# Deep learning model for needle electromyography electrodiagnosis in comparison with physician assessment: A retrospective study

Ilhan Yoo[a†], Jaesung Yoo[b†], Dongmin Kim[c], Ina Youn[d], Hyodong Kim[a], Michelle Youn[a], Jun Hee Won[e], Woosup Cho[e], Youho Myong[e], Sehoon Kim[e], Ri Yu[e], Sung-Min Kim[f], Kwangsoo Kim[g], Seung-Bo Lee[h*] and Keewon Kim[e*]

[a] Department of Neurology, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul, Republic of Korea

[b] School of Electrical Engineering, Korea University, Seoul, Republic of Korea

[c] Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea

[d] Department of Computer Science, New York University, New York, USA

[e] Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[f] Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

[g] Transdisciplinary Department of Medicine & Advanced Technology, Seoul National University Hospital, Seoul, Republic of Korea

[h] Department of Medical Informatics, Keimyung University School of Medicine, Daegu, Republic of Korea

† These Authors contributed equally to this work.

* Co-corresponding authors

## Address for Correspondence

Keewon Kim, MD, PhD

Department of Rehabilitation Medicine, Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul, Republic of Korea

Tel: +82-2-2072-0744

Fax: +82-2-6072-5244

E-mail: keewonkimm.d@gmail.com

**And**

Seung-Bo Lee, PhD

Department of Medical Informatics, Keimyung University School of Medicine, Daegu, South Korea

Tel: +82-10-8840-9483

E-mail: koreateam23@gmail.com

# Abstract

**Background**

Electromyography (EMG) plays an important role in the diagnosis of neuromuscular diseases by identifying the characteristic abnormalities of the waveform. However, it has some limitations in that there are discrepancies in the interpretations of electrodiagnosis results among physicians, and the accuracy of electrodiagnosis using EMG relies on the experience of physicians. To overcome these limitations of EMG, we investigated whether the deep-learning approach can better classify EMG waveforms into neuropathy, myopathy, and normal physicians' electrodiagnosis results.

**Methods and findings**

Needle EMG data (58 patients, 382 muscles) from the Seoul National University Hospital database from June 2015 to July 2020 were used. A one-dimensional convolutional neural network model was used as the deep-learning model. The deep-learning model and the six physicians classified and electro-diagnosed the EMG waveforms as myopathy, neuropathy, or normal.

The accuracy, sensitivity, specificity, positive predictive value and F1 score of the deep learning model were 0.720, 0.715, 0.858, 0.726, and 0.715, respectively, and the physicians' mean scores were 0.537, 0.527, 0.770, 0.582, and 0.511, respectively. The performance of the deep-learning model for predicting myopathy, neuropathy, and normal state was also evaluated using the area under the receiver operating characteristic curve, and the results were 0.874 (95% confidence interval [CI] 0.858–0.889), 0.781 (95% CI 0.723–0.839), and 0.847 (95% CI 0.836–0.858), respectively.

**Conclusions**

This study demonstrated that deep-learning could contribute to interpreting the EMG of patients with neuromuscular disease on behalf of physicians and assist physicians' decision-making regarding diagnosing patients with neuromuscular disease. Large prospective cohort studies with more diverse neuromuscular diseases will further improve the performance of deep-learning-based EMG

interpretation in the future.

# Author summary

**Why was this study done?**

· Electromyography (EMG) is an important electrophysiological test that is widely performed when diagnosing patients with neuromuscular disease.

· Despite the importance of EMG in diagnosis, the accuracy of EMG electrodiagnosis is not sufficiently precise, and there are often discrepancies among examiners. Therefore, a more objective and accurate method of EMG analysis is required.

· Although EMG in volitional state plays important role in diagnosing neuromuscular disease, previous studies have mainly focused on analysis of gestures using surface EMG or analysis of the abnormal spontaneous activities from needle EMG during resting state. Therefore, additional research is needed for clinical application of deep learning-based electrodiagnosis of neuromuscular disease.

