR) Burg method. These frequency spectrums are further classified using the artificial neural network and support vector machine classification techniques. Which classification approach works the best for migraine diagnosis can be determined using these categorization findings.Charles, Andrew [4], the migraine aura is a spectacular neurologic occurrence with intricate neuronal and vascular mechanisms that may have significant effects on diagnosis and treatment. The biology of migraine and its ideal treatment can be better understood with a more in-depth study of its clinical characteristics, comorbidities, patterns of propagation in the human brain, and unique responses to therapy.

Chong, Catherine D., Nathan Gaw, Yinlin Fu, Jing Li, Teresa Wu, and Todd J. Schwedt [5], The classification of migraine using rs-fMRI sheds light on the altered pain circuits in migraine and may help in the creation of a novel, noninvasive migraine biomarker. Longer disease duration may cause remodeling of brain circuitry, as shown by the fact that migraineurs with longer disease duration were categorized more accurately than migraineurs with shorter disease duration.

Ang, Kai Keng, Cuntai Guan, Kerry Lee, Jie Qi Lee, Shoko Nioka, and Britton Chance [6], Near-infrared spectroscopy (NIRS) uses light in the near-infrared range to enable non-invasive monitoring of cerebral haemoglobin oxygenation in human participants over the entire skull. This study presents a unique BCI for identifying changes brought on by increases in the scale of inputs used in a mental arithmetic problem using data from single-trial NIRS brain signals. 20 healthy individuals performed mental arithmetic tasks at three levels while their haemoglobin responses were tracked. Following that, precision in discriminating between one difficulty level and another is provided using 55-fold cross-validations on the collected data. With an overall average accuracy of 71.2%, the results suggested that the suggested NIRS-based BCI may be useful in determining the degree of difficulty of problems encountered by mental arithmetic problem solvers.

Garcia-Chimeno, Yolanda, Begonya Garcia-Zapirain, Marian Gomez-Beldarrain, Begonya Fernandez-Ruanova, and Juan Carlos Garcia-Monco [7], The proposed feature selection committee technique enhanced the efficiency of classifiers for migraine diagnosis when compared to individual feature selection approaches, leading to a robust system that achieved above 90% success across all models. The results imply that the offered methodologies may prove useful in assisting professionals in the categorization of migraines in patients undergoing magnetic resonance imaging. The proposed committee-based feature selection technique demonstrated the greatest improvements in recognition rate when the migraine category was taken into account. The accuracy of categorization into 3 groups improved while applying the Naive Bayes classifier, the Support Vector Machine classifier, and boosting, going from 67 to 93%, 90 to 95%, and 93 to 94%, respectively. Analgesics and their effects, as well as the characteristics of pain. It was discovered that characteristics related to pain, analgesics, and the left uncinate brain were deemed most useful for classification (related to the pain and emotions).

Jackowski, Konrad, Dariusz Jankowski, Dragan Simić, and Svetlana Simić [8], A reliable migraine diagnosis is a challenging decision. The paper also presents an ensemble classifier technique for headache assessment. Here, it is anticipated that the system will quickly diagnose a problem using simply the information gathered during the questionnaire. The majority of traditional classification algorithms could not be applied under such an assumption since they could not achieve a respectable degree of accuracy. It is chosen to use an ensemble solution as a result. As a result, we used a two-stage technique. First, a sizable pool of basic classifiers was created. By choosing algorithms with various kinds, structures, and learning methods, its diversity was guaranteed. Second, we chose the ensemble's ideal size and its members by employing exhaustive search techniques. Studies done on information acquired at the University of Novi Sad showed that the proposed system far outperformed all conventional approaches. Also covered are analyses of diversity and precision correlations for systems that have undergone testing.

Sanchez-Sanchez, Paola A., José Rafael García-González, and Juan Manuel Rúa Ascar [9] This article describes the creation of an approach for classifying migraines using models of artificial neural networks. The findings support those observed when multiple models provided in the literature are evaluated, showing that artificial neural networks can achieve higher precision and accuracy than other classification models frequently employed in machine learning. In the initial studies, 24 migraine diagnosis-related variables were used, and the artificial neural network model's precision level was 97%. In a subsequent testing step, the number of variables was reduced to 18, producing a precision of 98%. This shows that in addition to this, the artificial neural network model is effective at categorizing the different types of migraine accurately, but then also that it may be improved by accounting for a smaller number of factors that have a serious influence on the classification

