2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT) | 979-8-3503-1341-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/EASCT59475.2023.10393599

Migraine Categorization using the Scatter Search and Random Forest Classifier

Mary Litta David K, Vengala Venkata Sai Sharmili
Department of Computer Science & Engineering
School of Engineering
Dayananda Sagar University
Bengaluru, India
marylitta@gmail.com, sharmili.vengala@gmail.com

Tina Babu, Rekha R Nair
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
tinababup@gmail.com, rekhasanju.sanju@gmail.com

***Abstract*—Migraine attacks manifest through recurring episodes of intense headaches accompanied by autonomic nervous system dysfunction indicators, leading to various symptoms. Automatic classification of various classes of migraine is very much essential for the secondary opinion. This study categorizes a migraine dataset into seven distinct classes using features obtained through scatter search. To address the class balance issue, Synthetic Minority Oversampling Technique (SMOTE) was applied. Further, the optimal features were selected using Scatter Search. Subsequently, classification was executed using the Random Forest (RF) classifier. The proposed model was tested on a migraine dataset containing 400 instances and 24 features, resulting in a feature reduction of around 50%. Remarkably, the achieved accuracy of 98.26% surpassed that of the raw dataset. As a result, the proposed model outperforms existing methodologies, potentially offering an additional perspective for medical professionals.**

***Index Terms*—migraine, search, feature reduction, classification**

I. Introduction

This document shows migraine and its different types and its diagnosis is based on typical symptoms. Migraine is a neurological disorder characterized by recurring headaches that can be severe. A typical migraine data set may include different data points and features that capture different aspects of the condition. Headache is a prevalent symptom associated with nervous system disorders, characterized by pain occurring in the head. It can affect individuals of all ages, races, and economic backgrounds, with a higher prevalence in women. Nearly half of the global population experiences headaches, making it the third most commonly reported symptom worldwide. By studying these datasets, researchers aim to improve diagnostic accuracy, identify personalized treatment approaches, develop preventive strategies, and enhance the overall management of migraines.

Machine Learning has brought many changes in the field of research in areas of image processing [1], [2], natural language processing, healthcare [3] etc. Over the past decade, researchers have explored various machine learning classification algorithms to categorize different types of headaches. Vandewiele et al. conducted a study using multiple algorithms and recommended the Decision Tree model for classification [4]. Similarly, Krawczyk et al. compared algorithms and found that the Random Forest achieved the highest accuracy in headache classification after implementing feature selection techniques [5]. Various algorithms and synthetic datasets based on The International Classification of Headache Disorders (ICHD-2) were used by Aljaaf et al. and with their results also indicating that the Decision Tree method had the highest accuracy [6].

The main objective of this reserach is to develop a framework that could categorize the telecom churn dataset. The contribution of this are on many levels as follows:

- SMOTE was applied to the migraine dataset to address the issue of class imbalance and thereby improving the accuracy.
- The optimal features are extracted with the Cfs Subset evaluation and Scatter search. Thus the most prominent features are selected for the categorization.
- Many machine learning techniques were analysed for the comparitative analysis and RF classifier was giving the promising performance.

The rest of the manuscript is organized with the literature survey in Section II. The proposed framework is discussed in Section III, followed by experimental results and discussion in IV. The manuscript is concluded in Section V.

II. Literature Survey

Artificial intelligence (AI) has found considerable success in various fields such as law, regulation, plant disease, and medical issues [1]. Notably, it has demonstrated its effectiveness in tasks like diagnosing hypertension, discovering new drugs, and identifying nephropathy in newborns [7]. Moreover, AI has been applied to assess the risk of cancer-associated thrombosis and to calculate the risk of colon cancer progression [3], [8]. The ongoing development of AI applications includes areas like Alzheimer's and Parkinson's disease diagnosis, as well as brain tumor classification [9].