**What did the researchers do and find?**

· The EMG data from 58 patients were used in this study. Patients were classified by the deep-learning model and were electro-diagnosed by the physicians as myopathy, neuropathy, or normal. After classification and electrodiagnosis, the classification results of the deep-learning model were compared with the electrodiagnosis results obtained by neurologists and rehabilitation medicine doctors.

· The accuracy of classifying EMG data using our deep-learning model was 0.720, and the average accuracy of electrodiagnosis by physicians was 0.537.

· Classification using our deep-learning model was more accurate and faster than electrodiagnosis by physicians.

**What do these findings mean?**

- Deep-learning can electrodiagnose EMG on behalf of a physician and can be used to assist physicians in diagnosing patients with neuromuscular disease.

- The performance of deep-learning could be further enhanced by conducting large-scale prospective study that includes patients with more diverse and specific types of neuromuscular diseases.

# Introduction

Electromyography (EMG) is an electrophysiological test that records the electrical activity generated from nerves, muscles, and neuromuscular junctions via a needle inserted into the muscle or surface electrode during resting and volitional states [1-6]. Disorders of the peripheral nerves and muscles are characterized by abnormalities in EMG signals, which reflect the anatomical and physiological states of the peripheral nerves and muscles [1-6].

The signal recorded during muscle contraction, consists of the motor unit action potential (MUAP), which can only be measured from muscles in a volitional state and not in a resting state; EMG signals during the resting state cannot distinguish between neuropathy and myopathy. MUAP is an essential key in distinguishing whether a patient has neuropathy or myopathy [1, 5-12]. Additionally, when neuromuscular disease is in an early state or in a mild abnormal state, abnormal findings are found only in MUAP during a volitional state, not during a resting state [13, 14]. Patients with peripheral neuropathy commonly show MUAP with high amplitudes, long durations, and reduced recruitment, whereas patients with myopathy commonly exhibit MUAP with small amplitudes, short durations, and early recruitment on EMG evaluation [1, 5-12]. Needle EMG (nEMG) is a type of EMG that uses a needle as an electrode and is useful for recording MUAPs from individual motor units and deep muscles [15].

Although EMG plays an important role in diagnosing normal, neuropathy or myopathy in a patient, it has some limitations. First, the accuracy of EMG diagnosis relies on the proficiency of the examiner. Previous studies have reported that the sensitivity of EMG in the diagnosis of neuropathy, myopathy, and normal was 47–83%, specificity was 73–81% and interrater reliability was 62–81% [16-18]. Second, considerable time and effort are required to accurately detect abnormalities in EMG signals. The prevalence of neuropathy and myopathy continues to rise, creating a burden on physicians with regards to the increased number of EMG that require electrodiagnosis [19-22]. An accurate, efficient, and automated approach may be helpful for clinical diagnosis using nEMG.

Recently, deep-learning has been used to analyze large datasets in many fields, and it has also been applied to clinical data, including waveform and time series data, such as electrocardiography and electroencephalography [23-26]. There have also been a number of studies regarding deep-learning in medical applications where the performance of the deep-learning model was similar to or better than that of humans [27-30]. In previous EMG studies, machine learning has been mainly used to analyze the abnormal spontaneous activities that are present on EMG during the resting state or gestures that are not related to the diagnosis of neuromuscular diseases using surface EMG [31-35]. To the best of our knowledge, few studies have used using machine learning to analyze nEMG signals during volitional states.

To overcome the diagnostic limitations of EMG and investigate the feasibility of machine learning on nEMG during a volitional state, we developed a deep-learning model that can classify patients as having myopathy, neuropathy or being normal state based on volitional state nEMG signals. We retrospectively reviewed the nEMG data of patients with peripheral neuropathy or myopathy as well as those without any neuromuscular disease. The classification results of the deep-learning model and the electrodiagnosis results of the six physicians were compared.

# Methods

## Study design and preparation

The nEMG signal data consisted of 58 patients who visited Seoul National University Hospital between June 2015 and July 2020. Each patient was labeled as having peripheral neuropathy, myopathy, or normal based on the final diagnosis. This study was approved by the Institutional Review Board of Seoul National University Hospital (No. 2008-055-1147) and was conducted in accordance with the Declaration of Helsinki and its later amendments. Informed consent was waved because this study was a retrospective analysis and patient's private information was anonymized before analysis.

nEMG evaluation was performed using a Nicolet EDX EMG system and a monopolar needle electrode. The filter settings were set at 20 Hz (low cut) and 10 kHz (high cut). The nEMG signals were measured at a sampling rate of 48 kHz. The results of the final 10 s of the nEMG were stored in an EMG machine and used for analysis.

