AIS ARTIFICIAL INTELLIGENCE STUDIES

Analysis of EMG Signals Using Stacking Model and Deep-MLP in the Detection of Neuromuscular Disorders

İsmail Çalıkuşu[*a], **Volkan Kaya**[b]

[a,*] Nevşehir Hacı Bektaş Veli University, Vocational High School, Department of Electronic and Automation 050300 - Nevşehir, Türkiye
ORCID: 0000-0002-6640-7917

[b]Erzincan Binali Yıldırım University, Faculty of Engineering and Architecture, Department of Computer Engineering 24002-Erzincan, Türkiye
ORCID: 0000-0001-6940-3260

[*] Corresponding author.
e-mail: ismailcalikusu@nevsehir.edu.tr

ABSTRACT

The similarity of clinical symptoms in neuromuscular diseases, particularly myopathy and neuropathy, significantly complicates diagnosis based on superficial electromyography (sEMG) signals. While deep learning approaches in the literature offer high success rates, their low explainability limits clinical confidence. This study aims to develop a diagnostic system with both high accuracy and interpretability using handcrafted features extracted from sEMG signals and hybrid artificial intelligence models. The study utilized a dataset of 241 participants with recordings from the Biceps Brachii and Deltoid muscles, encompassing healthy, myopathy, and neuropathy classes. Forty-three features encompassing time, frequency, and nonlinear dynamics were extracted from the signals, and the 20 most distinctive features were identified using the Mutual Information method. SVM, Random Forest, Deep-MLP, and a Stacking architecture combining these models were used in the classification phase. Experimental results showed that the Stacking model exhibited the best performance with 80.26% accuracy in the three-class classification. In binary distinctions, a 93.17% success rate was achieved in the "Healthy-Neuropathy" classification. Furthermore, in the most difficult-to-distinct pair, "Myopathy-Neuropathy," the Deep-MLP model successfully modelled the heterogeneous nature of myopathic signals with up to 91% accuracy. These findings demonstrate that multidimensional feature sets and ensemble learning methods offer a non-invasive, reliable, and clinically interpretable solution for the early diagnosis of neuromuscular diseases.

Keywords: EMG, Neuromuscular Diseases, Machine Learning, Stacking, Deep-MLP.

Anahtar Kelimeler: EMG, Nöromüsküler Hastalıklar, Makine Öğrenmesi, Stacking, Deep-MLP.

Submitted: 15.12.2025
Revised: 24.12.2025
Accepted: 25.12.2025

doi:10.30855/ais.2025.08.02.06

Nöromüsküler Bozuklukların Tespitinde Yığınlama Modeli ve Deep-MLP Kullanılarak EMG Sinyallerinin Analizi

ÖZ

Nöromüsküler hastalıkların, özellikle myopati ve nöropatinin klinik semptomlarının benzerliği, yüzeysel elektromiyografi (sEMG) sinyalleri üzerinden yapılan tanıyı belirgin bir şekilde zorlaştırmaktadır. Literatürdeki derin öğrenme yaklaşımları yüksek oranda başarı sunsa da düşük açıklanabilirlikleri klinik güveni sınırlamaktadır. Bu çalışma, sEMG sinyallerinden çıkarılan el yapımı (handcrafted) öznitelikler ve hibrit yapay zekâ modelleri ile hem yüksek doğruluklu hem de yorumlanabilir bir tanı sistemi geliştirmeyi amaçlamaktadır. Çalışmada, Biceps Brachii ve Deltoid kaslarından alınan kayıtlara sahip; sağlıklı, myopati ve nöropati sınıflarını içeren 241 katılımcılı bir veri seti kullanılmıştır. Sinyallerden zaman, frekans ve lineer olmayan dinamikleri kapsayan 43 öznitelik çıkarılmış, Mutual Information yöntemiyle en ayırt edici 20 öznitelik belirlenmiştir. Sınıflandırma aşamasında SVM, Random Forest, Deep-MLP ve bu modelleri birleştiren Stacking mimarisi kullanılmıştır. Deneysel sonuçlar, üç sınıflı sınıflandırmada Stacking modelinin %80,26 doğrulukla en iyi performansı sergilediğini göstermiştir. İkili ayrımlarda ise "Sağlıklı-Nöropati" sınıflandırmasında %93,17 başarı elde edilmiştir. Ayrıca, ayırt edilmesi en güç olan "Myopati-Nöropati" ikilisinde Deep-MLP modeli %91'e varan doğrulukla myopatik sinyallerin heterojen yapısını başarıyla modellemiştir. Bu bulgular, çok boyutlu öznitelik setlerinin ve topluluk öğrenme yöntemlerinin, nöromüsküler hastalıkların erken tanısında invazif olmayan, güvenilir ve klinik olarak yorumlanabilir bir çözüm sunduğunu kanıtlamaktadır.

To cite this article: I. Calikusu and V. Kaya, "Analysis of EMG Signals Using Stacking Model and Deep-MLP in the Detection of Neuromuscular Disorders," *Artificial Intelligence Studies,* vol.8, no.2, pp. 158-169, 2025. doi:10.30855/ais.2025.08.02.06

1. Introduction

Neuromuscular diseases (NMDs) are disorders resulting from persistent and functional impairments in muscle fibers, motor neurons, and the neuromuscular junction. These diseases cause significant changes in motor unit activation patterns. Myopathic and neuropathic disorders exhibit similar clinical symptoms such as muscle weakness and fatigue, which makes their surface EMG (sEMG) diagnosis difficult [1, 2]. In recent years, due to the non-invasive nature of SEMG and its direct correlation with muscle activity, it has been increasingly used in the computer-assisted diagnosis of neuromuscular diseases [3, 4].

