ORIGINAL PAPER

JMS Journal of Medical Science

Deep neural networks to classify motor unit action potential signals acquired using needle electromyography

Isil Tatlidil
Karadeniz Technical University Faculty of Medicine, Department of Clinical Neurophysiology, Trabzon, Turkiye
https://orcid.org/0000-0002-9666-3647
Corresponding author: tatlidilisil173@hotmail.com

Murat Ekinci
Karadeniz Technical University Engineering Faculty, Department of Computer Engineering, Trabzon, Turkiye
https://orcid.org/0000-0001-9326-8425

Cavit Boz
Karadeniz Technical University Faculty of Medicine Department of Clinical Neurophysiology, Trabzon, Turkiye
https://orcid.org/0000-0003-0956-3304

https://doi.org/10.20883/medical.e1391

Keywords: deep learning, convolutional neural network, gate recurrent unit, needle electromyography, neuromuscular disorders

Received 2025-08-15
Accepted 2025-10-10
Published 2025-12-29

How to Cite: Tatlidil I, Ekinci M, Boz C. Deep neural networks to classify motor unit action potential signals acquired using needle electromyography. Journal of Medical Science. 2025 December;94(4);e1391. doi:10.20883/medical.e1391

ABSTRACT

Aim. To classify motor unit action potential patterns using a deep learning technique with high accuracy.
Material and methods. A dataset was compiled from three main groups of motor unit action potential patterns, including myopathy, neuropathy, and normal, as assessed by a clinical neurophysiologist during routine clinical assessments. After preprocessing the raw signals in the dataset, a total of 3,152 signal segments from 96 muscles of 26 individuals were divided into training and test sets. Deep learning network models were developed in Python using the Keras API in Jupyter Notebook.
Results. Among the deep learning models, a hybrid deep neural network model with a one-dimensional convolution layer as an input layer and four layers of gate recurrent units (1DCNN-GRU) achieved the highest accuracy rates. Ten-fold cross-validation resulted in a mean accuracy rate of 98.13 ± 1.05%.
Conclusions. Both conventional machine learning models and deep learning models could classify needle EMG patterns that belonged to three neuromuscular disorder groups with high accuracy. However, more clinical studies with larger datasets are needed for validation. In contrast to conventional machine learning techniques, deep learning models could receive signals as input data and automatically extract the required features. Therefore, they could facilitate the real-time implementation of the pattern recognition tasks in the future.

Introduction

Assessing audio-visually, needle electromyography (EMG) is a crucial diagnostic tool used to differentiate between the three main types of neuromuscular conditions: healthy, neuropathic, or myopathic [1]. During needle EMG procedures, potentials originating from needle insertion into a muscle, muscle cell membrane potentials while resting, and motor unit action potentials during

muscle contraction are recorded [2,3]. Although there are many experienced electromyographers or neurophysiologists, needle EMG examinations are still highly subjective, and, in some cases, reaching a consensus for certain kinds of patterns is unlikely [1,4]. Machine learning methods, intensive learning, have the potential to enhance standardisation in pattern recognition and may offer valuable diagnostic support to less experienced electromyographers in the future [5].

Machine learning methods have been extensively applied in recent clinical research to address challenges related to classification, decision-making, and prognosis prediction within clinical practice. [5,6]. As a subtype of artificial intelligence, machine learning has yielded promising results in biomedical image and signal processing studies, paving the way for the development of clinical decision support systems [7]. Various machine learning techniques are available for use in medical studies, which can be grouped into classic or conventional machine learning and deep learning techniques based on the data processing steps [5,8]. Traditional machine learning techniques share standard pipelines for processing image or time series data. After preprocessing data, quantitative analysis is applied to extract features. The best meaningful features for classification are then required to be chosen as the input data for a machine learning tool, such as a decision tree, support vector machine, logistic regression, or a small-scale artificial neural network, such as a multi-layer perceptron [6,9]. Although conventional machine learning techniques have hand-crafted feature extraction and selection steps in their data processing pipeline, deep learning techniques such as convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, long-short term memories (LSTM), and gate recurrent units (GRU) can automatically achieve these processing steps. Therefore, the image or signal presented in a time-series format can be provided as input data to the classification pipeline of an artificial neural network [1, 5, 8]. Direct use of signal data as input may streamline the automation process and enhance decision-making capabilities, enabling real-time implementation.

Despite several conventional machine learning studies showing promising results for classifying needle EMG signals, studies that implement deep learning for needle EMG signal classification are scarce [1,8]. Few studies have focused on classifying resting state membrane potentials [1]. Nodera et al. [10] classified resting-state potentials with 86% accuracy using Mel spectrograms as input data, and pre-trained models of convolutional neural networks for image recognition were employed. The accuracy reached up to 94% in the 19-layered Visual Geometry Group (VGG-19), a convolutional neural network model with 19 weighted layers, by using data augmentation techniques. Nam et al. [11] classified resting state potentials using Inception-v4 as a convolutional neural network, yielding results with 93% accuracy. The input data consisted of image files with .png extensions representing the resting state potentials. There are also a few studies with encouraging results that aim to classify motor unit action potentials during muscle contraction using deep learning techniques [1]. Sengur et al. [12] used a continuous wavelet transform (CWT) spectrogram and Pseudowigner-Wille distribution function for preprocessing. Two-dimensional spectrograms were used as input data for the two-dimensional convolutional neural network (2D-CNN) pipeline, which achieved 96.8% accuracy in classifying two groups: Amyotrophic Lateral Sclerosis (ALS) and normal. Yoo et al. [8] utilised nEMGnet, a one-dimensional residual convolutional neural network inspired by the Residual Neural Network (ResNet) and the Visual Geometry Group (VGG) neural network, for the classification of EMG segments, achieving 62.35 ± 4.60% accuracy in classifying normal, neurogenic, and myogenic segments. Then, the divide-and-vote algorithm was used to determine the disease labelling for each muscle and patient with 83% accuracy.

An RNN is a specialised form of deep neural network particularly well-suited for handling sequential data, such as time series [13]. They are mainly used to predict the following likely data. However, classifying signals or time series could also be possible by arranging the architecture [14]. LSTM and GRU are subtypes of RNNs that yield successful results in surface EMG studies [15, 16]. Our study aims to classify nEMG data with motor unit action potentials into three categories, standard, neuropathy, and myopathy, using deep learning techniques with high accuracy.

