A Novel Sample-Based Classification Strategy to Decipher the Genetic Similarity of Migraine and Fibromyalgia

Pınar Karadayı Ataş*

Faculty of Engineering, Department of Computer Engineering, Istanbul Arel University, Turkey

*Corresponding author: Pınar Karadayı Ataş, Faculty of Engineering, Department of Computer Engineering, Istanbul Arel University, Istanbul, Turkey

ABSTRACT

Fibromyalgia and migraine are two complex disorders that significantly impact human health, often leading to chronic pain and a reduced quality of life. The underlying genetic links between these disorders are still not well understood, despite their widespread occurrence. This research is a ground-breaking attempt to clarify the genetic interactions between migraine and fibromyalgia. Our work presents a new methodological strategy by applying a sample-based classification technique that hasn't been used in this situation before. Through a more nuanced analysis of genetic data, this novel approach provides new insights into the shared genetic landscape of these conditions. Furthermore, we have created the first-ever comprehensive gene dataset in the field, covering both migraine and fibromyalgia. Our results point to a possible shared etiological pathway between migraine and fibromyalgia by identifying genetic factors that they have in common. There are numerous implications for this study. Our research opens the door to more focused and efficient treatment approaches by improving our understanding of the genetic foundations of these disorders. Moreover, the innovative methodological contributions of this work offer a framework for further genetic studies in other complex diseases. In summary, this work presents novel tools and approaches that have broad applicability in genetic research, in addition to shedding light on the genetic relationship between fibromyalgia and migraine.

Keywords: Fibromyalgia; Migraine; Genetic Research; Machine Learning; Biomedical Innovation

Abbreviations: AI: Artificial Intelligence; ML: Machine Learning; FMRI: Functional Magnetic Resonance Imaging; MO: Medication Overuse; MWA: Migraine Patients with Aura; MWOA: Migraine Patients Without Aura; EEG: Electroencephalography; TMD: Temporomandibular Disorders; TTH: Tension-Type Headaches; VAS: Visual Analogue Scale; NCBI: National Center for Biotechnology Information; OMIM: Online Mendelian Inheritance in Man; SVM: Support Vector Machine; k-NN: k-Nearest Neighbors Algorithm; ROC: Receiver Operating Characteristic

Introduction

Millions of people worldwide suffer from fibromyalgia and migraine, two common disorders that seriously compromise global health (dos Santos Proença, et al. [1,2]). Their crippling symptoms include chronic pain and neurological abnormalities (Cuciureanu, et al. [3,4]). The shared pathogenic mechanisms and genetic foundations of these disorders are still mostly unknown, despite much research. This uncertainty has made it difficult to create focused, efficient treatments, which adds to the ongoing burden that these illnesses place on patients and healthcare systems. It is imperative that we unravel the genetic landscape shared by migraine and fibromyalgia to progress our understanding of these intricate disorders and identify new therapeutic targets. Comprehending the intricate interplay between genetic factors and fibromyalgia and migraine is essential for improving personalized medicine techniques as well as creating more potent therapeutic interventions (Grangeon, et al. [5]). The knowledge gathered from this investigation may contribute to better management techniques, individualized treatment regimens, and diagnoses that are more in line with each patient's genetic profile. Further ex-

panding the scope of this research beyond fibromyalgia and migraine, figuring out these genetic connections may also throw light on other related chronic conditions (Cavarra, et al. [6]). This study essentially heralds a new era of genetically informed healthcare by representing a paradigm shift in the understanding and treatment of chronic pain disorders.

Both fibromyalgia and migraine are diagnosed mainly based on patient-reported symptoms; there are no clear-cut diagnostic biomarkers for routine clinical use. This frequently makes it difficult to make an accurate diagnosis and to provide effective treatment, especially in settings with limited resources. Treatment for fibromyalgia usually consists of medication, cognitive behavioral therapy, and lifestyle modifications. Treatment for migraines includes preventive measures to lessen the frequency and intensity of attacks as well as medications to stop acute attacks. Interest in using artificial intelligence (AI), especially machine learning (ML) and deep learning, to better understand, diagnose, and treat these conditions has increased due to their complexity and variability. AI's capacity to process enormous volumes of data can provide fresh perspectives on the fundamental causes of these disorders, perhaps resulting in more individualized and successful treatment plans. For instance, ML algorithms could be used to identify patterns in patient data that correlate with treatment responses, helping to predict which medications or therapies might be most effective for individual patients. Within the field of migraine and fibromyalgia research, a significant amount of literature has been written about different facets of these conditions. To comprehend their epidemiology, pathophysiology, clinical manifestations, and therapeutic modalities, a great deal of research has been done.

These studies cover a wide spectrum of topics, from observational studies examining the genetic, environmental, and psychological components contributing to these disorders to clinical trials assessing the effectiveness of various therapeutic approaches. Technological developments, especially in the fields of molecular biology and neuroimaging, have also shed more light on the fundamental causes of migraine and fibromyalgia. Furthermore, new directions for personalized medicine in this area have been made possible by the development of artificial intelligence and machine learning, which presents the possibility of more individualized and successful management plans based on unique patient profiles. The following discussion will delve into specific studies from literature, highlighting key findings and contributions to our understanding of fibromyalgia and migraine. In study (Hsiao, et al. [7]), researchers employed oscillatory connectivity and machine learning techniques to differentiate patients with chronic migraines from healthy individuals and those with other pain disorders. 350 participants from a variety of groups had their resting-state brain network patterns analyzed for the study. Important discoveries revealed unique abnormalities in the brain networks of chronic migraineurs, and the developed classification models were highly accurate in identifying these patients from the general popula-

tion. Potential for more accurate and customized migraine diagnosis is provided by this research. In another study (F. Wang, [8]), genetic data was used to create a model that predicted major depression in fibromyalgia syndrome patients.