A certified neurologist and certified rehabilitation medicine doctor reviewed the nEMG data and confirmed the diagnosis of all patients. After extracting the nEMG signal stored as numerical data in the EMG device, it was converted into waveform data using MATLAB (version R2020b; The MathWorks Inc., Natick, Massachusetts, United States). Among the created waveform data, electromyographic artifacts that occurred from the movement of the needle electrode or the patient, were removed at the beginning and end of the EMG data. The noise in the remaining data was preserved. Based on the elbow joint of the upper extremity and the knee joint of the lower extremity, muscles close to the trunk were labeled as proximal muscles and those far from the trunk as distal muscles.

## Electrodiagnosis by physicians

In this study, a web-based nEMG signal labeling platform was developed to allow residents who belonged to different institutions to electrodiagnose patients. After de-identifying the patient

information, the nEMG waveform data were displayed on the screen in a similar manner to the EMG data presented on the EMG device (S1 Fig).

Two neurology residents and four rehabilitation medicine residents assessed the nEMG signal of the muscles examined for each patient and electro-diagnosed patients as myopathy, neuropathy, or normal without clinical information other than nEMG results. When the physician clicked the anonymized patient ID, the nEMG waveform data were presented on the screen along with a sound similar to the sound from the EMG device. The real-time waveform data were presented on the left, and the waveform data were stacked on the right for 500 μs. Physicians were allowed to change the voltage amplitude ticks of the screen to 100, 200, 500 μV, 1, and 2 mV. Physicians first annotated each nEMG signal from different muscles and thereafter diagnosed patients based on the muscle annotation results. The physicians' electrodiagnosis results were stored within the platform.

## Classification by deep-learning model

The nEMG signals were down-sampled to 10 kHz to reduce computational complexity. Each data was partitioned into windowed segments with lengths of 0.4 s and hop sizes of 0.1 s. The length and hop size of the windows were heuristically selected. After slicing, there were 3664, 2700, and 1706 segments from patients with neuropathy, myopathy, and normal patients, respectively.

A one-dimensional convolutional neural network (CNN) was used as the deep-learning model [36]. The CNN was designed after the motives of ResNes and VGG, which are well-performing image-classification models [37, 38]. The CNN comprised seven spatial reduction blocks, five residual blocks, and fully connected layers (S2 Fig). The spatial reduction block consisted of convolutional layers, batch normalization, rectified linear unit (ReLU), and max pooling. The residual block contained similar layers with added residual connections [37]. The fully connected layers consisted of 512, 256, 64, and 16 hidden layer neurons with a leaky ReLU activation function. The Softmax function was applied to the final three-output layer. The model received raw signal segments and predicted the probabilities of three class (myopathy, neuropathy, normal).

Hyper-parameters were determined empirically with a learning rate of $10^{-3}$, batch size of 32, and 100 epochs. Early stopping was performed by evaluating the accuracy of the validation set every 30 updates, and the patience value was set to 100 [39]. Cross-entropy loss was used as the loss function, with class weights applied inversely proportional to the number of signal segments from the training set. Deep-learning performance was measured through a $5 \times 3$-fold cross validation ($5 \times 3$ CV) due to small number of subjects.

The deep-learning model used two stages to classify patients: first, it received multiple segments sliced from a single muscle's nEMG data and returned the three-class probability of the segment being myopathy, neuropathy, or normal. The probabilities of the segments were averaged to obtain a three-class probability for each muscle. After the probabilities of all the muscles from each patient were computed, a feature vector was produced for the patient using the individual muscle probabilities. The two methods were compared to generate the feature vectors. Another method was to average the muscle probabilities per muscle location label (proximal or distal) to produce a six-dimensional vector. When the patient did not have any muscle probability for a muscle location label, a mean value of 1/3 was imputed to prevent undesired bias. The performance of the deep-learning model presented in this article was obtained by averaging all muscle probabilities. The second stage receives the generated feature vector and classifies the three-class diagnosis of the subject using a logistic regression classifier.