Differentiating between migraines with aura, migraines without aura, and other types of migraines and headaches is determined by applying specific criteria outlined by the International Headache Society [10]. Diagnosis of migraines involves evaluating the patient's medical history and symptoms, along with conducting a physical and neurological

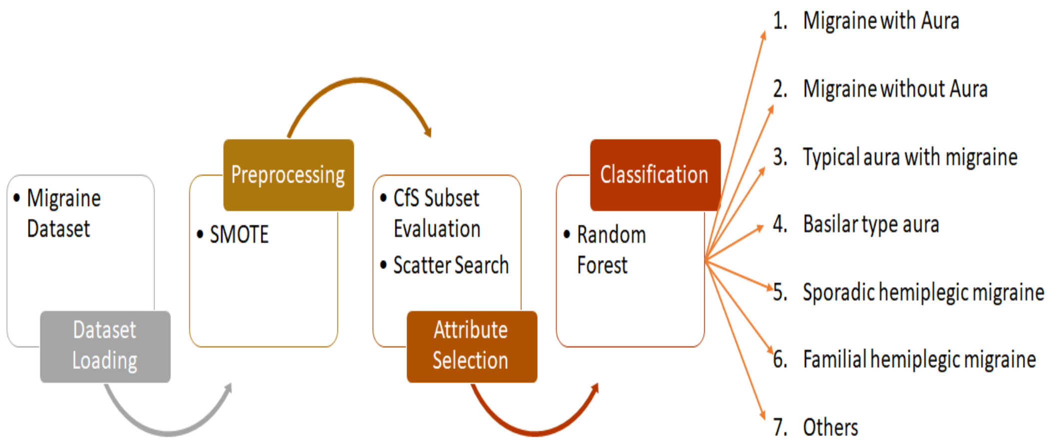

Fig. 1. Methodology of the Proposed Model

assessment. In certain cases, specialized tests like magnetic resonance imaging, tomography, electroencephalogram, and lumbar puncture might also be conducted [11].

Employing artificial neural networks [12] reduces diagnostic time and enhances accessibility by relying on patient-reported symptoms instead of medical devices. The study aims to achieve higher accuracy than the previous 70% benchmark, leading to better healthcare decisions and improved patient outcomes [13].

The study focuses on using artificial intelligence and supervised learning techniques to facilitate the prompt identification of several migraine varieties [14]. It seeks to create a classification framework that correctly ascertains the precise migraine type from symptoms and past health records. The study's significance lies in introducing an indirect method for classification, bypassing the need for brain wave measurements or sensors. There are various steps in the methodical categorization process, promising potential advancements in migraine diagnosis and patient care [15]. By employing this approach, the researchers aim to develop an effective and accurate system for classifying different types of migraines, facilitating early identification and appropriate medical intervention for patients.

Thus a proper classification system for migraine classification is essential with the abstract features. The performance also needs to be enhanced with the reduced feature set.

III. Proposed Methodology

The proposed methodology is illustrated in Figure 1. It consists of various phases such as preprocessing, feature selection using searching techniques, and finally classification. The phases are as follows

A. Dataset Experimented

The migraine dataset contains various features and is categorized into seven different migraine classes [12]. The study focused on classifying different migraine types, including Migraine with aura, Typical aura with migraine, Typical aura without migraine, Basilar-type aura, Sporadic hemiplegic migraine, Feminine hemiplegic migraine, and others. Notably, significant differences were observed among the classes concerning the distribution of specific symptoms: Headache intensity: Class 3 exhibited the highest frequency of severe headache intensity. Unilateral pain: Unilateral pain was the most commonly observed symptom, and class 1 had the lowest frequency of worsening with ordinary physical activity. Worsening with routine physical activity: Class 3 had the highest frequency of this symptom, while it was the least common in class 1. Nausea: Class 3 showed the highest frequency of nausea. Vomiting: The frequencies of vomiting were significantly different between class 1 and class 3. Photophobia and phonophobia: These symptoms were most common in class 2. Quality of pulsation: No significant differences were observed among the three classes for this symptom. Further analyses through multiple comparisons revealed significant differences in the frequencies of specific symptoms between different pairs of classes:

Significant differences in symptom frequency between class 1 and class 2 indicate distinct characteristics, crucial for accurate classification and targeted treatment. Distinct symptom frequency variations between class 1 and class 3 highlight important differences for accurate classification and targeted treatment. Headache intensity, photophobia, and phonophobia exhibit significant differences between class 2 and class 3, crucial for accurate classification and treatment strategies. These findings provide valuable insights into the distinctive symptom profiles among the different migraine classes, which can aid in the accurate classification and diagnosis of migraine types.