Material and methods

Data acquisition

We included EMG traces acquired using concentric needle EMG recordings on a Natus Neurology Nicolet EMG system with an AT2+6 amplifier in this study. We retrospectively reviewed the recorded needle EMG signals obtained in routine clinical practice from December 2022 to June 2023, together with the EMG reports of clinical neurophysiologists. We received a total of 155 EMG signal traces corresponding to 134 muscles in 33 patients. We applied labelling at the patient and muscle levels according to the EMG reports in the archive. We excluded cranio-bulbar muscles due to the shorter duration and smaller amplitude of motor unit action potentials compared with limb muscles [3,17]. Among 33 subjects, 15 were normal, 11 had pathologies causing neurogenic motor unit action potentials, and seven had myopathies with myogenic action potentials. Among the 15 normal subjects, we excluded one subject due to an increased percentage of polyphasic motor unit action potential findings and another two subjects because the study was performed only on cranio-bulbar muscles. In the neurogenic group, we excluded one patient because the study focused only on cranio-bulbar muscles, and two patients due to being in the acute to subacute phase of neuropathy, as no changes in motor unit action potential morphology related to reinnervation were observed. We also excluded muscles that did not represent labelled patterns.

Finally, we included EMG traces of 29 muscles from seven individuals with myopathy, 24 muscles from eight individuals with neuropathy, and 43 muscles from 11 normal individuals in the study. Detailed information about the subjects is presented in **Table 1A–1E**. The recording parameters included a 30-50 Hz high-pass filter, a 10 kHz low-pass filter, and a sampling rate of 48,000 per second. We extracted EMG data from the Synerqy software system in .txt format. We reformatted the .txt files for use with a program developed in WinForms Visual Studio C++ to review traces in 100-msec sweep times. We also used this pro-

Table 1A. Demographics of the cases according to records.

Demographics	Female	Male	Total	Age (Mean/Median/Range)
Myogenic	2	5	7	52.8/56/21-80
Neurogenic	2	6	8	52.6/52.5/38-72
Normal	5	6	11	45.4/54/12-76
Total	9	17	26	49.6/53.5/12-80

Table 1B. Sampled muscle records from the upper limb and shoulder

Muscles sampled	PM	Is	Ss	TB	D	BB	EDC	FDS	FCU	FDI	ADM	APB
Myogenic	4	-	2	-	4	5	4	-	-	1	-	-
Neurogenic	-	-	1	-	2	-	-	-	1	1	1	2
Normal	1	1	1	1	6	5	2	1	1	5	-	1
Total	5	1	4	1	12	10	6	1	2	7	1	3

PM = Pectoralis major, Ss = Supraspinatus, Is = Infraspinatus, Triceps Brachi = TB, D = Deltoid, BB = biceps brachi, EDC = Extensor digitorum Communis, FDS = Flexor Digitorum Superficialis, FCU = Flexor Carpi Ulnaris, FDI = First Dorsal interosseus, ADM = Abductor Digiti Minimi, APB = Abductor Pollicis Brevis

Table 1C. Sampled muscle records from the lower limb and hip.

Muscles sampled	Ip	GlMax	GlMed	TFL	RF	VM	AM	TA	GcM	TP	EHL
Myogenic	7	-	-	-	1	1	-	-	-	-	-
Neurogenic	-	2	2	1	-	1	1	5	2	1	1
Normal	1	-	-	-	1	4	-	5	6	1	-
Total	8	2	2	1	2	6	1	10	8	2	1

Ip = Iliopsoas, RF = Rectus Femoris, VM = Vastus Medialis, AM = Adductor Magnus, TA = Tibialis Anterior, GcM = Gastrocnemius Medialis, TP = Tibialis Posterior, EHL = Extensor Hallucis Longus, TFL = Tensor Fascia Lata, GlMax = Gluteus Maximus, GlMed = Gluteus Medius

Table 1D. Number of neurogenic conditions diagnosed.

Neurogenic Pathologies	n
Chronic L5-S1 radiculopathy	1
Subacute-Chronic L4-L5 radiculopathy	1
Chronic L4-L5 radiculopathy	1
Chronic L3-L4, L4-L5 radiculopathy	1
Chronic C5-C6 radiculopathy	2
Chronic median nerve injury	1
Chronic lower truncus injury	1

Table 1E. Number of myogenic conditions diagnosed.

Myogenic Pathologies	n
Myopathy alone	5
Myopathy with fibrillations	1
Myopathy with myotonia	1

gram for preprocessing and labelling. The study received approval from the ethics committee of Karadeniz Technical University.

Preprocessing

We down-sampled raw needle EMG signals to 9600 Hz to reduce computational complexity. For the down-sampling procedure, we used an algorithm that calculates the second derivative of each signal. The signal with the lowest second derivative value was chosen as the representative signal for each of the five consecutive signals in the trace. As a result, the local minimums and maximums of high-frequency patterns – containing details regarding small motor unit action potentials, turns, and phases, which are essential characteristics of myopathy – were retained during the down-sampling process (see **Figure 1**). We also filtered the signal traces using a 50-Hz high-pass filter to remove redundant low-frequency information digitally. Each nEMG data sample recorded from the labelled muscle has traces that are 4 to 20 seconds long. Two neurophysiologists (IT and CB) reviewed all nEMG traces and selected segments for labelling that were artefact-free and at least 0.5 seconds long. We sliced the extracted traces into 100-msec segments with 50-msec overlap. Following processing, the neurophysiologists conducted a thorough review of each signal segment. They removed any traces from the dataset that did not contain characteristic features of the designated neuromuscular disorder group. We also excluded the traces with far-field potentials. Most of the signal segments had 945-960 consecutive signal series. We determined the standard signal vector length to be 940 to ensure a consistent size for each signal segment. Eight signal segments, three of them belonging to the normal group, and five belonging to the neurogenic group, required zero padding at the end as a result of being shorter than 940. Finally, we selected 1221 normal, 718 neurogenic, and 1212 myogenic EMG signal segments as the training and test sets. Thus, we generated a dataset consisting of 3,152 signal rows with 940 columns.