The number of EMG-based automated classification studies has also increased significantly in the last 5 years. In these studies, machine learning (ML) and deep learning (DL) methods have been shown to offer high accuracy in this field [5]. Classification accuracy levels of 95-98% have been achieved with classical ML models such as Random Forest (RF), Support Vector Machines (SVM), and Multilayer Perceptrons (MLP) [6, 7]. On the other hand, wavelet-based features and ensemble approaches have brought these success rates closer to 99% [8, 9]. In more recent studies, accuracy performance up to 99% has been achieved in healthy-Myopathy-Neuropathy classification with convolutional and recurrent neural networks (CNN, LSTM, ConvLSTM) [10, 11].

Despite the high accuracy values reported in the studies mentioned above in the literature, some fundamental methodological limitations are noteworthy. A large part of deep learning studies are based on abstract features generated from wavelet coefficients or time-frequency representations. The physiological equivalent and clinical interpretability of these representative features are often limited [10, 12]. In addition, these approaches have disadvantages such as high computation time, the need for large datasets, and low model transparency [13].

A significant portion of classical ML-based studies focus on a limited number of time-domain features and do not systematically address feature redundancy and information overlap [5, 13]. It is emphasized that the heterogeneous nature of myopathic EMG signals, low-amplitude but irregular motor unit patterns, and inter-class overlaps make the distinction between myopathy and neuropathy still the most difficult classification problem in the literature [15, 16].

Due to the reasons presented in the literature above, a significant gap emerges. Many of the current studies only consider the energy and amplitude levels of the EMG signal, while not offering a holistic feature framework that encompasses temporal variability, statistical distribution, event-based measures, and signal complexity. Furthermore, despite the strong representational capabilities of deep learning methods, the limited interpretability of these approaches is inferior to the physiological explainability provided by handcrafted features. It appears that this area is not sufficiently utilized. Furthermore, the number of studies that comparatively and systematically evaluate the behavior of different learning paradigms on the same feature space is limited. These shortcomings restrict both the reliable analysis of model performance and the interpretability of the obtained results in a clinical context. This study aims to fill the above gap in the literature. A multidimensional feature set, including amplitude, energy, derivative-based variation, statistical distribution, nonlinear, and model-based features, was obtained from EMG signals using Matlab 2023 rb [17]. The extracted feature vector was analyzed by classification using different algorithms such as SVM (RBF), Random Forest, Deep-MLP, Stacking (SVM+RF+MLP), and Graph-Enhanced SVM [18]. In addition, the potential of using autoencoder-based representations together with hand-made features was also considered. In this respect, the study presents an interpretable, repeatable, and platform-independent EMG classification design that aims to eliminate the problem of high accuracy but low interpretability frequently observed in the literature. The entire process was carried out using Matlab 2023 rb. Transparency and reproducibility were ensured by using Python libraries [19, 20].

2. Material and Methods

2.1. Dataset

In this study, a publicly available and peer-reviewed dataset containing electromyographic recordings of healthy individuals and myopathic and neuropathic patients was used [21]. The dataset used includes raw invasive EMG (iEMG) signals obtained from the biceps brachii and deltoid muscles during isometric muscle contraction. The dataset was published on August 22, 2023, and is accessible with Version 1 and DOI: 10.17632/543xpjycj9.1.

The dataset was collected at the Department of Neurology, Mustapha Pacha University Hospital, Algeria, between October 2015 and April 2019, under the supervision of specialist clinicians and with interdisciplinary collaboration. The study process was carried out in collaboration with the Spoken Communication and Signal Processing Laboratory and the Instrumentation Laboratory within the Faculty of Electronics and Informatics at Houari Boumediene University of Science and Technology. Records from a total of 241 participants are divided into three classes:

(i) Healthy (Healthy control group),
(ii) Myopathy (Myopathy patients), and
(iii) Neuropathy (Neuropathy patients).

The healthy control group consisted of 50 volunteers (23 women, 27 men) aged 17–56 years (mean 27.9 ± 9.5) with no history of neuromuscular or musculoskeletal disease. The myopathy group included 98 patients (55 women, 43 men) aged 4–78 years (mean 41.6 ± 16.8), and included different myopathic subtypes such as myotonia and polymyositis. The neuropathy group consisted of 93 patients (53 women, 40 men) aged 2–81 years (mean 46.5 ± 16.3) and included neuropathic conditions such as anterior horn diseases, multiple sclerosis, and radicular damage. The dataset exhibited a heterogeneous structure in terms of disease severity and clinical findings.

EMG recordings were obtained using a MATRIX LIGHT 2-channel EMG EP Headbox (Micromed SPA) system with MYOLINE brand gold-tipped disposable concentric needle electrodes. SystemPlus Evolution 1.04 software was used in the data collection process. The recorded EMG signals were in the millivolt (mV) amplitude range and covered the frequency band of 16 Hz – 5000 Hz. The signals were sampled at a very high sampling frequency of 32,768 Hz; this allowed for the capture of temporal details of motor unit action potentials with high resolution.

The dataset's index structure consists of a hierarchical structure where recordings of the biceps brachii and deltoid muscles are stored in separate folders. Each muscle folder is further subdivided into subfolders corresponding to the Healthy, Myopathy, and Neuropathy classes. EMG signals are stored in .asc format, and the file naming scheme includes record number, patient ID, muscle group, and class information. For example, the expression “EMG004_02_LD_Hea” indicates that the fourth EMG recording of the second patient was taken from the left deltoid muscle and belongs to the healthy class.