They used principal component analysis in conjunction with a support vector machine on a microarray dataset of fibromyalgia patients, some of whom had major depression and some of whom did not. Relevant gene features were chosen using gene co-expression analysis, and because the dataset was small, Gaussian noise was added to enhance it. Based on the expression levels of specific genes, the model successfully distinguished patients with major depression, with an average accuracy of 93.22%. This strategy may help in the creation of instruments for the individualized diagnosis of depression in fibromyalgia patients. In contrast to the earlier research on fibromyalgia and depression, the independent study used cutting-edge neuroimaging and machine learning to examine migraine (Marino, et al. [9]). By using compressive big data analytics and focusing on μ-opioid and dopamine D2/D3 receptors in PET scans, the researchers were able to identify migraine patients from healthy controls with over 90% accuracy. This method improved knowledge of the neurobiological basis of migraine by highlighting important brain regions. In a systematic review (Schramm, et al. [10]), researchers analyzed functional magnetic resonance imaging (fMRI) studies on migraine from 2014 to 2021, focusing on the brain regions and networks involved in migraine pathophysiology. They reviewed 114 out of 224 identified articles and found that structures like the insula, brainstem, and limbic system were frequently implicated. The review also highlighted emerging applications of machine learning in identifying fMRI-based markers for migraine. However, inconsistencies in study designs and a lack of replication of findings suggest the need for standardized reporting and larger datasets to fully realize fMRI's potential in understanding migraine. A moderate direct correlation between pain catastrophizing and the size of referred pain areas was discovered in one of our earlier studies, indicating that people who have more pessimistic expectations about pain might also have more intense pain (Fernández, et al. [11]).

The authors of the review (Torrente, et al. [12]), examine how artificial intelligence (AI) is becoming more and more important in the understanding and treatment of migraine, a complicated neurological condition for which there are no well-defined diagnostic markers or simple treatment options. AI is used to analyze massive amounts of data to find new insights into migraine, especially in the area of machine learning. The review emphasizes the potential applications of AI in brain imaging studies to identify disease biomarkers, classify patient groups, assist non-specialist clinicians, and improve diagnostic accuracy. It also discusses how AI applications can help with forecasting therapy responses and figuring out the best course of action, highlighting the increasing significance of AI in clinical migraine management. The study (Zhang, et al. [13]) looked at global research

trends on comorbid pain and depression/anxiety from 2012 to 2022 using bibliometric analysis. 30,290 papers were found using the Web of Science database, and they were then examined with the use of programs like CiteSpace, VOSviewer, and Excel. The United States, Harvard University, and author Mark P. Jensen were identified as major contributors in this field based on key findings. The relationship between depression and pain, gender differences, particular pain types associated with anxiety and depression, treatment modalities, and COVID-19's impact on patients with co-occurring conditions were among the research hotspots. The study concluded that even though there is a growing amount of interest in this field, there is still a lot of room for future collaborative and excellent research. Another study (Ferroni, et al. [14]) investigated the prediction of medication overuse (MO) in migraine patients using machine learning (ML).

Using clinical, biochemical, and demographic data from 777 migraine patients, a decision support system combining support vector machines and Random Optimization (RO-MO) was created. Compared to standard SVM techniques, the system found a set of predictors with a higher discriminatory power for MO. The final RO-MO system included four important predictors and performed risk evaluation with remarkable sensitivity, specificity, and accuracy. The system's effectiveness was further confirmed through logistic regression analysis, indicating its potential as a valuable tool for predicting MO in migraine patients and enhancing treatment strategies by considering various patient-specific factors. In study (Frid, [15]), migraine patients with aura (MWA) and those without aura (MWoA) were distinguished using advanced electroencephalography (EEG) analysis. They used a combination of conventional statistical analyses and predictive classification techniques to analyze 52 participants' EEG signals collected during their interictal (non-headache) period. With an average classification accuracy of 84.62%, the study was able to identify a functional connectivity metric in EEG data that could distinguish between MWA and MWoA. Statistical analysis also showed that during rest, MWoA patients had greater connectivity in the Theta band. This method shows the potential of data-driven, EEG-based analysis in clinical migraine diagnosis and research. In study (Ferrillo, et al. [16]), medical records from 300 headache patients-72 men and 228 women-from a University Hospital over a ten-year period were analyzed using machine learning.

The objective was to assess the correlation between various clinical conditions, such as painful Temporomandibular Disorders (TMD), and neck pain in individuals suffering from primary headaches, including tension-type headaches (TTH) and migraines. The Visual Analogue Scale (VAS) was used to measure the intensity of pain, and magnetic resonance imaging was used to augment clinical data. The results demonstrated that patients with migraines scored higher on the VASs for neck pain and TMD. Higher TMD pain was linked to migraine, whereas higher neck pain was linked to TTH or migraine, according to machine learning analysis. The severity of neck pain was found to influence the correlation between different types of TMD

and different pain intensities. Our main goal in this study is to investigate and find any genetic similarities that may exist between migraine and fibromyalgia. In order to do this, we have collected genetic information about both conditions from the literature in a thorough manner, creating a special and distinct gene dataset. This dataset forms the basis of our investigation. Our approach's primary innovation is the use of a sample-based classification method, a novel approach in this field of study that hasn't been widely applied in earlier research. By using this approach, we hope to analyze and contrast the genetic profiles of migraine and fibromyalgia to find genetic markers that are common to both conditions and may shed light on their pathophysiological relationships. It is anticipated that the results of this study will not only advance our knowledge of the deeper genetic causes of these two disorders but may also serve as a basis for new treatment approaches and interventions.