B. Preprocessing using SMOTE

The study uses SMOTE, a resampling technique, to preprocess the data and address class imbalance. SMOTE generates synthetic samples to balance the class distribution, improving the reliability and representativeness of the training dataset for the classification model. Its purpose is to address the overfitting issue that arises with random resampling. When dealing with imbalanced data, where one class is significantly underrepresented compared to others, SMOTE helps in elevating the minority group to create a more balanced distribution. By applying SMOTE, the class values become balanced, ensuring that the classifier model performs better on the data. If unbalanced data is not preprocessed, the classifier may have reduced performance. The majority class predictions tend to dominate, while the minority classes may be considered noise and ignored. As a consequence, the model becomes biased towards the majority class, leading to suboptimal predictions and accuracy on the minority classes. SMOTE helps in mitigating this issue and improves the overall performance of the classifier, especially when dealing with imbalanced datasets.

The following definition applies to the distance $delta$ that separates two comparable feature values:

$$\delta(F_1, F_2) = \sum_{i=1}^{n} |\frac{S_{1i}}{S_i} - \frac{S_{2i}}{S_2}|^k \tag{1}$$

F_1 and F_2 are the two comparable value of features in the equation above. The overall number of instances of the feature value F_1 is S_1, and the amount of instances of the feature value F_1 for the class i is S_{1i}. S_2i and S_2 can be treated using a similar convention. Constant k is typically configured to 1. The matrix containing the value disparities associated with every nominal characteristic in the provided collection of feature vectors is determined using the aforementioned formula [16].

C. Feature Selection

Among the many features or attributes present in the dataset, the most significant ones need to be selected. The following techniques were used for this.

1) Cfs Subset Evaluation: The CfsSubsetEval algorithm is designed to identify a subset of features that complement each other well. The Correlation-based Feature Selection algorithm evaluates the significance of a subset of attributes by considering the predictive ability of each feature and the level of redundancy between them. It aims to identify subsets of features that exhibit strong correlation with the class while maintaining low mutual correlation among the features. This approach helps in selecting the most informative and relevant features for the classification task, leading to improved accuracy and efficiency in the model's performance. The CFS filter approach was employed with four different search methods: PSO Search, Evolutionary search, scatter search, and Tabu search. The evaluation of the selected feature subsets was performed using five different classifiers. The results of the study demonstrated that the feature set chosen by the CFS filter significantly improved the classification performance of all five classifiers compared to their performance with the original feature set. The selected subset of features is more relevant and informative for the classification task. Using various subset evaluation algorithms, such as CfsSubsetEval, can aid in identifying the most important attributes for CFS classification. The combination of the CFS filter approach and different search methods contributes to enhancing the accuracy and effectiveness of the classification models in this study [17].

2) Scatter search (SS): The Scatter Search algorithm is a metaheuristic optimization technique that aims to find solutions to complex optimization problems by exploring and combining a diverse set of solutions in order to improve upon initial solutions [18]. It employs a combination of solution generation, solution evaluation, and solution combination steps to iteratively refine the solutions towards an optimal or near-optimal solution. The algorithm is inspired by evolutionary processes and employs a strategy of diversification and intensification to explore the solution space effectively and converge to better solutions. Here is a step-by-step outline of the Scatter Search algorithm:

1) Initialization:
   - Generate a diverse set of initial solutions (population).
   - Evaluate the solutions using the objective function.
2) Solution Combination:
   - Select pairs of solutions from the population.
   - Create new solutions by combining elements of the selected pairs.
   - These combined solutions can be created using various methods like arithmetic mean, weighted average, crossover, etc.
3) Solution Evaluation:
   - Evaluate the newly generated solutions using the objective function.
   - Rank the combined solutions based on their fitness (objective function values).
4) Reference Set Update:
   - Maintain a reference set of non-dominated (Pareto-optimal) solutions.
   - Update the reference set with the new solutions if they are non-dominated and not already in the set.
5) Solution Subset Selection:
   - Select a subset of solutions from the reference set.
   - The subset can be selected based on various criteria like diversity, density, and coverage.
6) Intensification:
   - Focus on improving solutions in the selected subset.
   - Apply local search or other optimization methods to refine the selected solutions.
7) Diversification:
   - Generate diverse solutions by introducing randomness or perturbation to the selected subset.
   - This helps to explore different regions of the solution space.
8) Solution Replacement:
   - Replace a portion of the current population with the refined and diversified solutions.
   - Maintain the diversity and quality of the population.
9) Termination Criteria:

TABLE I
PERFORMANCE MEASURES WITH THE ORIGINAL DATASET

Classifier	Accuracy(%)	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Naïve Bayes	94.20	0.942	0.009	0.946	0.942	0.943	0.934	0.997	0.985
SVM	95.65	0.957	0.007	0.956	0.957	0.956	0.949	0.99	0.933
Decision Tree	80.83	0.8	0.33	0.872	0.852	0.857	0.836	0.98	0.913
Random Forest	97.68	0.977	0.004	0.978	0.977	0.977	0.973	1	0.999
MLP	98.00	98.26	0.983	0.003	0.984	0.983	0.983	0.98	0.995

TABLE II
PERFORMANCE MEASURES OF THE PROPOSED MODEL

Classifier	Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Naïve Bayes	89.85%	0.899	0.018	0.907	0.899	0.891	0.882	0.985	0.94
SVM	90.14%	0.901	0.017	0.904	0.901	0.897	0.885	0.975	0.858
Decision Tree	80.29%	0.803	0.032	0.806	0.803	0.802	0.771	0.966	0.848
MLP	92.17%	0.922	0.013	0.922	0.922	0.921	0.909	0.987	0.954
Random Forest	98.26%	0.983	0.003	0.983	0.983	0.983	0.980	0.999	0.996

- Check if termination criteria are met (e.g., maximum number of iterations or evaluations).
- If not, repeat steps 2 to 8.

10) Final Solution:
    - Return the best solution(s) obtained throughout the iterations.

D. Classification using Random Forest (RF) Classifier

The Random Forest classifier is an ensemble learning method that combines multiple decision trees to improve classification accuracy and generalization [19]. Each tree in the forest is constructed using a random subset of the training data and a random subset of features. Here's a simplified mathematical representation of the Random Forest classifier: Let

- X be the feature matrix of size $n \times m$, where n is the number of samples and m is the number of features.
- Y be the target vector of size $n \times 1$, containing the class labels for each sample.

The following defines the RF algorithm

1) Choose the number of decision trees to be created (N) and a subset of features to be considered at each split (m_{feat}).
2) For each of the decision tree
    - Randomly select a subset of n_{sample} samples from X (with replacement).
    - Randomly select m_{feat} features from X for buidling the tree.
    - Build a decision tree using the selected data and features.
3) For classification
    a) Input a new data point x_{new} of size $1 \times m$ to be classified.
    b) For each of the N decision trees:
        - Traverse the tree using the x_{new} features.
        - Assign the class label of the leaf node as the predicted class.
    c) Calculate the mode (most frequent class label) of the predicted classes from all trees as the final predicted class label.

In mathematical terms, the Random Forest classifier combines multiple decision trees T_i using a majority voting mechanism:

$$RF(x_{new}) = mode\{T_1(x_{new}), T_2(x_{new}), ..., T_N(x_{new})\} \tag{2}$$

where $RF(x_{new})$ is the final predicted class label for the input x_{new}. The Random Forest algorithm leverages the strength of multiple decision trees to reduce overfitting, improve accuracy, and handle high-dimensional data. It's a versatile and powerful classification method widely used in various domains.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed model was experimented with a migraine dataset consisting of 24 features and 400 instances. The dataset used in the study comprises seven distinct classes, each representing a specific type of migraine. In this study, the classification task involves categorizing migraines into several distinct classes, such as Typical aura with migraine, Typical aura without migraine, Migraine without aura, Basilar type aura, Sporadic hemiplegic migraine, Familial hemiplegic migraine, and others. To train and evaluate the classification model, the dataset is split into two subsets: a training set containing 80% of the data and a test set with the remaining 20%.