Deep Learning Network Model and Training

We performed deep learning using Python 3.7.16 in Jupyter Notebook. We designed deep neural networks using LSTM, GRU, and CNN in various architectures with the Keras library and a TensorFlow 2.6.0 backend in Python. We utilised the classification model to categorise EMG signal segments into three groups: neurogenic, myogenic, and normal. Therefore, the output layer consisted of three neurons, representing three neuromuscular disease groups according to the one-against-all principle. Preprocessed EMG signal segments, recorded as a CSV table, were provided as input data. Initially, we used 80% of labelled signal segments as the training set and the remainder as the test set. Then, avoiding overfitting, we also applied ten-fold cross-valida-

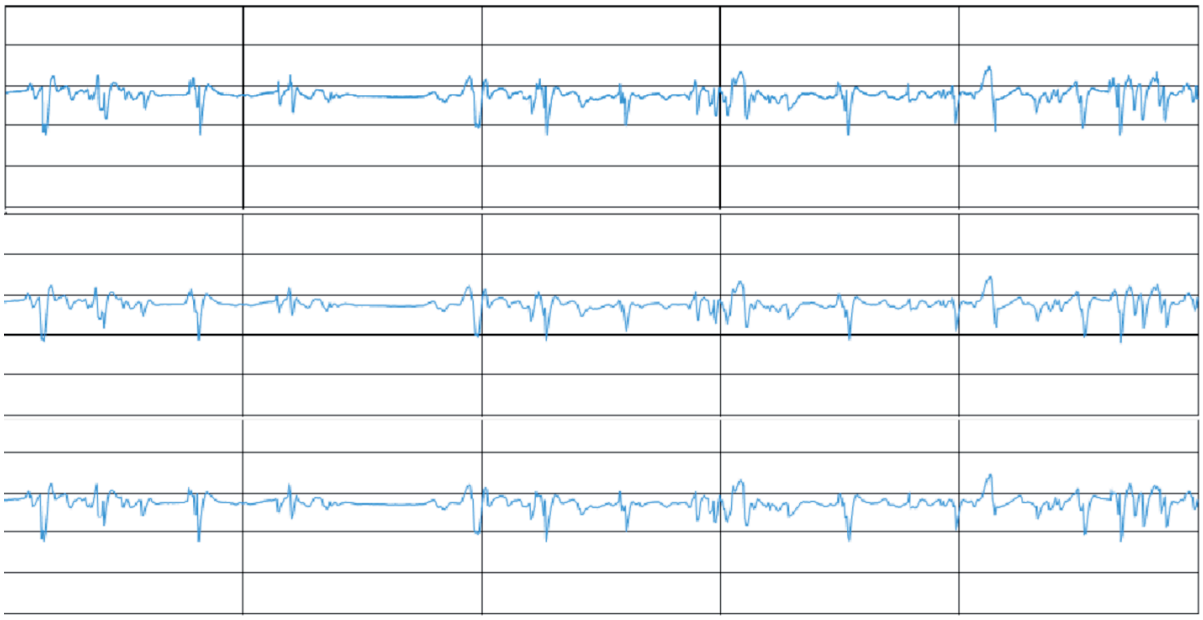

Figure 1. We down-sampled the signal segments to 9600 Hz to reduce computational complexity. However, during the down-sampling procedure, high-frequency information could be missed, which is particularly relevant for the recognition of myopathy forms. Each trace segment contains 100-msec-long signals. The raw trace, the first trace, contains high-frequency details that may be overlooked by traditional down-sampling methods, such as average pooling, which is seen in the second trace. By computing the second derivative at each signal point, peaks can be identified and selected as representative points during downsampling, as shown in the third trace. This approach enables greater preservation of high-frequency information compared with conventional techniques.

tion with randomly chosen training and test sets to assess the performance of the models. We performed the training and testing procedure on a CUDA-enabled NVIDIA GPU (GeForce RTX 3060) running on the Windows 11 operating system.

Statistics

We performed the statistics using the R Studio software package, version 4.4.1. For continuous variables, we conducted the Shapiro-Wilk test to check the normality of data distribution. We used Student's t-test for normally distributed continuous variables and the Mann-Whitney U test for non-normally distributed continuous variables to evaluate the differences between the groups. Statistical significance was considered significant if $p < 0.05$.

Results

We tested various deep learning architectures and optimised the best-performing models using the grid search algorithm. Finally, we obtained the mean accuracy rates of the ten-fold cross-validation results and assessed the models through statistical analysis.

We developed two main models. One was a residual convolutional network, similar to nEMGnet from Yoo et al., and the other was a multilayer RNN model that utilised either a unidirectional or bidirectional GRU or LSTM. We achieved the highest mean accuracy rate (98.13 ± 1.05%) using ten-fold cross-validation when testing the model 1D-CNN+4-layer-unidirectional GRU without max pooling. Statistical analysis revealed that the accuracy rate of the model 1D-CNN+4-layer-unidirectional GRU without maxpooling was significantly higher than 5-layer CNN ($p = 0.024$), 11-layer CNN without residual layers ($p = 0.005$), 4-layer unidirectional LSTM ($p < 0.001$), 4-layer-bidirectional-LSTM ($p < 0.001$), 4-layer-unidirectional GRU ($p = 0.012$), 3-layer-1D-CNN+2-layer-unidirectional GRU ($p = 0.028$), 1-layer CNN-2-layer-unidirectional GRU ($p = 0.003$), 2-layer CNN-1-layer-bidirectional GRU ($p = 0.01$), and 2-layer CNN-1-layer-unidirectional LSTM ($p < 0.001$). We summarised all results obtained with ten-fold cross-validation on the dataset, which was split randomly into a train set and a test set, and the

p-values for the models compared to 1-layer-1-D-CNN+4-layer-unidirectional GRU are presented in Table 2. For the best-performing model, 1D-CNN+4-layer-unidirectional GRU without max pooling, we present the normalised confusion matrix of the ten-fold cross-validation results, along with the mean and standard deviation, in **Figure 2**. The mean and standard deviation of precision, recall, and F1 scores are also presented in **Table 3**. The pipeline and the detailed archi-