This dataset is a powerful reference for the development and validation of machine learning and artificial intelligence-based neuromuscular disease diagnostic models because it allows for the systematic examination of electromyographic differences between healthy and pathological individuals and includes a large population encompassing different ages, genders, and disease severity levels.

2.2. Preprocessing and Segmentation

Since raw EMG signals contain electrical noise interference sources such as motion artifacts, mains interference, and electrode-skin impedance variation, these noises were removed using a band-pass filter between 15-400 Hz. After the filtering process, the signals were separated into windows using a sliding window approach for feature extraction. Segmentation ensured the accurate calculation of the statistical and spectral characteristics of the signal within each window. All these processes were performed in Matlab 2023 and Python using NumPy/SciPy-based signal functions.

2.3. Feature Extraction

The success of EMG-based classification studies depends significantly on accurately representing the signal and extracting distinctive features. Here, the features encompass a broad set of handcrafted characteristics that consider not only the amplitude and energy levels of the EMG signal but also different perspectives such as temporal variability, waveform dynamics, statistical and event-based metrics, and complexity/fractal structure. This allows for the simultaneous consideration of both quantitative (intensity and energy) and qualitative (disorder, complexity, and transition dynamics) aspects of muscle activation.

Temporal and energy-based features (RMS, MAV, AE, VAR, SD, MMAV/MMAV2, MSR, ASS, ASM, and Log Detector) reveal differences in activation levels between different movements and pathological conditions regarding the contraction intensity and energy distribution of the EMG signal. In addition, waveform and change-based features (WL, AAC, DAMV, DASDV, DVV, LogDAMV, and LogDASDV) capture differences between consecutive samples and rapid amplitude changes, while also helping to differentiate signals exhibiting different dynamic behaviors at similar energy levels. Event count-based features (ZC, SSC, FZC, and WA) provide strong information about frequency-related characteristics through the trend and sign changes of the signal.

Finally, statistical distribution and complexity-based features (KURT, SKEW, IQR, MAD, LogCOV, CARD, AR coefficients, TM, etc.) VO, MFL, LogTKEO, and MPR represent the irregularity and temporal dependence of the muscle signal distribution structure at a higher level. Fractal and energy operator-based criteria play a distinctive role in capturing sudden changes in muscle activation and pathological motor unit behaviours.

With all the features mentioned above, this study's multidimensional feature set considers the EMG signal not from a single perspective, but from a holistic viewpoint encompassing its various physiological and statistical dimensions. Thus, stronger and more stable differentiation is achieved, even in cases where there is high overlap between classes.

2.4. Feature Selection

In this study, a total of 43 different features were obtained from EMG signals in the time, frequency, and time-frequency domains. To optimize classification performance and reduce computational costs, the Mutual Information method was applied to this feature pool. As a result of the analysis, the 20 features with the highest-class discrimination (highest correlation) were identified, and the models were trained with this reduced feature set.

2.5. Classification Models

Five different AI-based classification models were used to categorize EMG signals into Healthy, Myopathy, and Neuropathy classes. The basic and hybrid algorithms used in the study are described below.

2.5.1. Support Vector Machines (SVM) and Graphics-Based Approaches

The SVM algorithm, which provides high success in cases where feature data cannot be linearly separated, was implemented with the Radial Basis Function (RBF) kernel in this study. In addition, Graph-Enhanced SVM method was also included in the comparative analyses in order to better model the structural relationships between data points [22].

2.5.2. Random Forest

The Random Forest algorithm, an ensemble learning method, was used because it produces stable results in high-variance feature data and is resistant to overfitting. The model was run by creating numerous decision trees and applying the majority voting principle.

2.5.3. Deep Multilayer Sensor

Unlike traditional machine learning methods, a Deep-MLP (Multilayer Perceptron) network with a deep architecture was designed to model complex and nonlinear relationships in the feature dataset. This model has been used to learn distinguishing patterns, especially in classification problems with high structural heterogeneity (e.g., differentiating between myopathy and neuropathy).

2.5.4. Stacking Models

A stacking architecture has been developed that combines heterogeneous learning paradigms to increase the generalization capacity of individual models. In this architecture, SVM, Random Forest, and Deep-MLP algorithms act as "base learners," while the outputs of these models are combined by a higher-level "meta-learner." This approach aims to create a more robust prediction mechanism in EMG-based disease classification by integrating the strengths of different algorithms.

2.6. Experimental Design

Classification experiments were conducted using two main scenarios:

1. Three-Class Classification: The aim was to simultaneously separate the Healthy, Myopathy, and Neuropathy classes using the entire dataset.

2. Binary Classification: To examine the distinguishing features of diseases in more detail, "One-and-All" and pairwise comparison (Healthy and Neuropathy, Healthy and Myopathy, Myopathy and Neuropathy) scenarios were applied.

2.7. Evaluation Criteria

The performance of the developed models was evaluated using Accuracy, Macro-F1 Score, and Confusion Matrix metrics. Specifically, the F1 score was used to measure the class-based sensitivity of the model in unbalanced data distributions, and complexity matrices were used to analyze inter-class confusion.

3. Results and Discussions

This study identified the most distinctive and defining features among 43 features obtained from surface EMG signals. These features included evaluating the performance of various classification methods and verifying the classification success through visual outputs.