Silberstein, Stephen D., and Susan L. Hutchinson [10] this paper states that women who are sensitive, a reduction in oestrogen is linked to an intensification of migraines. Such women are frequently characterized with MRM or PMM, and variations in circulating hormone levels may also cause them to have additional non-headache symptoms (e.g., anxiety, depression, dysphoria, among others). In order to correctly identify the type of headache and recommend the best solutions, alternative diagnosis is essential. In order to properly inform treatment choices, frequency, severity, non-pain symptoms, and disability must all be carefully evaluated in relation to hormonal effects. Many MRM patients can benefit from monotherapy, which effectively treats all of their attacks regardless of whether they are related to menstruation or not. However, not all women may benefit from such a strategy; others may require short- or long-term prophylaxis in addition to emergency care and life-saving drugs. The use of polytherapy to treat concurrent disorders or symptoms is possible (e.g., PMDD, depression, anxiety, among others). Although it is now a widely recognized clinical phenomena that fluctuations in circulating oestrogen levels can cause migraines, moderating plasma oestrogen levels as a form of treatment is currently being investigated. Long-term, double-blind, randomized controlled trials are required to better determine the best approach and most efficient medications for treating migraines brought on by changes in oestrogen levels.

III. METHODOLOGY

The proposed system is implemented by primarily selecting the significant features which are highly correlated, and they are used as input parameters for the customized Keras sequential ANN model which is stacked with a KerasClassifier estimator, resulting in accurate prediction of migraine classification model. The whole system is implemented on a Gradio User Interface.

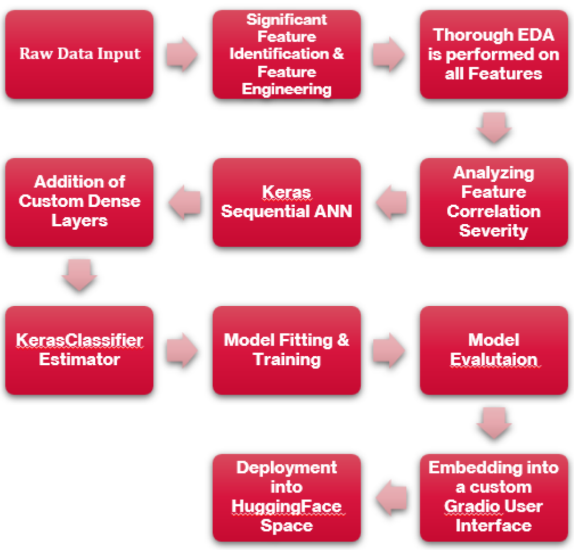

Fig. 1. Proposed System Framework.

A. *Keras Sequential Neural Network*

A prevalent neural network design for classification problems is the Keras Sequential model. The Sequential model is intended to be a stack of layers, each of which operates on the input data in a particular way. The input data passes successively through the layers, with the output of one layer being the input for the following layer.

Quantitatively, the conceptual model of the sequential model for classification is as follows:

$y = f(Wx + b)$

here,

W - is the weight matrix

b - is the bias vector

f - is the activation function

x - is the input data

y - is the output prediction

The Sequential model is composed of a number of layers, each of which has a type and a set of variables that describe it. The dense layer and the activation layer are the two most typical types of layers utilized in the Sequential model for classification. The dense layer, a fully - connected layer, transforms the input data linearly using a bias vector and a weight matrix. The activation function is then applied to the dense layer's output to provide nonlinearity to the model.

ReLu and Softmax are the activation functions used in this sequential model's Dense layers.

ReLU (Rectified Linear Unit) is a simple and effective activation function that has become a standard choice in deep learning. It is easy to compute, has a simple derivative, and is effective at introducing non-linearity into the model. It is defined mathematically as:

$f(x) = max(0, x)$

where x is the input to the function and f(x) is the output.

The derivative of ReLU can be expressed as:

$f'(x) = 1 \text{ if } x > 0, \; f'(x) = 0 \text{ if } x <= 0$

The mathematical derivation of ReLU is straightforward. It is simply a piecewise function that sets negative inputs to zero. The function is continuous everywhere except at the origin, where it has a corner.

The piecewise definition of ReLU can be written as:

$f(x) = \{x \text{ if } x >= 0; \; 0 \text{ if } x < 0\}$

The softmax function maps the input values to a probability distribution over classes, with each value in the output vector representing the probability of the input belonging to a particular class. The class with the highest probability is considered the predicted class for the input. The softmax function is an important activation function in classification tasks, as it enables the neural network to output probability estimates for each possible class, which can be used to make informed decisions.