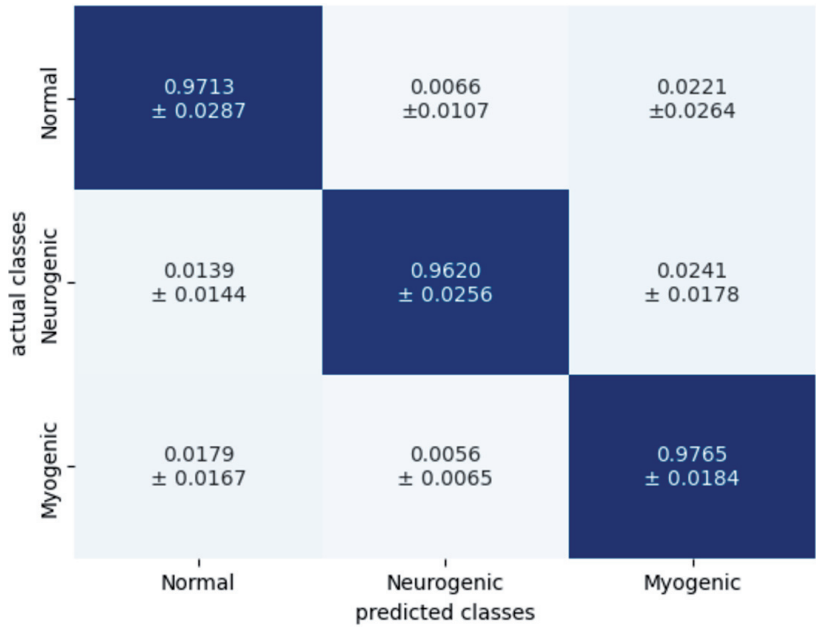

Figure 2. Confusion matrix of the optimised 1D-CNN-4-layer GRU model obtained through ten-fold cross-validation with normalised means and standard deviation.

Table 2. The optimised deep learning models and accuracy rates with ten-fold cross-validation.

Deep learning model	Epochs	10-fold cross-validation results (accuracy)	P*
5-layer CNN	50	96.55 ± 1.60 %	0.024
11-layer CNN without residual layers	50	95.71 ± 2.23 %	0.005
11-layer CNN with 3 residual layers	50	97.1 ± 1.05 %	0.052
4-layer-unidirectional-LSTM	70	95.23 ± 1.23%	<0.001
4-layer-bidirectional-LSTM	90	94.64 ± 1.87%	<0.001
4-layer -unidirectional -GRU	30	96.13 ± 1.82%	0.012
4-layer- bidirectional-GRU	35	97.04 ± 2.18 %	0.116
1-layer-1D-CNN+4-layer-unidirectional GRU	90	97.77 ± 1.08%	0.487
1 layer-1D-CNN+4-layer-unidirectional GRU (without maxpooling)	90	98.13 ± 1.05%	NA
1-layer-1D-CNN+4-layer unidirectional LSTM (without maxpooling)	250	95.52 ± 3.46%	0.052
1-layer-1D-CNN+4-layer-bidirectional GRU (without maxpooling)	90	97.81 ± 0.96%	0.513
1-layer-1D-CNN+3-layer-unidirectional GRU	80	96.47 ± 2.54%	0.093
2-layer-1D-CNN+2-layer-unidirectional GRU	70	97.82 ± 0.59%	0.445
2-layer-1D-CNN+3-layer-unidirectional GRU	90	97.77 ± 1.45%	0.878
3-layer-1D-CNN+3-layer-unidirectional GRU	70	96.98 ± 1.87%	0.145
3-layer-1D-CNN+2-layer-unidirectional GRU	90	97.02 ± 0.92%	0.028
1-layer CNN-2-layer-unidirectional GRU	70	96.50 ± 1.16%	**0.003**
2-layer CNN-1-layer-unidirectional GRU	70	97.92 ± 0.82%	0.661
2-layer CNN-1-layer-bidirectional GRU	70	96.27 ± 1.60%	0.010
2-layer CNN-1-layer-unidirectional LSTM	90	96.03 ± 1.08%	<0.001

* P-value for the models in comparison with 1-layer, -1D-CNN+4-layer-unidirectional GRU

Table 3. Precision, recall, and F1 scores for each group in the 1d-CNN-4-layer GRU model.

	Precision	Recall	F1-scores
Normal	0.9778 ± 0.0174	0.9845 ± 0.0139	0.9800 ± 0.0118
Neurogenic	0.9926 ± 0.0118	0.9761 ± 0.0162	0.9842 ± 0.0081
Myogenic	0.9787 ± 0.0148	0.9817 ± 0.0235	0.9800 ± 0.015
Accuracy	0.9813 ± 0.0105	0.9813 ± 0.0105	0.9813 ± 0.0105

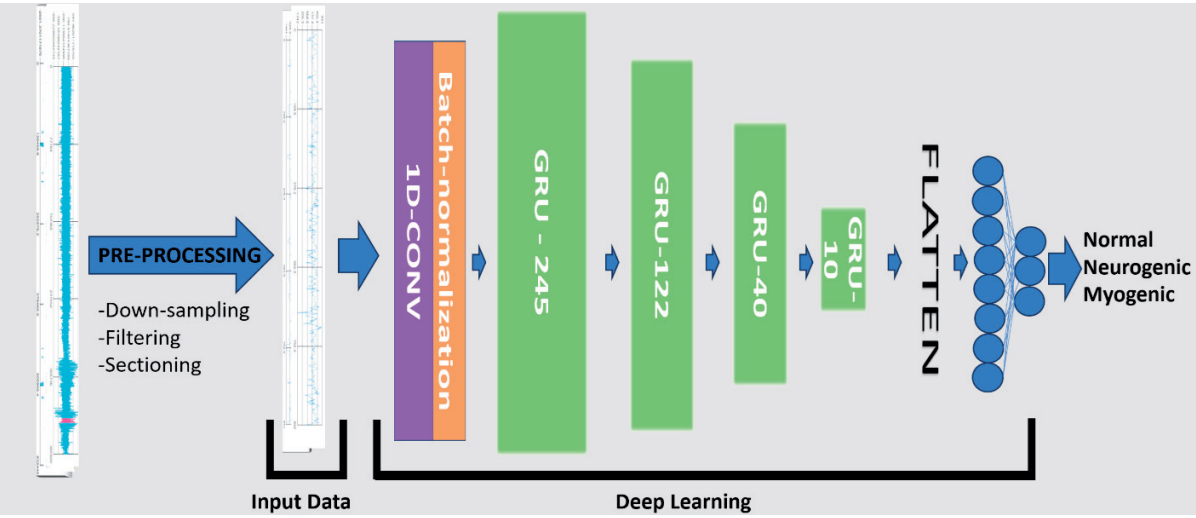

Figure 3. The deep learning pipeline achieved a mean accuracy rate of 98.13 ± 1.05% with ten-fold cross-validation. A similar architecture was also designed using LSTM and bilateral GRU instead of unilateral GRU.

tecture of the model are shown in **Figures 3** and **4**, respectively. The residual convolutional neural network pipeline and architecture are presented in **Figure 5**.