3.1. Feature Selection Findings

In this study, 43 different statistical, frequency, and time features were extracted. A Mutual Information-based feature selection algorithm was applied to these features, and the 20 features with the highest information gain were selected. The majority of these features are based on frequency content (e.g., mean frequency, median frequency) and amplitude-based statistics. The relationships between the selected features were examined through the Correlation Matrix of the Selected 20 Features presented in Figure 1.

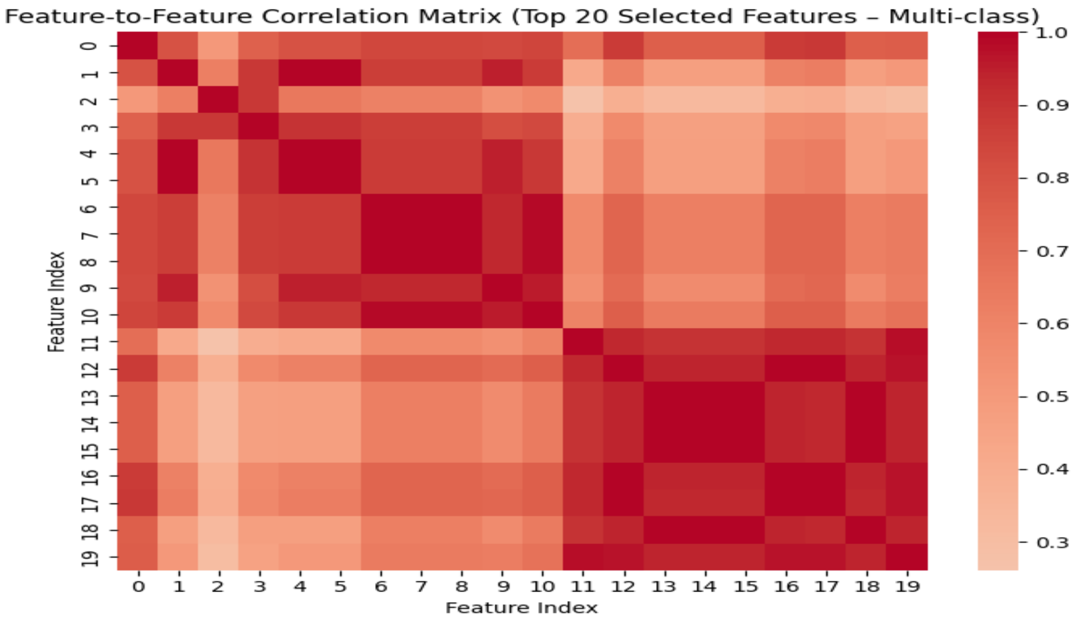

Figure 1. Correlation Matrix of the Selected 20 Features

Figure 1 shows the feature matrix of the 20 features with the best correlation among the top 20 EMG features selected using the mutual information method. The feature matrix in the figure shows that amplitude and energy-based features (such as RMS, MAV, MMAV, IEMG, DASDV, SSI, AAC) show high correlation with each other. These features fundamentally reflect the amplitude distribution of the signal, the intensity of muscle activation, and the total energy level. Therefore, they carry similar physiological information and naturally group together in a cluster. In contrast, non-linear energy operators (MSR, ASS, ASM) and statistical dispersion features (Skewness, Variance, Standard Deviation, Interquartile Range, MAD) exhibit lower correlations compared to amplitude-based features. This suggests that these features provide additional and independent information by capturing the irregularity, statistical shape, and complexity dimensions of the EMG signal. Frequency/dynamic-based features, such as Zero Crossing and AAC, are also more isolated compared to other clusters. Overall, the correlation matrix clearly shows that the selected 20 features present a balanced structure with both common and complementary information, thus enhancing the discriminative power of the classification models.

3.2 Three-Class Classification Results (Healthy – Myopathy – Neuropathy)

In this study, a triple classification (Healthy, myopathy, and neuropathy) was performed using 20 features with the highest correlation, selected from 43 different EMG features obtained from EMG signals using the Mutual Information method. The results, which were subjected to five different artificial intelligence classification processes and had accuracy rates ranging from 76.88% to 80.26%, are presented in Table 1. The highest accuracy performance (80.26%) and F1 score (75.81%) were obtained using the Stacking model, where the outputs of SVM, Random Forest, and Deep-MLP models were combined by a high-level meta-learner. This performance increase demonstrates that combining multiple heterogeneous learning paradigms provides a stronger generalization capacity in EMG-based disease classification.

Table 1. Three Class Classification Results (Healthy–Myopathy–Neuropathy)

Method	Accuracy (%)	Macro-F1 (%)
SVM (RBF)	77.92	72.15
Random Forest	78.70	74.19
Deep-MLP	76.88	72.51
Stacking (SVM+RF+MLP)	80.26	75.89
Graph-Enhanced SVM	78.70	73.27

Figure 2 shows the confusion matrix of the three-class model. Examination of Figure 2 reveals that the Healthy class exhibits high sensitivity (93%), while the Neuropathy class similarly provides high accuracy (82–83%). The sensitivity in the Myopathy class is significantly lower, nearly around 50%. This result confirms that myopathic EMG signals have high variance and that some samples exhibit similar characteristics to normal or neuropathic structures, making differentiation more difficult.