The softmax function is mathematically defined as:

$f(x_i) = e^\wedge(x_i) / sum(e^\wedge(x_i))$

where *xi* is the input value for the i-th neuron in the output layer, and j ranges over all neurons in the output layer. The denominator in the equation is the sum of the exponential values of all neurons in the output layer. This normalization term ensures that the output values of the softmax function sum to 1, which is a requirement for probability distributions.

Categorical cross-entropy is a widely used loss function in neural networks for multi-class classification tasks. It is commonly used in conjunction with the softmax activation function in the output layer.

The categorical cross-entropy loss function is defined as:

$$L(y, f(x)) = -sum(y_i * log(f(x)\, f(x)_i))$$

where *y* is the true probability distribution over the classes, *f(x)* is the predicted probability distribution over the classes, and *i* ranges over all possible classes. The disparity between both the true distribution and the expected distribution is measured by the loss function. The loss function is minimized during training by adjusting the weights and biases of the neural network through backpropagation. The backpropagation algorithm calculates the gradients of the loss function with respect to the network parameters and uses them to update the weights and biases to minimize the loss function.

The categorical cross-entropy loss function is widely used because it provides a smooth and continuous measure of the difference between the predicted and true probability distributions. It also has the desirable property of being additive, which means that the total loss for a batch of inputs can be computed as the sum of the losses for each individual input.

The Adam optimizer combines the advantages of two other optimization algorithms, AdaGrad and RMSProp. Like AdaGrad, it keeps track of the sum of squared gradients of the parameters, which helps to scale the learning rate for each individual parameter.

The update rule for the Adam optimizer is given by the following equations:

$$m_t = beta_1 * m_{\{t-1\}} + (1 - beta_1) * g_t$$

$$v_t = beta_2 * v_{\{t-1\}} + (1 - beta_2) * g_{t^\wedge 2}$$

$$m_t = m_t / (1 - beta_1 {}^\wedge t)$$

$$v_t = v_t / (1 - beta_2 {}^\wedge t)$$

$$w_t = w_{\{t-1\}} - alpha * m^t / (sqrt(v^t) + epsilon)$$

where,

- g_t is the gradient of the loss function at time step *t*
- m_t and v_t are the first and second moments of the gradient, respectively
- $beta_1$ and $beta_2$ are the exponential decay rates for the moments
- m^t and v^t are bias-corrected estimates of the moments
- *alpha* is the learning rate
- *epsilon* is a small value added to the denominator to prevent division by zero
- w_t is the updated parameter value at time step **t**

The Adam optimizer uses the moving averages of the first and second moments of the gradient to adjust the learning rate for each parameter. The first moment, m^t, estimates the mean of the gradients, while the second moment, v_t, estimates the variance of the gradients. The bias correction terms, m^t and v^t, correct for the fact that the estimates are biased towards zero in the early stages of training.

The KerasClassifier estimator is a wrapper for the Keras Sequential model that allows it to be used with scikit-learn. It takes the base model as an input and allows for the specification of training parameters such as batch size, number of epochs, and verbose level.

B. Implementation

1) *Data Collection:*

Data was collected from the *Code Ocean capsule*, in which the symptoms and the severity of each patient's symptoms were securely gathered and converted to a csv file for processing and quick access to the data.

2) *Feature Identification and Feature Engineering:*

Primarily, 24 significant features have been identified and then the data has been cleaned by handling missing values using data imputation methods, and then the categorical data has been encoded based on the number of classes, wherein we used label encoding and replace methods.

3) *Exploratory Data Analysis*:

The correlation between features and the target variable has been analyzed using various Exploratory Data Analysis methodologies. The various charts/plots used in the EDA process are:

- Percentage Plots
- Histogram
- Violin Plot
- Box Plot
- Distribution plot with KDE and boxplots
- Count Plots

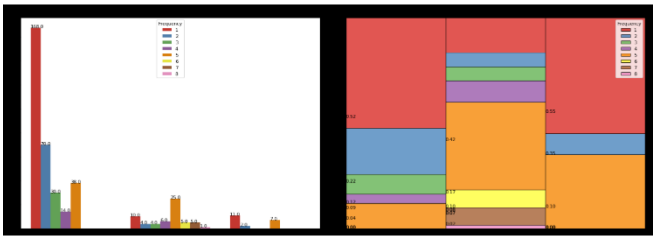

Fig. 2. Percent plot of frequency vs migraine type.