Discussion

Receiving needle EMG signal segments as input data, our study is the first to classify needle EMG data using a deep learning technique, achieving high accuracy. Both the residual convolutional neural networks and the 1D-CNN-4-layer GRU model without max pooling yielded high accuracy results. In the literature, although CNN is a successful deep learning tool receiving raw images as input, LSTM and GRU, as subtypes of RNN, show more promising results in time series data, such as electroencephalography [18]. Accordingly, our study mainly focused on RNN-based models to achieve high accuracy rates. Notably, GRU-based models showed slightly, and in some cases, statistically significantly better performance than LSTM-based models. Although the accuracy rates of LSTM-based models surpass those of GRU-based models in most deep learning studies in the literature, GRU-based models have also been reported to show superior success rates in some deep learning studies. The small size of the data set and the low dimension of the input data could account for the better accuracy results of GRU-based models [19,20]. The improved accuracy rates observed in unidirectional models may also be attributed to similar underlying factors.

Despite the success of RNN in time series data, residual CNN models also yielded promising results in our study. The residual CNN model architecture we designed was inspired by the architecture of nEMGnet, as described in the study by Yoo et al. [8]. Because the signal segment length for the input layer was 4-fold shorter than nEMGnet, the same number of layers would cause a negative output dimension. Therefore, the number of layers was fewer than that of nEMGnet. However, in our study, results for the accuracy of segment-based classification are significantly higher compared with the 62.35 ± 4.60% reported by Yoo et al. [8]. This may be due to both the character-

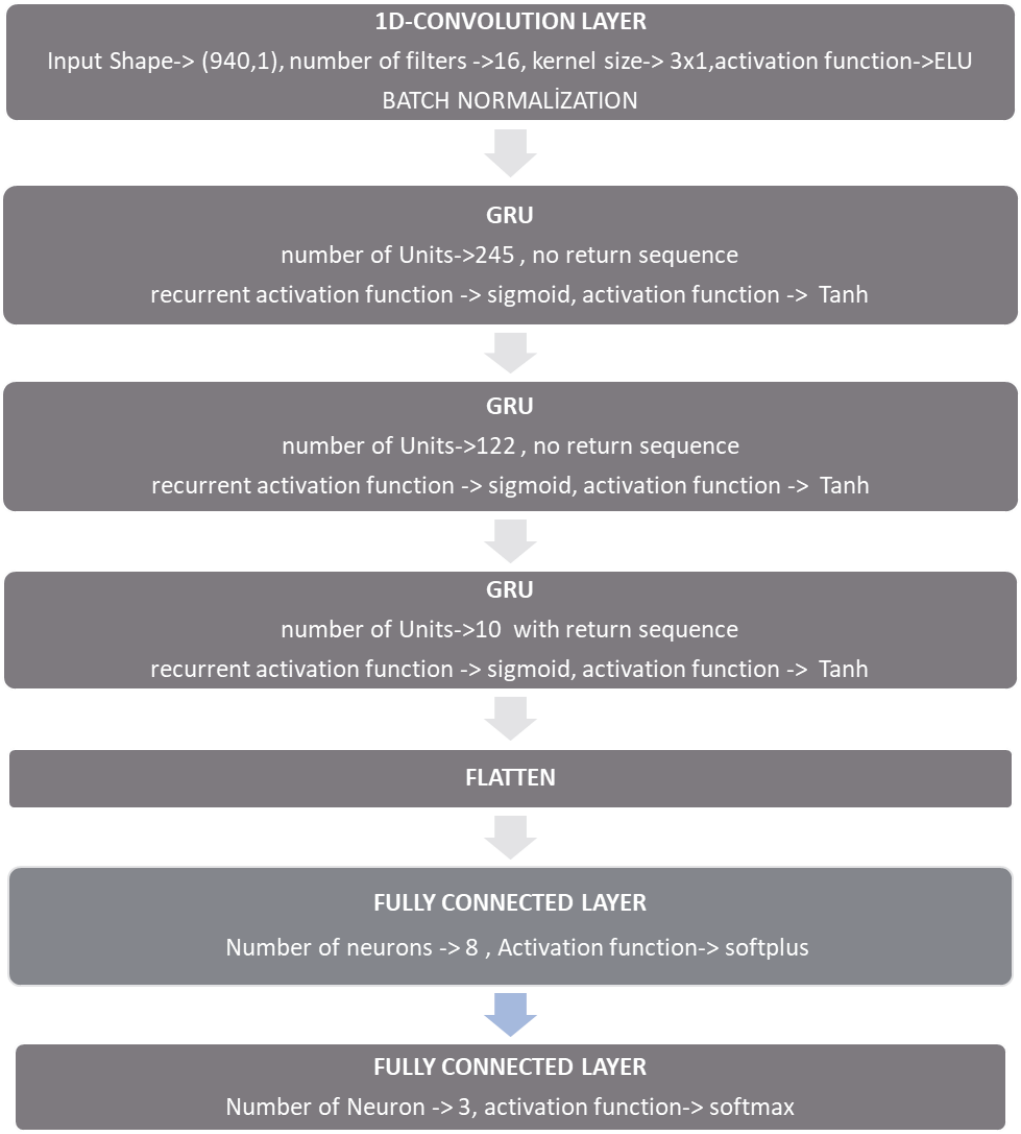

Figure 4. Detailed architecture of the 1D-CNN-4-layer GRU model deep learning model.