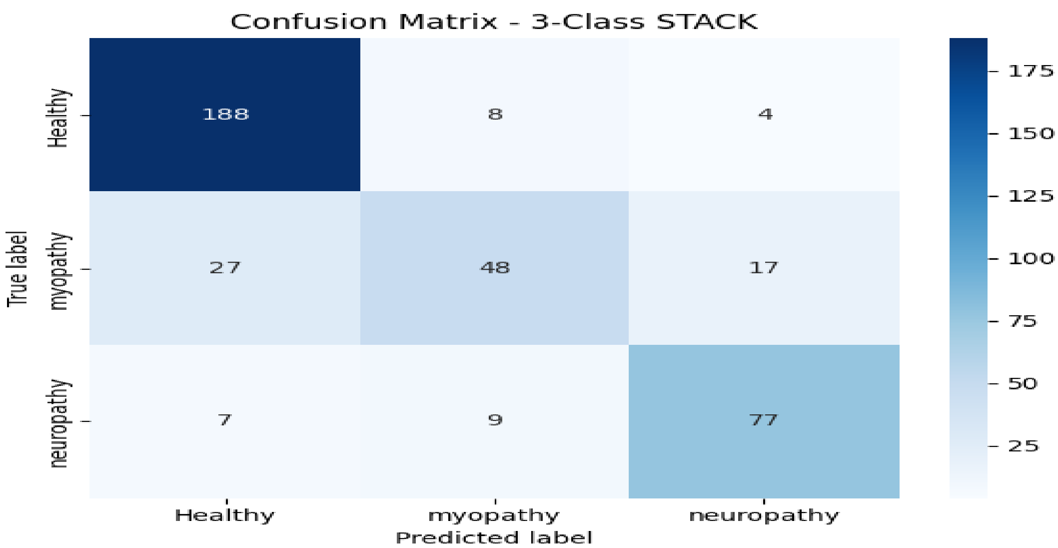

Figure 2. Three-Class Confusion Matrix of the Stacking Method.

According to the results in Figure 2, Myopathy appears to be the most challenging class in the three-class classification problem. Furthermore, the Stacking method demonstrates a significant contribution to classification performance by providing over 80% accuracy.

3.3. Binary Classification Results (Healthy and Others, Myopathy and Others, Neuropathy and Others)

In this study, the 20 most informative feature models for each binary classification task were trained using this feature set.

3.3.1 Healthy and Others

This study first attempted to determine the extent to which healthy individuals differed from other myopathy and neuropathy patients in a pairwise analysis. It was observed that the "Healthy" class exhibited more regular, stable, and low-noise EMG patterns compared to both the Myopathy and Neuropathy disease classes. This structural integrity of the EMG signals from healthy individuals resulted in more consistent and distinct amplitude, spectral, and frequency-based features. Thus, the differentiation of the "Healthy" class from the myopathy and neuropathy groups was achieved with higher accuracy. For the classification process, the 20 most informative features were re-selecting each time using the mutual information method, and the models were trained on this feature dataset. The obtained accuracy of 87.27% and Macro-F1 score of 87.27% demonstrate an improvement in the classification performance of EMG signals from healthy individuals.

Table 2. Classification of healthy and other groups.

Metric	**Value**
Accuracy	87.27%
Macro-F1	87.27%

3.3.2 Myopathy and Others

Table 3 presents the results of a classification study conducted to demonstrate how distinct the myopathy group is from the other two groups. The accuracy of this classification was 83.64%, and the Macro-F1 value was 75.91%. The high heterogeneity within the myopathy group's internal structure made this classification difficult. The frequent occurrence of short-duration motor unit potentials, lower amplitude levels, and increasing irregularity of the signal over time in myopathic EMG signals blurred the lines between healthy and neuropathic samples. Therefore, the classification success of the myopathy class compared to other groups is low.

Table 3. Classification of myopathy and other groups

Metric	**Value**
Accuracy	83.64%
Macro-F1	75.91%

3.3.3 Neuropathy and Others

The accuracy and Macro-F1 values obtained from the classification of the neuropathy class with the other two classes are presented in Table 4. As can be seen from this table, neuropathic samples can be distinguished quite well from the other groups. The classification resulted in an accuracy of 88.05% and a Macro-F1 of 84.04%. The slowing of nerve conduction velocity, irregularities in motor unit activation signal patterns, and shifts in frequency characteristics observed in neuropathic EMG signals have enabled this class to be distinguished from the other two groups with a distinct pattern.

Table 4. Classification of neuropathic and other groups

Metric	**Value**
Accuracy	88.05%
Macro-F1	84.04%

Within these three classes, the results clearly show that neuropathy is the easiest to differentiate, while myopathy is the most difficult. The more consistent and discriminatory characteristics of neuropathic signals improved the classification performance. On the other hand, the high variance and irregular patterns in myopathic signals were found to be the main factors limiting model accuracy.

3.4. Pairwise Comparisons with Artificial Intelligence

This study also examines, from a different perspective, how binary classification changes not only with classical classification methods but also with different new artificial intelligence models. In this study, the discriminative power of pairwise comparisons shows significant differences, especially in the Healthy, Myopathy, and Neuropathy groups, where signal variation is evident according to class characteristics. The artificial intelligence-based results obtained for each class pair are presented in detail below.

3.4.1. Healthy and Neuropathy

The highest accuracy in the classification between healthy and neuropathy classes was obtained with the Random Forest and Deep-MLP models, at 93.17%. The models were able to easily make this distinction because the signal frequency content, motor unit synchronization, and action potential pattern are quite different in the healthy and neuropathy groups. The results obtained for this comparison are presented in Table 5.