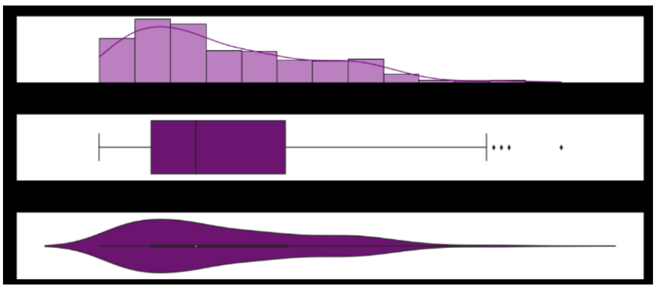

Fig. 3. Histogram, Boxplot and Violin plot of intensity feature.

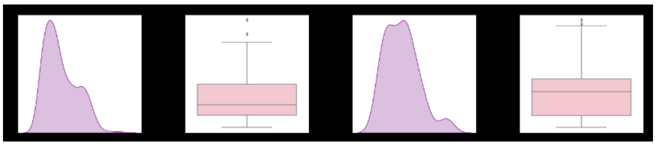

Fig. 4. Distribution plot with KDE and Boxplots for location feature.

Fig. 5. Count plot for migraine types.

4) Keras Sequential ANN:

This is a baseline Keras Sequential model for multi-class classification. The model consists of two Dense layers with 14 units each, both using the Rectified Linear Unit (ReLU) activation function, in addition to the Baseline Model. The output layer has 3 units and uses the Softmax activation function, which is commonly used in multi-class classification tasks. The model is compiled with the categorical cross-entropy loss function, which is a measure of the dissimilarity between the predicted probabilities and the true probabilities, and the Adam optimizer, which is a popular stochastic gradient descent algorithm that uses adaptive learning rates and momentum to update the model parameters.

5) Keras Estimator:

An instance of the KerasClassifier class is a wrapper for the Keras Sequential model that allows it to be used with scikit-learn. The build_fn parameter is set to the baseline_model function that defines the architecture of the model. This function is called by the KerasClassifier to build the model. The classifier is set to 100 epochs with batch size of 10 per epoch and verbose set to 'no logging' i.e 0.

6) Model Training and Evaluation:

Finally, the data is fit into the model and it is trained and tested with the specified hyperparameters, thus providing accurate and improvised results as output predictions. The output class labels are into 3 categories: Menstrual Migraine, Non-Menstrual Migraine, and Others (irrelevant to menstrual migraine).

7) Gradio User Interface:

The Gradio interface is designed to predict the type of migraine based on 24 different input variables. These variables include age, duration, frequency, location, character, intensity, and various other symptoms related to migraine. The interface provides a text box for each of the 24 input variables, where the user can enter the corresponding value. The input values are then passed through a pre-trained KerasClassifier model that has been trained on a dataset of migraine data. Once the input values are passed through the model, the predicted output is returned as a text output component in the Gradio interface. The predicted output is one of three types of migraine: Non-Menstrual Migraine, Menstrual Migraine, or Others. To make the Gradio interface more user-friendly, the interface provides clear instructions on how to provide the necessary inputs. Additionally, the interface includes visual elements, such as images and textboxes, to provide additional information about the project and its authors. The interface has three tabs, namely "Home", "Guidelines", and "Menstrual Migraine Model". The "Home" tab displays general information about the migraine predictor. The "Guidelines" tab provides instructions for providing input values to the interface. The "Menstrual Migraine Model" tab contains the interface for entering the input values and getting the predicted migraine type as output.