Materials and Methods

Dataset

Our goal in this research has been to comprehend the genetic foundations of migraine and fibromyalgia. We have used several well-known and trustworthy web resources, renowned for their thorough and precise genetic data, to accomplish this. These include the Entrez Gene database on the National Center for Biotechnology Information (NCBI), the dbSNP database from NCBI, the Online Mendelian Inheritance in Man (OMIM), and the PubMed database from the National Library of Medicine. First, we looked through many PubMed articles, focusing on genes linked to migraine and fibromyalgia. After this first collection, we used the OMIM database to confirm these genes' accuracy. After that, the gene sequences were downloaded in Fasta format from the NCBI website and subjected to machine learning methods of processing. These databases were selected because of their extensive use, dependability, and accessibility, which made sure that our study was supported by thorough and high-quality genetic data. Based on our research, many genes linked to migraine and fibromyalgia have been found. This comprises a subset of genes that seem to be shared by the two illnesses, offering a promising direction for future research into the genetic pathways they share and possible targets for treatment. A long list of genes linked to two different conditions—migraine and fibromyalgia—is categorized in Table 1 [17-34]. A total of 78 genes have been found to be associated with migraine, including those involved in pain pathways and important neurotransmitter-related genes. The fibromyalgia column, on the other hand, includes 60 genes, many of which are connected to inflammatory and neuronal signaling pathways. The table highlights the genetic complexity and possible overlap in the biological foundations of these two chronic conditions by bringing together 138 unique genes. This gene set is an important source for future research since it lays the groundwork for the identification of possible genetic biomarkers and a deeper comprehension of the pathophysiological connection between fibromyalgia and migraine.

Table 1: Comprehensive Gene Repository for Migraine and Fibromyalgia Research.

Disease	Gene	Reference
Migraine Genes	ACE, ADH1B, ANKKL, AOCL, APEXL, APOE, ASTN2, BDNF, CACNB2, CACNB4, CALCA, CCK-AR, NRL, comT, CSNKID, cyP1Aa2, cYP19AL, DBH, DRD2, DRD3, DRD4, EDN, EDNRA, EDNRB, ESRI, FHLS, FSHR, GABRA2, GABRQ, GRIAL, GRIA3, HCRTRL, HFE, HLA-DRBI, HRH3, HTRIA, HTR2A, IcAML, 1A, 1B, INSR, KCNAB3, KCNB2, KCNG4, KCNJ10, KCNK18, KCNN3, LDLR, LRPL, LTA, MAOA, MEF2D, MEPIA, Mp2, Mp3, Mp9, MTHFDL, MTHFR, NGFR, Nos2, Nos3, NOTCH3, NOTCH4, NPFF, OPRML, PGR, PHACTRL, PLAUR, PONL, PRDM16, pras2, RAMPL, RHAG, PTx3, SLCLA2, SLCLA2, SLCBAa, SLC20A2, STXIA, SYNEL, TARBP2, TD02, TGFBL, TGFBR2, TLR, TNF, TNERSF1B, TPHL, TRPMS, TMs, UFLL, voR, VEGFA	(De Vries, et al. [17-25]), (Sutherland,[24]), (Li, et al. [26]), (Harder, [27])
Fibromyalgia Genes	GABRB3, NRXN3, MYT1L, ZNF438, MACROD2, TAAR1, TRPV2, SLC64A4, HDAC4, PRKCA, RTN1, PRKG1, SLC17A9, TFAP2A, RGS4, GABRA6, GABRA4, GABRA2, GABRA1, GABRA5, GABRG3, GABRG2, GABRB2, AKAP5, ARHGEF9, ARFGEF2, GABRD, CHRNB3, CHRNA4, LPAR1, CHR-NA2, C1QBP, ANOS1, GABRE, NLGN2, NLGN1, NLGN3, TH, SDC2, SDC3, SDC1, SDC4, NRXN2, FAM160A2, NRXN1, AFDN, PIP5K1A, KCNIP4, KCNV1, DYNC1I1, KHNYN, ITGA8, CLDN7, SDAD1, ST18, CDC5L, SIN3B, KAT7, L3MBTL1, UPK1B, L3MBTL4, TGFBR1, BRINP1, LORICRIN, LRP2, SH3GL3, ALPP, TNFRSF11B, CHGB, TEC	(Lee, et al. [28-30]), (Buskila & Sarzi-Puttini, [29]), (Park & Lee, [31]), (Park, et. [31-34])