The model experimented with the original dataset and classifiers is presented in Table I. As shown in Table I, the model shows an accuracy of 96.25% for the Naïve Bayes classifier and 96.23% for the SVM classifier for 24 attributes from the data. The True Positive (TP) Rate and False Positive (FP) Rate for the Naïve Bayes classifier are 0.963 and 0.056. TP Rate and FP Rate for the SVM classifier are 0.962 and 0.006. Multilayer Perceptron and Random Forest classifiers show accuracy of 92.50% and 90% respectively. The decision tree classifier shows the lowest accuracy i.e. 80%.

Figure 2 shows a comparison of accuracies by applying SMOTE for different classifiers with the original dataset. In

TABLE III
COMPARATIVE ANALYSIS ON RANDOM FOREST CLASSIFIER WITH VARIOUS SEARCH TECHNIQUES

Search Methods	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
PSO Search	0.977	0.004	0.978	0.977	0.977	0.973	1	0.999
Scatter Search	0.983	0.003	0.983	0.983	0.983	0.98	0.999	0.996
Tabu Search	0.957	0.007	0.958	0.957	0.957	0.95	0.995	0.983
Evolutionary Search	0.983	0.003	0.983	0.983	0.983	0.98	0.998	0.997

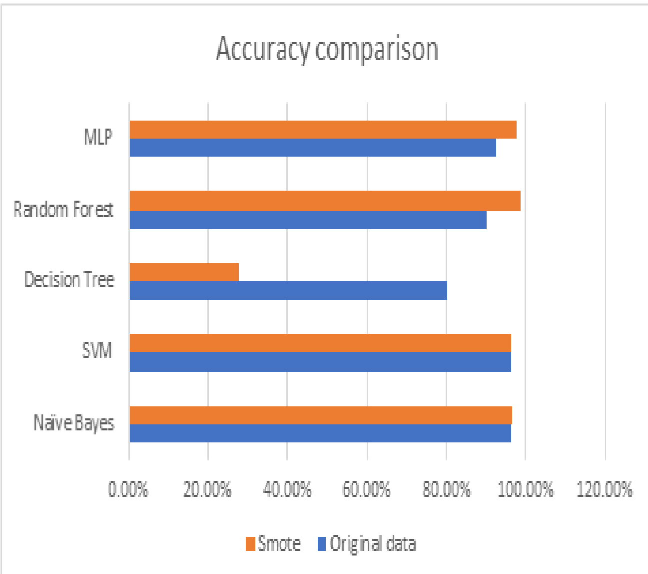

Fig. 2. Accuracy comparison with various models on original Dataset and SMOTE dataset.

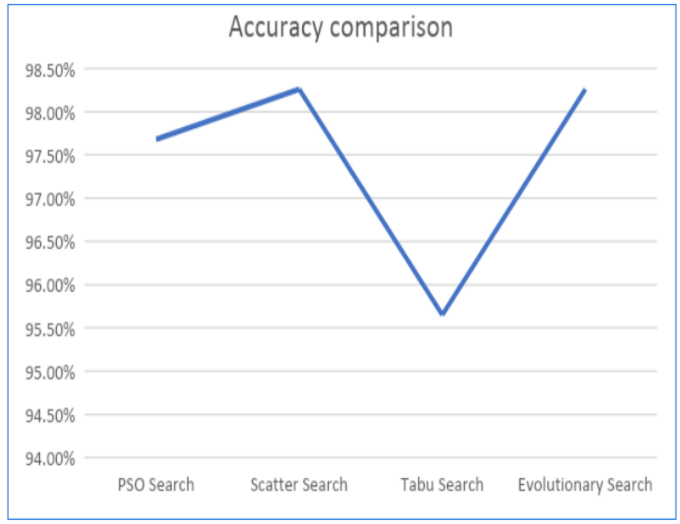

Fig. 3. Accuracy comparison using various search methods.

multilayer perceptron (MLP) the accuracy is 97.68% and after applying SMOTE its 92.50%. Using the decision tree classifier, the accuracy is 80% for original data and 87.83% accuracy after applying smote. After applying SMOTE, some classifiers increased their accuracy but the Random Forest classifier gives more accuracy compared to the other four classifiers. An outstanding accuracy of 98.55% was achieved by the Random Forest classifier. The TP Rate and FP Rate for the Random Forest classifier are 0.986 and 0.002 (Table II). In this classifier, ROC area is 1, and PRC area is 0.999.