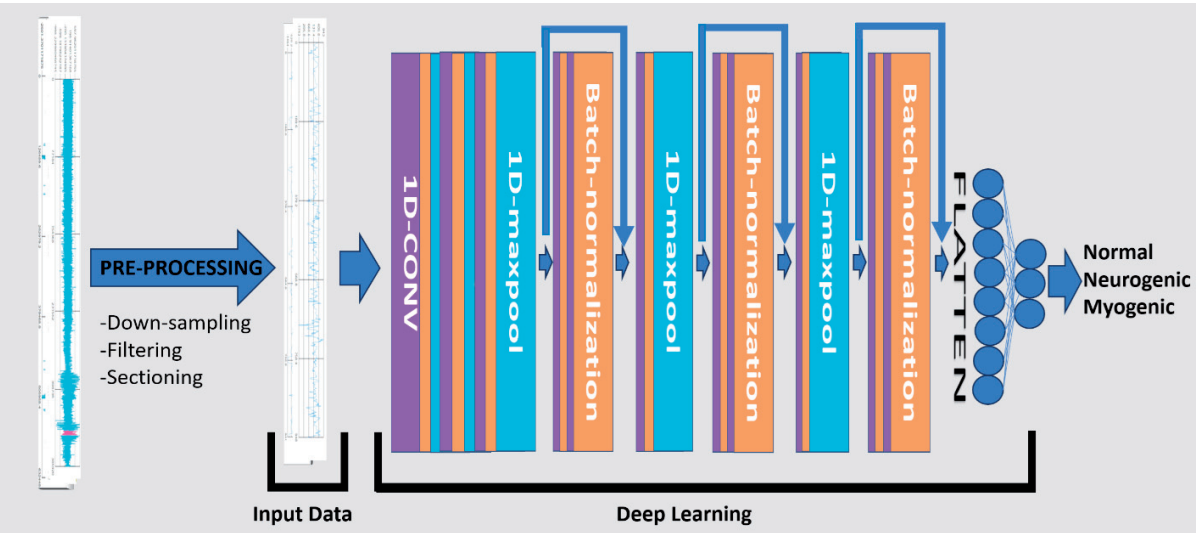

Figure 5. The residual convolutional neural network pipeline achieved an accuracy of 97.1 ± 1.05%.

istics of the dataset and nuances in the architectural design. Overall, adding residual layers could improve the performance of convolutional neural networks, as indicated by the statistical analysis of the study. Although the suggested model, model 1D-CNN+4-layer-unidirectional GRU without maxpooling, had significantly better performance than other CNN models, there was no statistically significant difference between the proposed model and the 11-layer CNN with residual layers.

On the other hand, Sengur et al. [12] achieved 96.8% accuracy in classifying normal and ALS groups using deep learning, which was higher than in other deep learning studies. However, the input data was not the EMG trace itself. EMG signal segments were converted into two-dimensional spectrograms through the CWT and Pseudowigner-Wille distribution function implementation.

Despite our study having some promising results, conventional machine learning techniques have yielded better accuracy. Extracting features from time-frequency analysis, Samantha et al. [21] achieved 99.5% accuracy in classifying ALS and normal groups using a genetic algorithm. Roy et al. [22] achieved 100% accuracy in discriminating between healthy and neuropathy conditions using a support vector machine and feature extraction methods based on time-frequency analysis. Dubey et al. [23] employed the Hilbert transform for feature extraction, achieving 99.5% accuracy with feedforward neural networks in classifying neuropathy, myopathy, and normal groups. Kamali et al. [24] achieved 100% accuracy using a random forest algorithm on samples from the tibialis anterior muscle for the classification of neuropathy, myopathy, and normal groups. They also employed time-frequency analysis for feature extraction. As seen in previous studies, feature extraction from time-frequency analysis yields the best results. Therefore, it may be pertinent to deepen the understanding of deep learning methodologies, as well as to design and refine their automatic feature extraction capabilities within the scope of time-frequency analysis. Further optimisation of deep learning models is necessary to achieve higher accuracy rates.

In our study, we created the dataset by retrospectively using recorded data. Similar to publicly available datasets, it consisted of needle EMG signals recorded from the muscles of 26 subjects: seven individuals with myopathy, eight individuals with neuropathy, and 11 healthy individuals. The number of subjects for each group was similar to the EMGLab dataset, a publicly available dataset used in most studies [1]. In contrast to the EMGLab dataset, our dataset did not contain patients with motor neuron disease. However, patients with radiculopathy and nerve palsies have typical clinical features and needle EMG patterns easily classified by an expert.

In contrast to the EMGLab dataset, since the study was retrospective, all patients had records from various muscles in our dataset. Additionally, depending on the patient's cooperation during the recorded part of the study, different recruitment patterns may be observed in the signal segments. On one hand, the involvement of various muscles and a wide variety of recruitment patterns could make the classification task more challenging. On the other hand, standardised recordings from the same muscle could help achieve more difficult tasks, such as disease subgroup classifications.

Despite deep neural networks being considered black boxes in terms of their feature extraction capabilities, we arranged the input parameters based on some electrophysiologic parameters to optimise the results. Considering that the sweep time of an EMG device screen is commonly adjusted to 5 msec/div or 10 msec/div, and a 100-msec trace provides an idea about motor unit action potential firing rates and morphology, we determined the length of EMG segments to be 100 msec. Most previous studies also used 100-msec segments [8]. The residual convolutional network architecture was similar to that of nEMGnet in the study by Yoo et al. [8]. However, we preferred smaller sizes of convolutional filter kernels to process high-frequency information. The design of the 1-D CNN-4-layer GRU model also contains architectural nuances. We estimated the number of units in the first RNN layer by considering specific parameters of the motor unit action potentials. The approximate motor unit action potential duration varies between 5 and 15 msec; it is shorter in myopathy (1-5 msec) but longer in neuropathy (15-25 msec) [3]. As 245 units of RNN capture approximately 25 milliseconds, it could be adequate for processing all types of motor unit action potential morphologies.