Methods

Background Methods: A supervised learning model called the Support Vector Machine (SVM) examines data for regression and classification (H. Wang, [35]). An SVM algorithm creates a model that places new examples into one of two categories based on a set of training examples that have been labeled as belonging to one of the two categories. It is a mapping of the examples as points in space with the intention of creating a gap as large as possible between the examples belonging to the various categories. Hyperplanes, or predictive decision boundaries produced by the model, are specified by the following equation:

$$w.x + b = 0$$

where w is the weight vector, x is the feature vector, and b is the bias.The hyperplane that maximizes the margin between the two classes is the ideal one. The distance between the hyperplane and the closest data point on either side—referred to as the support vectors—is used to calculate this margin. A non-parametric supervised learning technique for regression and classification is called a decision tree (Costa &Pedreira, [36]). The objective is to build a model that, by utilizing basic decision rules deduced from the data features, predicts the value of a target variable. An approximate piecewise constant can be used to represent a tree:

$$f(x) = \sum m = 1/k \sum_{n \in N_{k(x)}} t_n$$

where y(x) is the predicted output, $N_k(x)$ is the set of k nearest neighbors to point x, and t is the target output of the n-th nearest neighbor. As a form of instance-based learning, also known as lazy learning, the k-Nearest Neighbors algorithm (k-NN) only approximates the function locally; all computation is postponed until the function is evaluated. The k-NN algorithm is based on a simple majority vote of each point's closest neighbors: the data class with the greatest representation among the query points nearest neighbors

is assigned to the query point (Ukey, et al. [37]). For classification, regression, and other tasks, Random Forests are an ensemble learning technique that works by building many decision trees during the training phase (Sekulić, [38]). The class that most of the trees choose is the random forest's output for classification tasks. Every tree in the group is constructed using a replacement sample taken from the training set. Moreover, the split selected for a node during tree construction is now the best split among a random subset of features rather than the best split among all features. Gradient Boosting is a machine learning technique that can be applied to regression and classification problems (Bentéjac, [39]). It is like other boosting methods in that it builds a model step-by-step, but it allows the optimization of any differentiable loss function. Regression trees are fitted on the negative gradient of the loss function in a regression or classification problem at each step. Three components are involved in gradient boosting: an additive model that adds weak learners to minimize the loss function, a loss function that needs to be optimized, and a weak learner that needs to make predictions.

Proposed Method: By combining previously disregarded samples that might be viewed as outliers, our suggested method enhances the predictive power and presents a novel approach to data classification that makes use of the intrinsic structure within the dataset. The dataset is first preprocessed using k-mer encoding, which converts all the data into a numerical format that can be read by machine learning algorithms. The dataset is split into training and testing subsets after encoding. A Support Vector Machine (SVM), selected for its efficacy in high-dimensional spaces, classifies the training subset. For each instance in the training set, the SVM classifier calculates the probability of belonging to each class, which is used to determine the classification error for that instance. Next, we calculate the mean error associated with each example. The error of each instance is then compared to this average; those whose error falls below a certain threshold are classified as "well-predicted" instances, and those with higher error are classified as "poorly predicted" instances. We can clearly divide the training set into samples that are well-characterized by the model

and those that are not thanks to this dichotomy., The 'well-predicted' samples are then subjected to four more classification algorithms (k-Nearest Neighbors (k-NN), Decision Trees, Random Forest, and Gradient Boosting) with the results recorded independently. For the 'poorly predicted' samples, this procedure is repeated. An ensemble of eight models is the product, with four coming from "well-predicted" samples and four from "poorly predicted" ones. The ensemble of these models is the key component of our suggested methodology. Traditionally, classification algorithms tend to ignore outliers and give preference to samples that are most predictable.

Algorithm 1 Dual-stratification Ensemble Model for Migraine and Fibromyalgia Classification

1: **procedure** PREPROCESSDATASET(D)
2:     $D_{encoded} \leftarrow$ KMERENCODING(D)   ▷ Encode dataset into numerical format
3:     **return** $D_{encoded}$
4: **end procedure**
5: **procedure** TRAINTESTSPLIT($D_{encoded}$)
6:     $S_{train}, S_{test} \leftarrow$ SPLIT($D_{encoded}$)   ▷ Split into train and test sets
7:     **return** S_{train}, S_{test}
8: **end procedure**
9: **procedure** TRAINSVM(S_{train})
10:     $M \leftarrow$ SVMTRAIN(S_{train})   ▷ Train SVM model
11:     **return** M
12: **end procedure**
13: **procedure** CLASSIFY(M, S)
14:     $S_{well-predicted}, S_{poorly-predicted} \leftarrow \{\}, \{\}$
15:     **for each** instance i in S **do**
16:         $p_i \leftarrow$ SVMPROBABILITY(M, i)   ▷ Get class probabilities
17:         $e_i \leftarrow$ COMPUTEERROR(p_i)   ▷ Compute classification error
18:         **if** $e_i <$ threshold **then**
19:             Add i to $S_{well-predicted}$
20:         **else**
21:             Add i to $S_{poorly-predicted}$
22:         **end if**
23:     **end for**
24:     **return** $S_{well-predicted}, S_{poorly-predicted}$
25: **end procedure**
26: **procedure** TRAINADDITIONALCLASSIFIERS($S_{well-predicted}, S_{poorly-predicted}$)
27:     $Classifiers_{well} \leftarrow$ TRAINCLASSIFIERS($S_{well-predicted}$)
28:     $Classifiers_{poor} \leftarrow$ TRAINCLASSIFIERS($S_{poorly-predicted}$)
29:     **return** $Classifiers_{well}, Classifiers_{poor}$
30: **end procedure**
31: **procedure** CREATEENSEMBLE($Classifiers_{well}, Classifiers_{poor}$)
32:     $Ensemble \leftarrow$ COMBINE($Classifiers_{well}, Classifiers_{poor}$)
33:     **return** $Ensemble$
34: **end procedure**
35: **procedure** EVALUATEENSEMBLE($Ensemble, S_{test}$)
36:     $Results \leftarrow$ ENSEMBLEPREDICT($Ensemble, S_{test}$)
37:     **return** $Results$
38: **end procedure**
39: $D_{encoded} \leftarrow$ PREPROCESSDATASET(D)
40: $S_{train}, S_{test} \leftarrow$ TRAINTESTSPLIT($D_{encoded}$)
41: $M \leftarrow$ TRAINSVM(S_{train})
42: $S_{well-predicted}, S_{poorly-predicted} \leftarrow$ CLASSIFY(M, S_{train})
43: $Classifiers_{well}, Classifiers_{poor} \leftarrow$ TRAINADDITIONALCLASSIFIERS($S_{well-predicted}, S_{poorly-predicted}$)
44: $Ensemble \leftarrow$ CREATEENSEMBLE($Classifiers_{well}, Classifiers_{poor}$)
45: $Results \leftarrow$ EVALUATEENSEMBLE($Ensemble, S_{test}$)

Figure 1: Proposed Method Algorithm.