The proposed model performance in Table III-D shows the accuracies with various search techniques with RF classifier. 98.26% of accuracy is obtained for the RF Classifier for scatter search. The least performance is recorded for the Tabu search with 95.7%. The suggested model demonstrates promising accuracy when utilizing scatter search for RF Classifier. TP Rate and FP Rate are 0.983 and 0.007 for scatter search. ROC Area and PRC area in scatter search are 0.999 and 0.996.

The graph in Figure 3 compares the accuracy of different random forest classifier retrievals. The variance search method achieved high accuracy when features were reduced to 50%. The study uses a multi-class classifier to classify data into three groups: control, sporadic migraine, and chronic migraine. Feature selection's effectiveness is evaluated by classifying

TABLE IV
FEATURE ANALYSIS ON VARIOUS SEARCH TECHNIQUES

Search Methods	No. of Features Selected	% of Features reduced
Scatter Search	12	50%
Tabu Search	12	50%
PSO Search	15	37.50%
Evolutionary Search	16	33.33%

new data with each method. The goal is to identify relevant features for accurate classification. The graph provides insights into the impact of feature selection on classifier accuracy in this multi-class classification task.

Table IV illustrates the features extracted for various searches. Classifiers also provide better accuracy after feature reduction. Both Scatter Search and Tabu Search algorithms have independently selected 12 features and have a 50% reduction from the original features from the dataset. This represents a significant reduction in the number of features used for classification compared to the original dataset. The

TABLE V
PROPOSED MODEL COMPARISON

Techniques	Precision	Recall	F1-Score	Accuracy
Patrizia et. al [20]	0.474	0.786	0.591	78.7%
Garcia et. al [21]	0.93	0.92	0.92	95.0%
Kwon et. al [22]	0.8707	0.935	-	81.8%
Proposed	0.983	0.983	0.983	98.26%

percentage reduction in features is considerable, resulting in a more concise and efficient representation of the data, which can potentially improve the classifiers' performance and reduce computation time. 15 features were selected in the PSO search, reducing the features by 37.5%. Evolutionary search selects 16 features, reducing features by 33.33%. So the model used a Random Forest classifier which is a suitable classifier and Scatter search is a suitable search method for this migrated data.

Table V gives the proposed model comparison with other state of the art techniques. The proposed model have given good accuracy of 98.26% when compared with the techniques in [20]–[22]. In terms of Precision, recall and F-score, the proposed model have given considerable performance.

V. Conclusion

This study introduces a novel methodology for classifying migraines using optimised feature selection and RF classifier. SMOTE is applied for the class balancing and the relevant features are extracted with the scatter search, followed by classification with RF. The proposed model with 24 variables related to migraine diagnosis were used, leading to a remarkable precision level of 98.26% for the RF model. The features were reduced Subsequently, a second testing phase employed only 18 variables, resulting in an even higher precision of 98%. These outcomes not only confirm the effectiveness of artificial neural networks in accurately classifying different types of migraines but also highlight the further improvement by considering a reduced set of key variables that significantly impact the classification process.

Deep learning has promising applications in headache research, particularly in utilizing autoencoder networks. These networks are well-suited for managing correlated high-dimensional features and can even acquire resilient low-dimensional feature representations in the presence of noise. The authors envision the potential effectiveness of autoencoders in this context and intend to explore this avenue in the upcoming research endeavors.

References

[1] T. Babu, T. Singh, D. Gupta, and S. Hameed, "Colon cancer prediction on histological images using deep learning features and bayesian optimized svm," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 5, pp. 5275–5286, 2021.

[2] R. Haarika, T. Babu, and R. R. Nair, "Insect classification framework based on a novel fusion of high-level and shallow features," *Procedia Computer Science*, vol. 218, pp. 338–347, 2023, international Conference on Machine Learning and Data Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923000169

[3] T. Babu, T. Singh, D. Gupta, and S. Hameed, "Optimized cancer detection on various magnified histopathological colon imagesbased on dwt features and fcm clustering," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 1, pp. 1–17, 2022.