Table 5. Healthy and Neuropathy: A Dual Comparison

Method	Accuracy	Macro-F1
Random Forest	93.17%	92.03%
Deep-MLP	93.17%	92.12%
Stacking	92.17%	92.03%
Graph+SVM	92.49%	91.23%
Advanced-SVM	92.15%	90.97%

3.4.2. Healthy and Myopathy

The highest accuracy (86.99%) was achieved with the Deep MLP technique in the pairwise comparison classification of Healthy and Myopathy. The fact that EMG signals in Myopathy exhibit lower amplitude, shorter duration of motor unit activity, and more irregular waveform compared to healthy signals allowed the model to more easily capture distinctive patterns.

Table 6. Healthy and Myopathy: A Dual Comparison

Method	Accuracy	Macro-F1
RF	85.62%	82.01%
Deep-MLP	86.99%	84.52%
Stacking	85.62%	82.34%
Graph + SVM	85.96%	82.42%
Advanced-SVM	84.25%	81.05%

The distribution of classification success achieved with Deep-ML is visualized in the confusion matrix in Figure 3. Here, it is seen that the Healthy and Myopathy signals can be significantly separated; the model predicts the Healthy group with high accuracy. The high classification success rate in the Myopathy group, despite structural heterogeneity, demonstrates that Deep-MLP can successfully represent complex EMG patterns.

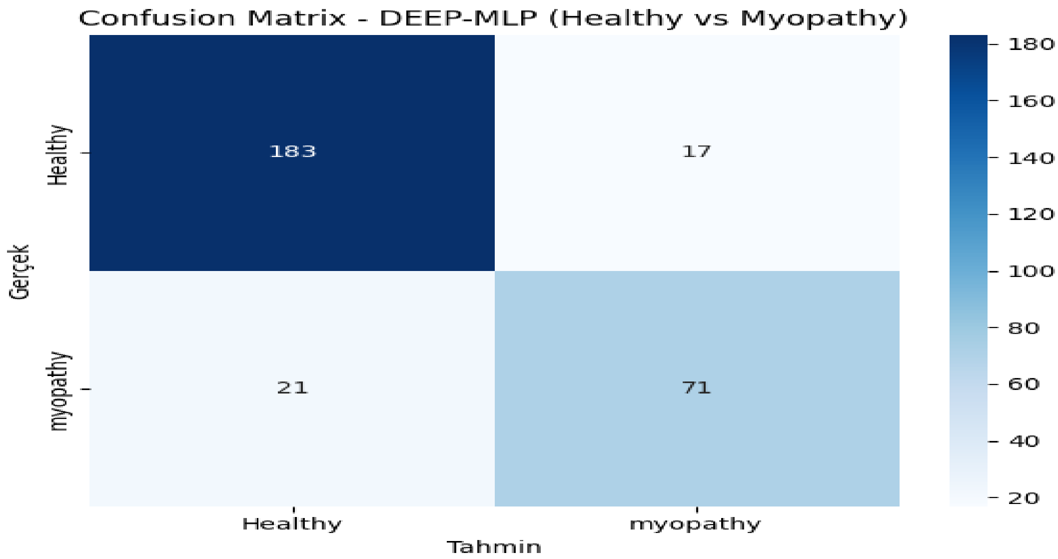

Figure 3. Deep MLP confusion matrix.

3.4.3. Myopathy and Neuropathy

Unlike the differentiation of healthy groups, the differentiation between myopathy and neuropathy is difficult because they share common EMG features. Both neuropathic and myopathic signals exhibit disruption, amplitude reduction, conduction anomalies, and temporal dysregulation. These shared pathological features disrupt the boundaries between classes, reducing classification success.

The results of the classification performance between myopathy and neuropathy are given in Table 7. As can be seen from the table, the Deep-MLP model showed the highest classification success with approximately 89–91% accuracy and 88–90% Macro-F1. Random Forest showed the second best performance with an accuracy level of 88%. Stacking and Graph+SVM methods exhibited classification success rates of approximately 87–89% and 86%, respectively. These findings demonstrate that deep learning-based methods offer a stronger representation capacity in differentiating between the two pathological signals.

Table 7. Myopathy and Neuropathy dual classification success.

Method	**Accuracy**	**Macro-F1**
Deep-MLP	≈89–91%	≈88–90%
Random Forest	≈88%	≈87%
Stacking	≈87–89%	≈86–88%
Graph + SVM	≈86%	≈85%

The classification error distribution and the model's actual-prediction performance are presented in the confusion matrix in Figure 4. Figure 4 shows that the model correctly predicted a significant portion of the Myopathy samples and achieved an even higher classification success rate in the Neuropathy class. The generally low margin of error confirms that the feature set for distinguishing both pathological groups contains sufficient information.