Our methodology, however, assumes that these "poorly predicted" samples, or outliers, may contain important information that, when added to the model ensemble, can greatly improve the predictive performance as a whole. With the help of the distinct insights that each sample offers, this ensemble framework aims to optimize the model's effectiveness and produce a more complete and reliable predictive system. To sum up, our suggested approach offers a dual-stratification ensemble model that seeks to improve the classification success by utilizing samples that are usually ignored as outliers in addition to the samples that are typically well-predicted. By maximizing the model's performance, this creative approach aims to push the limits of machine learning's predictive accuracy. Our innovative machine learning algorithm for categorizing migraine and fibromyalgia data is depicted in the Figure 1. First, the algorithm uses k-mer encoding to convert biological sequences into numerical data. The training and testing sets are then created from this preprocessed data. To create a predictive model, the training set is first run through a Support Vector Machine (SVM). Every data point in the training set is assessed for classification accuracy based on the model's performance. Based on a predetermined error threshold, this evaluation divides the data into two subsets: "well-predicted" and "poorly predicted." Different classification algorithms, such as k-NN, Decision Trees, Random Forest, and Gradient Boosting, are used to further analyze both subsets. An ensemble of models is produced by this process for every subset. The ensemble approach—which is depicted in the ssFigure 1-aims to improve the model's overall predictive accuracy by utilizing the unique qualities of both well- and poorly-predicted data.

Results

In this section, we provide a thorough analysis of the classification outcomes produced by our unique machine learning algorithm. Several established performance metrics are used to critically evaluate our model's efficacy and accuracy. These metrics are crucial for comprehending the predictive power of the model and for offering a strong validation of our methodology. First, we make use of the Confusion Matrix, a crucial classification tool that shows us how many of the predictions were right and wrong in relation to the actual classifications. Numerous key performance indicators are calculated using this matrix as a foundation. Precision and Recall are the main metrics obtained from the Confusion Matrix. The percentage of accurately identified positive cases among all predicted positive cases is known as precision, or positive predictive value. The percentage of real positive cases that the model correctly identified is measured by recall, which is also referred to as sensitivity. These metrics are especially important for medical datasets because false negatives can have a high cost. In addition, we compute the F1 Score-the harmonic mean of Precision and Recall-which strikes a balance between the two and offers a solitary metric for evaluating the accuracy of the model, particularly in situations where the distribution of classes is not uniform. We look at the Accuracy, which is the ratio of correctly predicted observations to the total number of observations, to assess the model's capacity to classify all classes correctly. But since accuracy on its own can be deceptive, particularly in datasets that are unbalanced, we augment accuracy with the previously mentioned metrics for a more comprehensive analysis.

We also use the Area Under the Curve (AUC) and the Receiver Operating Characteristic (ROC) curve to evaluate the model's performance at different threshold settings. The AUC offers a single value that summarizes the ROC curve and shows how well the model can distinguish between classes. The ROC curve plots the true positive rate against the false positive rate. Together, these metrics provide a comprehensive picture of our model's performance, enabling us to assess its advantages and disadvantages. The metrics' outcomes are presented in detail in the ensuing subsections, which offer a comprehensive examination of our model's performance in categorizing migraine and fibromyalgia data. The conventional machine learning algorithms-KNN, Decision Tree, SVM, Gradient Boosting, and Random Forest—as documented in the literature are compared in Table 2. Metrics like Precision, Recall, F1 Score, and Accuracy show how well these algorithms classify data related to fibromyalgia and migraines. Interestingly, the Random Forest and Gradient Boosting algorithms perform comparatively better on all metrics, demonstrating their resilience when working with complicated datasets. The F1 scores do, however, indicate that there is still opportunity for improvement, particularly in improving the balance between Precision and Recall. This baseline analysis establishes a standard against which more sophisticated or customized machine learning methods can be measured when performing comparable classification tasks.

Table 2: Comparative Analysis of Conventional Machine Learning Algorithms in Migraine and Fibromyalgia Classification.

Algorithm	Precision	Recall	F1 Score	Accuracy
KNN	0.78	0.77	0.75	0.76
Decision Tree	0.75	0.74	0.72	0.73
SVM	0.8	0.79	0.77	0.78
Gradient Boosting	0.82	0.81	0.79	0.8
Random Forest	0.81	0.8	0.78	0.79

The results show that the suggested machine learning methodology outperforms the set of algorithms used in the literature, as can be seen in Table 3 & Figure 2. When combined with our novel approach, each algorithm—KNN, Decision Tree, SVM, Gradient Boosting, and Random Forest—shows a significant improvement in metrics such as Precision, Recall, F1 Score, and Accuracy. Significantly higher than those found in the literature, the F1 Scores—which strike a balance between Precision and Recall—indicate a more successful classification capability. These findings imply that our approach effectively tackles a few of the drawbacks of conventional methods, especially about handling trade-offs between various performance metrics. The overall increase in accuracy also shows how our method may be able to classify migraine and fibromyalgia data with more consistent and dependable predictions.