Additionally, because high-frequency information plays a crucial role in the discrimination

of myopathy, neuropathy, and standard motor unit action potentials, and GRU layers are specialised in learning time series patterns, the model was further optimised by omitting the maxpooling layer, which might cause losses in the high-frequency information captured by the GRU units. In most studies, a flattened layer is used before GRU units. However, we did not use this due to its low accuracy rates in tests. Because the convolution layer produces output data as several time series equal to the number of convolution filters, flattening it may cause some information to be lost in the time series processed by the GRU layer, which could lead to low accuracy rates.

The retrospective design of the study did not permit a standardised assessment of motor unit potential recruitment patterns. This limitation also led to the exclusion of acute neuropathies from the study, conditions that require analysis of recruitment patterns for an accurate understanding [3]. When a muscle is slightly contracted, the smallest motor unit action potential first starts firing, and its frequency is approximately 5-8 Hz at the beginning. By increasing the force, the frequency increases by 3-5 Hz, then a second larger motor unit is recruited at <15 Hz [25]. By 50% of maximal voluntary contraction, almost all motor units are recruited in small limb muscles, and further voluntary contraction rates could increase the firing frequency of motor unit action potentials up to 50 Hz. In large limb muscles, motor units are recruited at least 90% of maximal voluntary contraction, while muscle firing frequency reaches up to 30-40 Hz [26]. Thus, a motor unit action potential can be fired one to five times in small muscles and one to four times in large muscles within a 100-msec period. Although this rate increases in myogenic muscle, the number of motor unit action potentials recruited reduces in neurogenic muscle [3]. Finally, a wide variety of voluntary motor unit contraction rates have been successfully classified using the proposed deep learning model with 245 hidden units, which can capture periods of approximately 25 milliseconds. Considering that patients may have poor cooperation in recruitment tasks in clinical practice, the high accuracy rate of the proposed model in classifying neuromuscular disorders offers hope for the real-time implementation of the model. Nevertheless, standardisation of recruitment patterns, particularly for training neural network models, could both increase accuracy and help classify neuropathic and myopathic pattern subtypes.

We included a wide variety of small and large muscles in the limbs, hips, and shoulders in our study. The age of the subjects also ranged from 12 to 80 years. Different limb muscles have slightly different ranges of amplitude, duration, or polyphasic rates [3]. Furthermore, with ageing, motor unit action potential morphology can also differ slightly in size, and action potentials with longer durations and higher amplitudes are observed in healthy older subjects [3], which may lead to lower accuracy rates in pattern recognition. Despite acquiring data recorded from various muscles from a highly heterogeneous patient population, we achieved high accuracy rates using the proposed model. Although this situation may be helpful for the real-life implementation of the model, standardising the muscles could aid in muscle-specific assessments and increase accuracy. Studying only with the tibialis anterior muscle, Roy et al. [22] and Kamali et al. [24] achieved 100% accuracy.

After labelling the muscle according to the EMG report results, both artefacts and EMG signal segments that did not represent typical patterns of labelled neuromuscular conditions were excluded. This approach may increase accuracy rates and could be considered a barrier to implementing the model for real-time use. However, the fact that most patients in the data set had acquired myopathy and radiculopathy, with patchy muscle involvement, particularly in some acquired myopathies and radiculopathies, could explain the standard motor unit action potential patterns in pathologic signal segments [3]. Labelling these segments as neuropathic or myopathic would reduce the quality of training and result in low accuracy rates. However, there may still be some subtle patterns of the pathology that could not be discerned by an expert but could be detected with deep learning systems [27].

This study has certain limitations. The small sample size is a primary constraint, which may increase the risk of overfitting. Additionally, the limited number of patients prevented the use of patient-based classification and the application of an external validation data set, as referenced by de Jonge et al. [1]. Furthermore, we were unable to use an independent external validation set. The results cannot be generalised to the entire

neuromuscular disorder classification because the patient population in our study was small and restricted to patients with radiculopathy, axonal nerve injury, and predominantly inflammatory myopathies. As a result, both the limited diversity of the neuromuscular disorders and the small size of the datasets reduce the generalizability of the results and increase the risk of overfitting. Moreover, the real-time implementation of automatic classification of neuromuscular disorders requires the discrimination of motor unit action potentials, artefacts, needle insertion potentials, and resting state potentials. In a retrospectively designed clinical study, Hubers et al. [17] classified signal segments with artefacts, resting-state potentials, and motor unit action potentials with 96% accuracy at the first step, and then the classification of motor unit action potentials was performed. Our study focused solely on classifying motor unit action potentials. We were unable to assess different levels of recruitment due to the small dataset size and the retrospective design of the study. For this reason, having only reduced recruitment as a pathologic finding, acute neuropathies were excluded. Therefore, assessing recruitment patterns in future studies would also improve real-life generalizability.

The current findings of this study are preliminary. Future work will focus on increasing the number of patients for patient-based classification and expanding the dataset to include a broader spectrum of pathologies, such as common genetic and inflammatory myopathies, motor neuron disease, and both acquired and genetic polyneuropathies. Additionally, plans include conducting multicenter studies to obtain external validation datasets and using labelling based on expert consensus, aiming to enhance the reliability of the results. Ultimately, after developing deep learning models for recognising motor unit action potential morphology, artefacts, resting-state potentials, and recruitment patterns, the objective is to create real-time systems with high accuracy.

Conclusion

Deep learning is actively used in most current studies for pattern recognition tasks. The automatic feature extraction ability of deep learning tools reduces the complexity of the machine learning procedure. Showing good performance on time-series data, both 1D-CNN-4-layer GRU and residual convolutional neural network models show promising results on needle-EMG data in terms of neuromuscular disease classification. However, more clinical studies with larger datasets are needed to validate the results for clinical application.

Disclosures

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

This retrospective study was reviewed and approved by the Ethical Committee of Karadeniz Technical University on October 3, 2024 (Registry number: E-82554930-050.01.04-567570).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Generative AI

The authors declare that they did not use generative AI in the production of this manuscript.

CRediT authorship contribution statement

Isil Tatlidil: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Software, Visualisation, Writing – original draft. **Murat Ekinci**: Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Cavit Boz**: Data curation, Resources, Supervision, Validation, Writing – review & editing.