[4] G. Vandewiele, F. De Backere, K. Lannoye, M. Berghe, O. Janssens, S. Hoecke, V. Keereman, K. Paemeleire, F. Ongenae, and F. De Turck, "A decision support system to follow up and diagnose primary headache patients using semantically enriched data," *BMC Medical Informatics and Decision Making*, vol. 18, 11 2018.

[5] B. Krawczyk, D. Simić, S. Simić, and M. Woźniak, "Automatic diagnosis of primary headaches by machine learning methods," *Open Medicine*, vol. 8, no. 2, pp. 157–165, 2013. [Online]. Available: https://doi.org/10.2478/s11536-012-0098-5

[6] A. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, P. Fergus, M. Al-Jumaily, and N. Radi, "A systematic comparison and evaluation of supervised machine learning classifiers using headache dataset," in *Advanced Intelligent Computing Theories and Applications*, D.-S. Huang and K. Han, Eds. Cham: Springer International Publishing, 2015, pp. 101–108.

[7] R. R. Nair, T. Singh, A. Basavapattana, and M. M. Pawar, "Multi-layer, multi-modal medical image intelligent fusion," *Multimedia Tools and Applications*, pp. 1–27, 2022.

[8] R. R. Nair, T. Babu, and T. Singh, "Multiresolution approach on medical image fusion by modified local energy," *Signal, Image and Video Processing*, 2023.

[9] R. R. Nair, T. Singh, R. Sankar, and K. Gunndu, "Multi-modal medical image fusion using lmf-gan-a maximum parameter infusion technique," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 5, pp. 5375–5386, 2021.

[10] A. Charles, "The migraine aura," *CONTINUUM: Lifelong Learning in Neurology*, vol. 24, pp. 1009–1022, 08 2018.

[11] R. Evans, "Diagnostic testing for migraine and other primary headaches," *Neurologic Clinics*, vol. 37, 08 2019.

[12] T. Babu and R. R. Nair, "Colon cancer prediction with transfer learning and k-means clustering," in *Frontiers of ICT in Healthcare*, J. K. Mandal and D. De, Eds. Singapore: Springer Nature Singapore, 2023, pp. 191–200.

[13] W. Lee, I. Min, K. I. Yang, D. Kim, C.-H. Yun, and M. Chu, "Classifying migraine subtypes and their characteristics by latent class analysis using data of a nation-wide population-based study," *Scientific Reports*, vol. 11, 11 2021.

[14] G. Tezel and U. Köse, "Headache disease diagnosis by using the clonal selection algorithm," 05 2011.

[15] Y. Woldeamanuel and R. Cowan, "Computerized migraine diagnostic tools: a systematic review," *Therapeutic Advances in Chronic Disease*, vol. 13, p. 204062232110652, 01 2022.

[16] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res. (JAIR)*, vol. 16, pp. 321–357, 06 2002.

[17] V. Kumar and D. Kumar, "A systematic review on firefly algorithm: Past, present, and future," *Archives of Computational Methods in Engineering*, p. 3269–3291, 04 2020.

[18] M. Laguna and R. Marti, *Scatter Search*, 01 2006, pp. 139–152.

[19] A. K. Dixit, R. R. Nair, and T. Babu, "Analysis and classification of restaurants based on rating with xgboost model," in *2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2022, pp. 1–6.

[20] P. Ferroni, F. M. Zanzotto, N. Scarpato, A. Spila, L. Fofi, G. Egeo, A. Rullo, R. Palmirotta, P. Barbanti, and F. Guadagni, "Machine learning approach to predict medication overuse in migraine patients," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1487–1496, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2001037020302956

[21] Y. Garcia-Chimeno, B. Zapirain, M. Gomez-Beldarrain, B. Fernandez-Ruanova, and J. García-Moncó, "Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data," *BMC Medical Informatics and Decision Making*, vol. 17, 04 2017.

[22] J. Kwon, H. Lee, S. Cho, C.-S. Chung, M. J. Lee, and H. Park, "Machine learning-based automated classification of headache disorders using patient-reported questionnaires," *Scientific Reports*, vol. 10, p. 14062, 08 2020.