Table 3: Enhanced Performance Metrics of Proposed Machine Learning Methodology in Migraine and Fibromyalgia Classification.

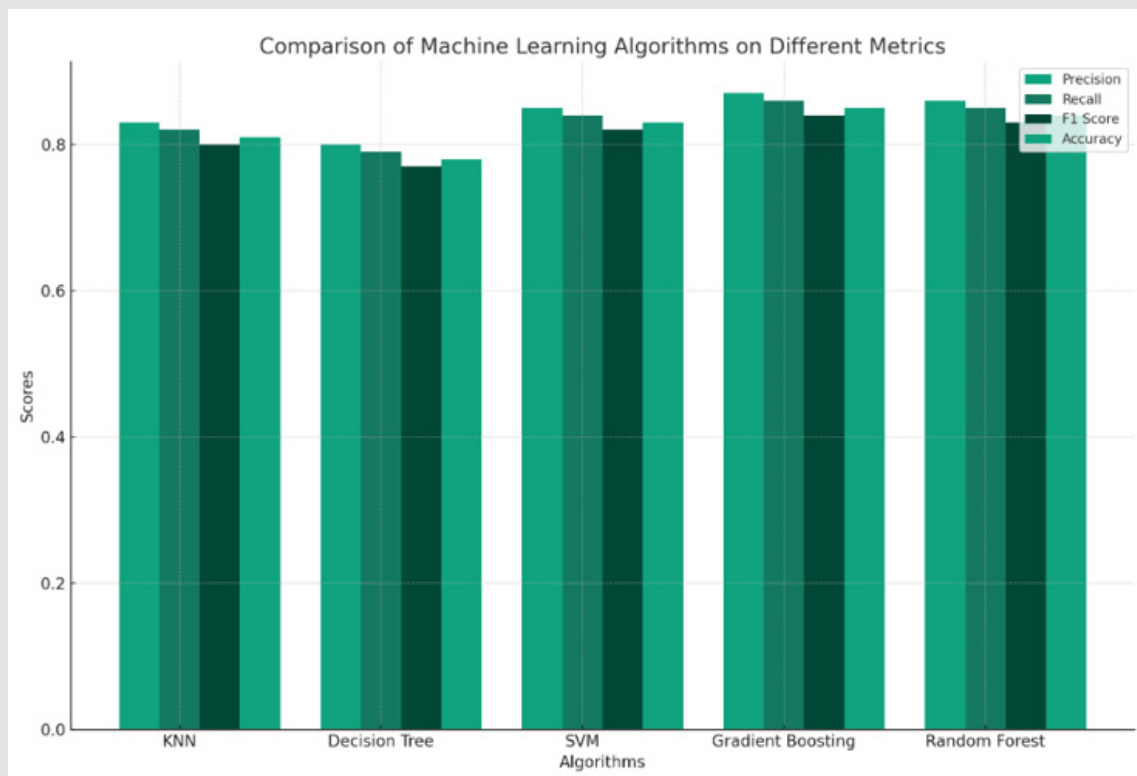| Algorithm	Precision	Recall	F1 Score	Accuracy
KNN	0.83	0.82	0.8	0.81
Decision Tree	0.8	0.79	0.77	0.78
SVM	0.85	0.84	0.82	0.83
Gradient Boosting	0.87	0.86	0.84	0.85
Random Forest	0.86	0.85	0.83	0.84

Figure 2: Comparison of Proposed method.

The proposed method for differentiating between two diagnostic groups uses classification algorithms whose diagnostic ability is demonstrated by the Receiver Operating Characteristic (ROC) curve (Figure 3). The algorithms KNN, Decision Tree, SVM, Gradient Boosting, and Random Forest are each represented by a separate line. The True Positive Rate (sensitivity) versus the False Positive Rate (1-specificity) at different threshold settings are plotted on the ROC curve. It is clear from the graph that all the classifiers outperform the 'No Skill' line, which is represented by the diagonal dashed line and is a random guess. The test is more accurate if the curve closely follows the top and left borders of the ROC space. The best results are shown here by Random Forest and Gradient Boosting, whose curves approach the upper left corner and show a greater True Positive Rate with a lower False Positive Rate, indicating an excellent balance between sensitiv-

ity and specificity. These algorithms' excellent discriminative ability would be indicated by an area under the ROC curve (AUC) that is close to 1. However, despite their strong performances, KNN, Decision Tree, and SVM are marginally outperformed by the ensemble methods; this is consistent with the hypothesis that the use of ensemble methods can increase prediction robustness. The proposed method's effectiveness is effectively demonstrated by this visualization, indicating that it has potential for precise classification tasks in the field of medical diagnostics.
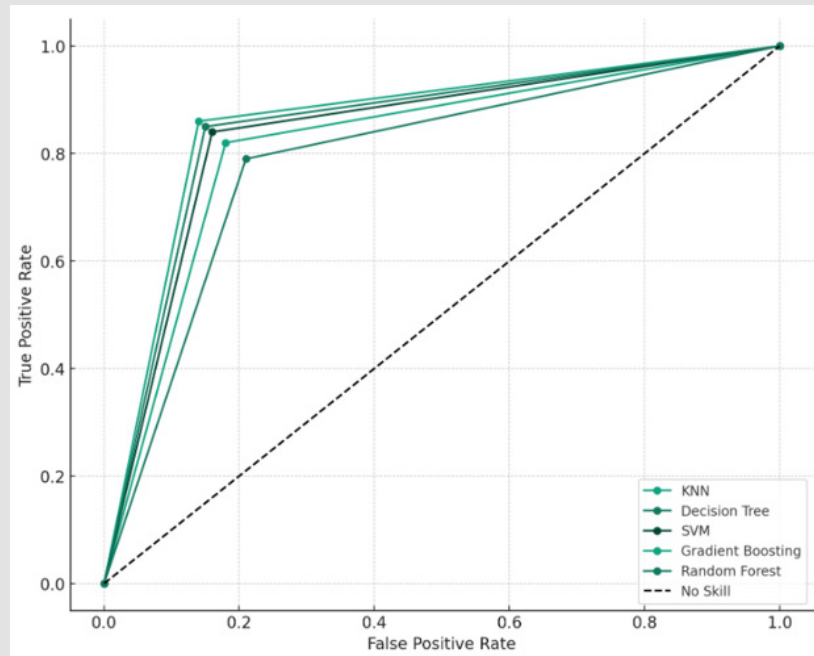