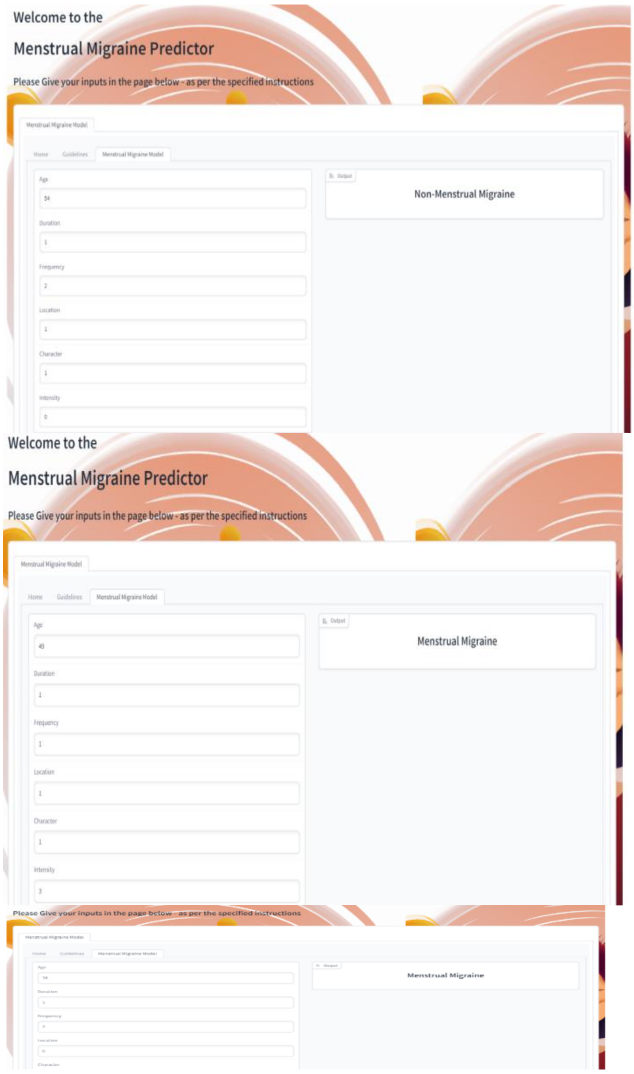

Fig 6 – model prediction output deployed on gradio UI predicting types of migraines for various user inputs.

IV. Results and Conclusion

Overall, the code defines a baseline Keras Sequential model with two hidden layers of 14 units each, ReLU activation functions, and a softmax output layer. Categorical cross-entropy loss function and the Adam optimizer are utilized to train the model. An accuracy of 88.72% was achieved by the model.

Based on the feature identification and feature engineering, it is anticipated that age plays a major role in menstrual migraine prediction and evaluating the symptom severity. The keras sequential model was optimized using custom dense layers along with keras estimator providing a recall value of 0.94, a precision of 0.89, and an accuracy of 88.72%.

V. FUTURE SCOPE

The developed interface can be integrated with period tracker applications, providing complete information to the users regarding their migraine health condition during menstruation and suggesting appropriate actions based on the predictions. The interface can also be connected to a database, such that the system can store any new inputs and leverage them to continuously improve our model performance. The system can be scaled up through utilizing cloud architecture.

VI. REFERENCES

[1] Ahmet, Alkan, and Batuhan Akben Selahaddin. "Use of k-means clustering in migraine detection by using EEG records under flash stimulation." International Journal of Physical Sciences 6, no. 4 (2011): 641-650.

[2] Akben, Selahaddin Batuhan, Abdulhamit Subasi, and Deniz Tuncel. "Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network." Journal of medical systems 36, no. 2 (2012): 925-931.

[3] Akben, Selahaddin Batuhan, Abdülhamit Subaşı, and Mahmut Kemal Kıymık. "Comparison of artificial neural network and support vector machine classification methods in diagnosis of migraine by using EEG." In 2010 IEEE 18th Signal Processing and Communications Applications Conference, pp. 637-640. IEEE, 2010.

[4] Charles, Andrew. "The migraine aura." Continuum: Lifelong Learning in Neurology 24, no. 4 (2018): 1009-1022.

[5] Chong, Catherine D., Nathan Gaw, Yinlin Fu, Jing Li, Teresa Wu, and Todd J. Schwedt. "Migraine classification using magnetic resonance imaging resting-state functional connectivity data." Cephalalgia 37, no. 9 (2017): 828-844.

[6] Ang, Kai Keng, Cuntai Guan, Kerry Lee, Jie Qi Lee, Shoko Nioka, and Britton Chance. "A brain-computer interface for mental arithmetic task from single-trial near-infrared spectroscopy brain signals." In 2010 20th International Conference on Pattern Recognition, pp. 3764-3767. IEEE, 2010.

[7] Garcia-Chimeno, Yolanda, Begonya Garcia-Zapirain, Marian Gomez-Beldarrain, Begonya Fernandez-Ruanova, and Juan Carlos Garcia-Monco. "Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data." BMC medical informatics and decision making 17, no. 1 (2017): 1-10.

[8] Jackowski, Konrad, Dariusz Jankowski, Dragan Simić, and Svetlana Simić. "Migraine diagnosis support system based on classifier ensemble." In International conference on ICT innovations, pp. 329-339. Springer, Cham, 2014.

[9] Sanchez-Sanchez, Paola A., José Rafael García-González, and Juan Manuel Rúa Ascar. "Automatic migraine classification using artificial neural networks." F1000Research 9 (2020).

[10] Silberstein, Stephen D., and Susan L. Hutchinson. "Diagnosis and treatment of the menstrual migraine patient." Headache: The Journal of Head and Face Pain 48 (2008): S115-S12