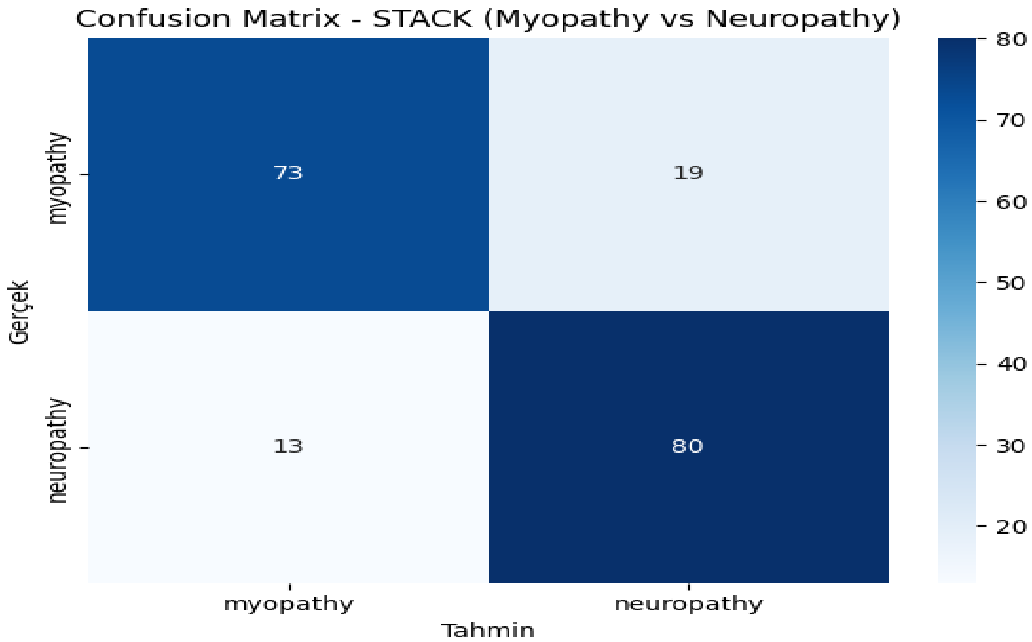

Figure 4. Myopathy–Neuropathy confusion matrix obtained using the stacking model.

3.5. General Assessment

This study shows that multidimensional time, frequency, and non-linear features selected from 43 different characteristics obtained from superficial EMG signals have very high classification accuracy in both triple and binary classifications. According to the triple classification results, the Stacking model was the most successful method with 80.26% accuracy. When examining the distinction between disease groups, it was determined that the Neuropathy class was the easiest to differentiate, while the Myopathy class was the most difficult. This result suggests that neuropathic EMG patterns contain more pronounced spectral and morphological differences, while myopathic signals have a more heterogeneous and irregular structure.

The binary classification results also revealed the diagnostic potential of EMG; particularly in the Healthy-Neuropathy distinction, a very high performance was achieved with 93.17% accuracy. All these findings clearly demonstrate that superficial EMG is a reliable biomarker in automated diagnostic processes for neuromuscular diseases, and that non-linear features make a critical contribution to model performance.

3.6. Discussion

In this study, 43 different features consisting of 43 different time-frequency and nonlinear features were obtained from feature extraction from sEMG signals. These features were then analyzed using the Mutual Information method, selecting the 20 most distinctive features to achieve success in classifying myopathy, neuropathy, and healthy individuals. The results show that multiple feature sets and hybrid learning architectures provide superior performance in understanding the complex and stochastic nature of EMG signals compared to single-approach methods. In particular, in the three-class (Health-Myopathy-Neuropathy) classification problem, the Stacking hybrid model achieved an accuracy rate of 80.26%, surpassing classical methods such as SVM and RF. This demonstrates that combining different learning algorithms (ensemble learning) increases generalization capability [6, 7, 23].

When class-based classification performances are examined pathophysiologically, a significant difference is observed between the high discriminability of the Neuropathy class and the classification difficulty of the Myopathy class. The reason for this is that in neuropathic processes, the loss of synchronization in motor unit firing rates and the slowing of nerve conduction lead to significant spectral shifts and irregularities in EMG signals, allowing the model to easily separate this class from the healthy group [1, 15]. However, the low sensitivity (around 50% sensitivity) of the Myopathy class may be due to the low signals resulting from atrophy in myopathic muscle fibers, showing similarity to the noise background of healthy signals or mild neuropathic findings [4, 16]. However, unlike related studies in the literature [11, 14], the Deep-MLP architecture used in this study achieved a success rate of up to 91% in the difficult-to-categorize Myopathy and Neuropathy distinction. This shows that the distinction is easier when classified with hand-made features in deep learning structures.

From another perspective, methodologically, this study is more transparent and explainable compared to the "end-to-end deep learning" approaches that are becoming widespread in the literature [10, 12]. Deep learning techniques create uncertainty in clinical assessment due to a black box problem unknown in the literature. In contrast, knowing the physiological counterparts of features such as entropy, wavelet energy, and statistical moments used in this study allows clinicians to easily interpret them [17]. The hybrid feature and modelling approach used in this study provides high accuracy, especially in the detection of early-stage neuromuscular diseases, while showing that the inclusion of nonlinear features in the model is critical for the diagnosis of heterogeneous groups such as myopia. In future studies, it is expected that the patient population will be expanded and spatiotemporal features will be integrated into the model to improve the performance in the myopathy class.

4. Conclusions

This study demonstrates that processing of time, frequency, and especially nonlinear features extracted from superficial EMG signals with hybrid artificial intelligence architectures provides high performance and clinical interpretability in the differential diagnosis of neuromuscular diseases. The Stacking model achieved an overall accuracy of 80.26%, trained on a set of 20 features optimized using the Mutual Information method, in three-class differentiation, and the accuracy reaching 91% of the Deep-MLP model in differentiating Myopathy from Neuropathy, considered the most challenging problem in the literature, proves the effectiveness of the proposed method in analyzing complex and heterogeneous pathological patterns. The findings show that this approach, which combines the representational power of deep learning with the transparency of hand-made features, can be integrated into clinical workflows as a non-invasive, low-cost, and reliable computer-aided diagnostic system and can make significant contributions to the early diagnosis processes of neuromuscular diseases.

Conflict of Interest Statement

The authors report no conflicts of interest.