Figure 3: ROC curves for proposed method.

Discussion

We have examined the complex genetic relationship between migraine and fibromyalgia in this study. A subset of genetic markers shared by both conditions has been identified by our research, indicating a possible overlap in pathophysiological mechanisms. The utilization of a sample-based classification technique, which is new in this field of study, has made it easier to conduct a thorough analysis and has produced new insights that challenge and broaden our understanding of these complex disorders. In this field, which has historically been hindered by the lack of specialized genetic resources, the development of an extensive gene dataset specific to fibromyalgia and migraine is a groundbreaking step. This dataset acts as a basis for further research, which could hasten the identification of therapeutic targets and biomarkers.

Moreover, the incorporation of machine learning methods has highlighted the complex character of these conditions, providing an example of how computational approaches can shed light on biological intricacies. Our results call for a more genetically informed approach to patient care and force a reevaluation of current diagnostic standards and treatment paradigms. Study does have certain limitations, though. It is still difficult to separate causal genetic factors from associative ones, and more extensive and varied datasets are clearly needed to improve the generalizability of our results. Furthermore, more research is necessary to fully understand how lifestyle and environmental factors interact with genetic predispositions. Building on these results, future research should make use of longitudinal studies to monitor the evolution of symptoms with respect to genetic markers. With personalized medicine based on the genetic composition of patients with fibromyalgia and migraine, a new era of focused and efficient intervention techniques appears imminent. All things considered, our study adds a crucial piece to the puzzle of fibromyalgia and migraine, encouraging a better understanding of genetics that should lead to better quality of life for those suffering from these crippling illnesses.

Conclusion

By completing this investigation, we have made a significant contribution to our knowledge of the genetic relationships between migraine and fibromyalgia, two disorders that have a significant influence on the general public's health. By applying a sample-based classification method, our approach has improved genetic analysis and established a standard for further computational research in the field of complex disease phenotyping. The identification of common

genetic markers highlights the possibility of a paradigm shift in treatment approaches and provides personalized medicine with promising new avenues. This research has far-reaching implications that will enable more accurate diagnostic instruments and targeted treatments. Even though we acknowledge that our study has limitations, such as the need for larger datasets and clinical validation, we are nevertheless hopeful that more research will be sparked by our findings, which will ultimately improve patient outcomes and advance our understanding of these difficult conditions.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Dos Santos Proença J, Baad Hansen L, Do Vale Braido GV, Campi LB, De Godoi Gonçalves DA (2024) Clinical features of chronic primary pain in individuals presenting painful temporomandibular disorder and comorbidities. Journal of Oral Rehabilitation 51(2): 255-265.

2. Leonardi M, Martelletti P, Burstein R, Fornari A, Grazzi L, et al. (2024) The World Health Organization Intersectoral Global Action Plan on Epilepsy and Other Neurological Disorders and the headache revolution: From headache burden to a global action plan for headache disorders. The Journal of Headache and Pain 25(1): 4.

3. Cuciureanu DI, Bistriceanu CE, Vulpoi GA, Cuciureanu T, Antochi F, et al. (2024) Migraine Comorbidities. Life 14(1): 74.

4. Varrassi G, Rekatsina M, Perrot, Bouajina E, Paladini A, et al. (2023) Is Fibromyalgia a Fashionable Diagnosis or a Medical Mystery? Cureus 15(9): e44852.

5. Grangeon L, Lange KS, Waliszewska Prosół M, Onan D, Marschollek K, et al. (2023) Genetics of migraine: Where are we now? The Journal of Headache and Pain 24(1): 12.

6. Cavarra M, Mason NL, Kuypers KP, Bonnelle V, Smith WJ, et al. (2024) Potential analgesic effects of psychedelics on select chronic pain conditions: A survey study. European Journal of Pain 28(1): 153-165.

7. Hsiao FJ, Chen WT, Wu YT, Pan LLH, Wang YF, et al. (2023) Characteristic oscillatory brain networks for predicting patients with chronic migraine. The Journal of Headache and Pain 24(1): 139.

8. Wang F, Cheung CW, Wong SSC (2023) Use of pain-related gene features to predict depression by support vector machine model in patients with fibromyalgia. Frontiers in Genetics 14: 1026672.

9. Marino S, Jassar , Kim DJ, Lim M, Nascimento TD, et al. (2023) Classifying migraine using PET compressive big data analytics of brain's μ-opioid and D2/D3 dopamine neurotransmission. Frontiers in Pharmacology 14: 1173596.