References

1. De Jonge S, Potters W V, Verhamme C. Artificial intelligence for automatic classification of needle EMG signals: A scoping review. Clin Neurophys-

iol. 2024 Mar;159:41–55. https://doi.org/10.1016/j.clinph.2023.12.134.
2. Rubin DI. Needle electromyography: Basic concepts. Handb Clin Neurol. 2019 Jan 1;160:243–56.
3. Preston DC, Shapiro BE. Electromyography and Neuromuscular Disorders. 4th ed. Philadelphia: Elsevier; 2021. 134–259 p.
4. Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve. 2006 Aug;34(2):238–41. https://doi.org/10.1002/mus.20554.
5. Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. Phys Med Biol. 2020 Mar 3;65(5):05TR01. https://doi.org/10.1088/1361-6560/ab6f51.
6. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol. 2020 Feb 27;9(2):14. https://doi.org/10.1167/tvst.9.2.14.
7. Caiafa CF, Sun Z, Tanaka T, Marti-Puig P, Solé-Casals J. Special Issue 'Machine Learning Methods for Biomedical Data Analysis'. Sensors (Basel). 2023 Nov 24;23(23). https://doi.org/10.3390/s23239377.
8. Yoo J, Yoo I, Youn I, Kim SM, Yu R, et al. Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability. Comput Methods Programs Biomed. 2022 Nov;226:107079. https://doi.org/10.1016/j.cmpb.2022.107079.
9. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. Comput Methods Programs Biomed. 2018 Jul;161:1–13. https://doi.org/10.1016/j.cmpb.2018.04.005.
10. Nodera H, Osaki Y, Yamazaki H, Mori A, Izumi Y, Kaji R. Deep learning for waveform identification of resting needle electromyography signals. Clin Neurophysiol. 2019 May;130(5):617–23. https://doi.org/10.1016/j.clinph.2019.01.024.
11. Nam S, Sohn MK, Kim HA, Kong HJ, Jung IY. Development of artificial intelligence to support needle electromyography diagnostic analysis. Healthc Inform Res. 2019;25(2):131. https://doi.org/10.4258/hir.2019.25.2.131.
12. Sengur A, Akbulut Y, Guo Y, Bajaj V. Classification of amyotrophic lateral sclerosis disease based on convolutional neural network and reinforcement sample learning algorithm. Health Inf Sci Syst. 2017 Dec 30;5(1):9. https://doi.org/10.1007/s13755-017-0029-6.
13. Zhang Z, He C, Yang K. A novel surface electromyographic signal-based hand gesture prediction using a recurrent neural network. Sensors. 2020 Jul 17;20(14):3994. https://doi.org/10.3390/s20143994.
14. Das S, Tariq A, Santos T, Kantareddy SS, Banerjee I. Recurrent neural networks (RNNs): Architectures, training tricks, and introduction to influential research. Newyork: Humana; 2023. https://doi.org/ 10.1007/978-1-0716-3195-9_4.
15. Keleş AD, Turksoy RT, Yucesoy CA. The use of non-normalized surface EMG and feature inputs for LSTM-based powered ankle prosthesis control algorithm development. Front Neurosci. 2023;17:1158280. https://doi.org/10.3389/fnins.2023.1158280.
16. Aviles M, Alvarez-Alvarado JM, Robles-Ocampo JB, Sevilla-Camacho PY, Rodríguez-Reséndiz J. Optimizing RNNs for EMG Signal Classification: A novel strategy using grey wolf optimization. bioengineering (Basel). 2024 Jan 13;11(1). https://doi.org/10.3390/bioengineering11010077.
17. Hubers D, Potters W, Paalvast O, de Jonge S, Doelkahar B, Tannemaat M et al. Artificial intelligence-based classification of motor unit action potentials in real-world needle EMG recordings. Clinical Neurophysiology. 2023 Dec;156:220–7. https://doi.org/10.1016/j.clinph.2023.10.008.
18. Khademi Z, Ebrahimi F, Kordy HM. A review of critical challenges in MI-BCI: From conventional to deep learning methods. J Neurosci Methods. 2023 Jan 1;383:109736. https://doi.org/10.1016/j.jneumeth.
19. Yang S, Yu X, Zhou Y. LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp review dataset as an example. In: 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). IEEE; 2020. p. 98–101.
20. Rivas F, Sierra-Garcia JE, Camara JM. Comparison of LSTM- and GRU-Type RNN networks for attention and meditation prediction on raw EEG data from low-cost headsets. Electronics (Basel). 2025 Feb 12;14(4):707. https://doi.org/10.3390/electronics14040707.
21. Samanta K, Chatterjee S, Bose R. Neuromuscular disease detection based on feature extraction from time–frequency images of EMG signals employing robust hyperbolic Stockwell transform. Int J Imaging Syst Technol. 2022 Jul 25;32(4):1251–62. https://doi.org/10.1002/ima.22709.
22. Roy SS, Dey D, Karmakar A, Roy AS, Ashutosh K, Choudhury NR. Detection of abnormal electromyograms employing DWT-based amplitude envelope analysis using Teager energy operator. Int J Biomed Eng Technol. 2022;40(3):224. https://doi.org/10.1504/ijbet.2022.126493
23. Dubey R, Kumar M, Upadhyay A, Pachori RB. Automated diagnosis of muscle diseases from EMG signals using an empirical mode decomposition-based method. Biomed Signal Process Control. 2022 Jan 1;71:103098. https://doi.org/10.1016/j.bspc.2021.103098
24. Kamali T, Stashuk DW. Electrophysiological muscle classification using multiple instance learning and unsupervised time and spectral domain analysis. IEEE Trans Biomed Eng. 2018 Nov;65(11):2494–502. https://doi.org/10.1109/TBME.2018.2802200.
25. Nandedkar SD, Barkhaus PE, Stålberg E V. Motor unit recruitment and firing rate at low force of contraction. Muscle Nerve. 2022 Dec;66(6):750–6. https://doi.org/10.1002/mus.27737.
26. Masakado Y. Motor unit firing behavior in man. Keio J Med. 1994 Sep;43(3):137–42. https://doi.org/10.2302/kjm.43.137.
27. Van Putten MJAM, Olbrich S, Arns M. Predicting sex from brain rhythms with deep learning. Sci Rep. 2018 Feb 15;8(1):3069. https://doi.org/10.1038/s41598-018-21495-7.