References

[1] A. Morozov, T. Sorokovikova, S. Naumova, and A. Kryukova, "Electroneuromyography as one of the methods of diagnostics of pathologies of different genesis (literature review)," Vrach, vol. 35, no. 4, pp. 23-28, 2024.
[2] W. Schulte-Mattler and C. Bischoff, "EMG und Neurografie–ein Update," Klinische Neurophysiologie, vol. 53, no. 02, pp. 107-119, 2022.
[3] C. Quintao et al., "Surface electromyography for testing motor dysfunction in amyotrophic lateral sclerosis," Neurophysiologie Clinique, vol. 51, no. 5, pp. 454-465, 2021.
[4] K. Hokkoku et al., "Frequency of EMG abnormalities in idiopathic inflammatory myopathies under the EULAR/ACR classification criteria," Medicine, vol. 103, no. 4, p. e37105, 2024.
[5] Dörterler, S., Dumlu, H., Özdemir, D., & Temurtaş, H. (2022). Melezlenmiş K-means ve Diferansiyel Gelişim Algoritmaları ile Kalp Hastalığının Teşhisi. In International Conference on Engineering and Applied Natural Sciences içinde (pp. 1840-1844). Konya.
[6] I. Yoo et al., "Deep learning-based algorithm versus physician judgement for diagnosis of myopathy and neuropathy from needle electromyography," medRxiv, p. 2023.01. 13.23284511, 2023.
[7] A. Achmamad, M. Elfezazi, A. Chehri, I. Ahmed, A. Jbari, and R. Saadane, "ML-based identification of neuromuscular disorder using EMG signals for emotional health application," ACM Transactions on Internet Technology, 2023.
[8] A. Subasi and S. M. Qaisar, "Surface EMG signal classification using TQWT, Bagging and Boosting for hand movement recognition," Journal of Ambient Intelligence and Humanized Computing, vol. 13, no. 7, pp. 3539-3554, 2022.
[9] E. Cepeda, N. N. Sánchez-Pozo, and L. M. Chamorro-Hernández, "Enhanced Predictive Modeling for Neuromuscular Disease Classification: A Comparative Assessment Using Gaussian Copula Denoising on Electromyo-graphic Data."
[10] M. Zaher, A. S. Ghoneim, L. Abdelhamid, and A. Atia, "Unlocking the potential of RNN and CNN models for accurate rehabilitation exercise classification on multi-datasets," Multimedia Tools and Applications, vol. 84, no. 3, pp. 1261-1301, 2025.
[11] E. Tuncer and E. Bolat, "LSTM-based approach for Classification of Myopathy and Normal Electromyogram (EMG) Data," Balkan Journal of Electrical and Computer Engineering, vol. 11, no. 3, 2023.
[12] S. Ouamri, M. Abed, and A. O. Ali, "Classification of intramuscular EMG signals using a separable compact support kernel-based timefrequency distribution and deep transfer learning," in 2023 International Symposium on Signals, Circuits and Systems (ISSCS), 2023: IEEE, pp. 1-4.

[13] A. Pavlidou, X. Liang, N. G. Arekhloo, H. Li, J. Marquetand, and H. Heidari, "Spontaneous muscle activity classification with delay-based reservoir computing," APL Machine Learning, vol. 1, no. 4, 2023.
[14] S. V. Patel, G. K. Kumar, I. Ali, G. Jyothi, K. Banumathi, and B. Shanthakumar, "Time-Domain Statistical Analysis of EMG Signals for Muscular Dysfunction Detection," Asian Journal of Electrical Sciences, vol. 13, no. 2, pp. 1-18, 2024.
[15] A. R. Verma and B. Gupta, "Detecting neuromuscular disorders using EMG signals based on TQWT features," Augmented Human Research, vol. 5, no. 1, p. 8, 2020.
[16] T. Kamali and D. W. Stashuk, "Transparent electrophysiological muscle classification from EMG signals using fuzzy-based multiple instance learning," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 4, pp. 842-849, 2020.
[17] S. Roldan-Vasco, A. Orozco-Duque, and J. R. Orozco-Arroyave, "Swallowing disorders analysis using surface EMG biomarkers and classification models," Digital Signal Processing, vol. 133, p. 103815, 2023.
[18] M. Emimal, W. J. Hans, T. Inbamalar, and N. M. Lindsay, "Classification of EMG signals with CNN features and voting ensemble classifier," Computer Methods in Biomechanics and Biomedical Engineering, vol. 28, no. 7, pp. 1042-1056, 2025.
[19] H. Lee, M. Jiang, J. Yang, Z. Yang, and Q. Zhao, "Decoding gestures in electromyography: spatiotemporal Graph neural networks for generalizable and interpretable classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024.
[20] Y. Wei, Y. Chen, and C. Ye, "Neuromuscular disease auxiliary diagnosis using a portable magnetomyographic system," Physiological Measurement, vol. 45, no. 9, p. 095001, 2024.
[21] A. MESSEKHER, S. MEKAOUI, M. KEDIR-TALHA, M. I. KEDIHA, and F. MOSTEFAOUI, "Electromyographic data recordings during isometric contractions of the biceps brachii and deltoid collected from patients and healthy subjects," Mendeley Data, vol. 1, p. 2023, 2023.
[22] S. Liang, J. Shao, J. Zhang, and B. Cui, "Graph-based non-sampling for knowledge graph enhanced recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 9, pp. 9462-9475, 2023.
[23] Dörterler, S., Dumlu, H., Özdemir, D., & Temurtaş, H. (2024). Hybridization of meta-heuristic algorithms with k-means for clustering analysis: Case of medical datasets. Gazi Mühendislik Bilimleri Dergisi, 10(1), 1-11.