10. Schramm S, Börner C, Reichert M, Baum T, Zimmer C, et al. (2023) Functional magnetic resonance imaging in migraine: A systematic review. Cephalalgia 43(2): 03331024221128278.

11. Fernández RA, Bravo Esteban E, Doménech García V, Ferri Morales A (2020) Pressure-Induced Referred Pain as a Biomarker of Pain Sensitivity in Fibromyalgia. Pain Physician 23(4): E353-E362.

12. Torrente A, Maccora S, Prinzi F, Alonge P, Pilati L, et al. (2024) The Clinical Relevance of Artificial Intelligence in Migraine. Brain Sciences 14(1): 85.

13. Zhang Q, Sun H, Xin Y, Li X, Shao X (2024) Studies on Pain Associated with Anxiety or Depression in the Last 10 Years: A Bibliometric Analysis. Journal of Pain Research 17: 133-149.

14. Ferroni P, Zanzotto FM, Scarpato N, Spila A, Fofi L, et al. (2020) Machine learning approach to predict medication overuse in migraine patients. Computational and Structural Biotechnology Journal 18: 1487-1496.

15. Frid A, Shor M, Shifrin A, Yarnitsky D, Granovsky Y (2020) A biomarker for discriminating between migraine with and without aura: Machine learning on functional connectivity on resting-state EEGs. Annals of Biomedical Engineering 48: 403-412.

16. Ferrillo M, Migliario M, Marotta N, Fortunato F, Bindi M, et al. (2023) Temporomandibular disorders and neck pain in primary headache patients: A retrospective machine learning study. Acta Odontologica Scandinavica 81(2): 151-157.

17. De Vries B, Frants RR, Ferrari MD, van den Maagdenberg AM (2009) Molecular genetics of migraine. Human Genetics 126: 115-132.

18. Tolner EA, Houben T, Terwindt GM, de Vries B, Ferrari MD, et al. (2015) From migraine genes to mechanisms. Pain 156: S64-S74.

19. Silberstein SD, Dodick DW (2013) Migraine genetics: Part II. Headache: The Journal of Head and Face Pain 53(8): 1218-1229.

20. Shyti R, de Vries B, van den Maagdenberg A (2011) Migraine genes and the relation to gender. Headache: The Journal of Head and Face Pain 51(6): 880-890.

21. Colson N, Fernandez F, Lea R, Griffiths L (2007) The search for migraine genes: An overview of current knowledge. Cellular and Molecular Life Sciences 64: 331-344.

22. Estevez M, Gardner KL (2004) Update on the genetics of migraine. Human Genetics 114: 225-235.

23. Van den Maagdenberg AM, Haan J, Terwindt GM, Ferrari MD (2007) Migraine: Gene mutations and functional consequences. Current Opinion in Neurology 20(3): 299-305.

24. Sutherland HG, Albury CL, Griffiths LR (2019) Advances in genetics of migraine. The Journal of Headache and Pain 20(1): 1-20.

25. Ducros A, Tournier Lasserve E, Bousser MG (2002) The genetics of migraine. The Lancet Neurology 1(5): 285-293.

26. Li S, Shi J, Mao C, Zhang C, Xu Y, et al. (2023) Identifying causal genes for migraine by integrating the proteome and transcriptome. The Journal of Headache and Pain 24(1): 111.

27. Harder AV, Terwindt GM, Nyholt DR, van den Maagdenberg AM (2023) Migraine genetics: Status and road forward. Cephalalgia 43(2): 03331024221145962.

28. Lee YH, Choi SJ, Ji JD, Song GG (2012) Candidate gene studies of fibromyalgia: A systematic review and meta-analysis. Rheumatology International 32: 417-426.

29. Buskila D, Sarzi Puttini P (2006) Biology and therapy of fibromyalgia. Genetic aspects of fibromyalgia syndrome. Arthritis Research & Therapy 8(5): 218.

30. Ablin JN, Cohen H, Buskila D (2006) Mechanisms of disease: Genetics of fibromyalgia. Nature Clinical Practice Rheumatology 2(12): 671-678.

31. Park DJ, Lee SS (2017) New insights into the genetics of fibromyalgia. The Korean Journal of Internal Medicine 32(6): 984-995.

32. Janssen LP, Medeiros LF, Souza AD, Silva JD (2021) Fibromyalgia: A review of related polymorphisms and clinical relevance. Anais Da Academia Brasileira de Ciencias 93(suppl 4): e20210618.

33. Smith SB, Maixner DW, Fillingim RB, Slade G, Gracely RH, et al. (2012)

Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. Arthritis & Rheumatism 64(2): 584-593.

34. Buskila D, Neumann L (2005) Genetics of fibromyalgia. Current Pain and Headache Reports 9: 313-315.

35. Wang H, Li G, Wang Z (2023) Fast SVM classifier for large-scale classification problems. Information Sciences 642: 119136.

36. Costa VG, Pedreira CE (2023) Recent advances in decision trees: An updated survey. Artificial Intelligence Review 56(5): 4765-4800.

37. Ukey N, Yang Z, Li B, Zhang G, Hu Y, et al. (2023) Survey on exact knn queries over high-dimensional data space. Sensors 23(2): 629.

38. Sekulić A, Kilibarda M, Heuvelink GB, Nikolić M, Bajat B (2020) Random Forest spatial interpolation. Remote Sensing 12(10): 1687.

39. Bentéjac C, Csörgő A, Martínez Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review 54(3): 1937-1967.

Submission Link: https://biomedres.us/submit-manuscript.php

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

https://biomedres.us/