After the model was trained, an explainable artificial intelligence technique named feature visualization was applied to visualize the features that the model focused on [40]. The initial feature was sampled from a Gaussian distribution N (0,0.5), and random jitter was applied as regularization with 12.5% of the signal segment length. The learning rate was $10^{-2}$ and gradient descent updates were applied 3000 times.

**Assessing the results of the deep-learning model and the physicians**

The performance of the deep-learning model and the physicians were evaluated using the following metrics: accuracy, F1 score, area under the receiver operating characteristic (ROC) curve, positive

predictive value (PPV; precision), sensitivity (recall), and specificity. Based on these metrics, we compared the results classified by two versions of the deep-leaning model with and without muscle location labels and averaged the results by six physicians.

**Statistical analysis**

Statistical analyses were performed using R statistical software (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria) and the Python programming language (version 3.6). Normality for continuous variables was assessed using the Shapiro-Wilk test. The differences among the groups for categorical variables were assessed using Pearson's $\chi^2$ test and those for continuous variables were assessed using the Kruskal–Wallis test. Data are expressed as mean ± standard deviation for continuous variables and number (%) for categorical variables. The degree of agreement was calculated between physicians and the deep-learning model and was expressed with the value of Fleiss kappa. All metrics except accuracy (F1 score, AUROC, PPV, recall, and specificity) were binary classification and were measured by averaging each class metric using the one-versus-rest method.

# Results

## Patients' characteristics

There were 20 patients without any neuromuscular disease; 19 patients with neuropathy, including radiculopathy, motor axonal polyneuropathy, and motor neuron disease; and 19 patients with myopathy, including muscular dystrophy and inflammatory myopathy. The number of nEMG data points used for analysis was 124, 161, and 97 for myopathy, neuropathy, and normal patients, respectively. The remaining demographic characteristics are presented in Table 1.

**Table 1. Demographic characteristics of patients and their needle electromyography data**

|  | Myopathy | Neuropathy | Normal | p-value |
|---|---|---|---|---|
| Number of patients | 19 | 19 | 20 | |
| Female, n (%) | 14 (73.7) | 12 (63.2) | 13 (65) | 0.761 |
| Age (mean±SD) | 52.2±20.1 | 58.4±15.1 | 60.2±16.9 | 0.329 |
| Number of nEMG data according to location of muscle (%) | | | | <0.001 |
| Distal muscles | 60 (48.4) | 97 (60.2) | 80 (82.5) | |
| Proximal muscles | 64 (51.6) | 64 (39.8) | 17 (17.5) | |
| Number of nEMG (mean±SD) | 6.53±3.82 | 8.47±4.59 | 4.85±1.93 | 0.006 |
| Total signal length (sec) | 313.54 | 423.12 | 204.31 | |

nEMG=needle electromyography

Data are expressed as mean±standard deviation.

## Performance results of deep-learning and physicians

The classification performance of the deep-learning algorithm was compared with the electrodiagnosis results of physicians using the following metrics: accuracy, sensitivity, specificity, PPV, and F1 score. Sliced segments less than 0.4 s were excluded because the deep leaning model only receives segment longer than 0.4 s; total of eight patients and 10 muscle nEMG data were excluded from analysis because these were not annotated by either the physicians or the deep-learning

model. The results of the deep-learning model that did not use muscle location information are presented as the above-mentioned metrics. The accuracy, sensitivity, specificity, PPV, and F1 score of the deep-learning model were 0.720±0.040, 0.715±0.050, 0.858±0.023, 0.726±0.032, and 0.715±0.044, respectively. That of the physicians were 0.537±0.069, 0.527±0.069, 0.770±0.034, 0.582±0.076, and 0.511±0.084, respectively (Table 2). The degree of agreement, expressed in Fleiss κ, between physicians was 0.26 and between physicians and the deep-learning model was 0.26 (S1 Table).

**Table 2. The electrodiagnosis results by physicians and classification results by the deep-learning model**

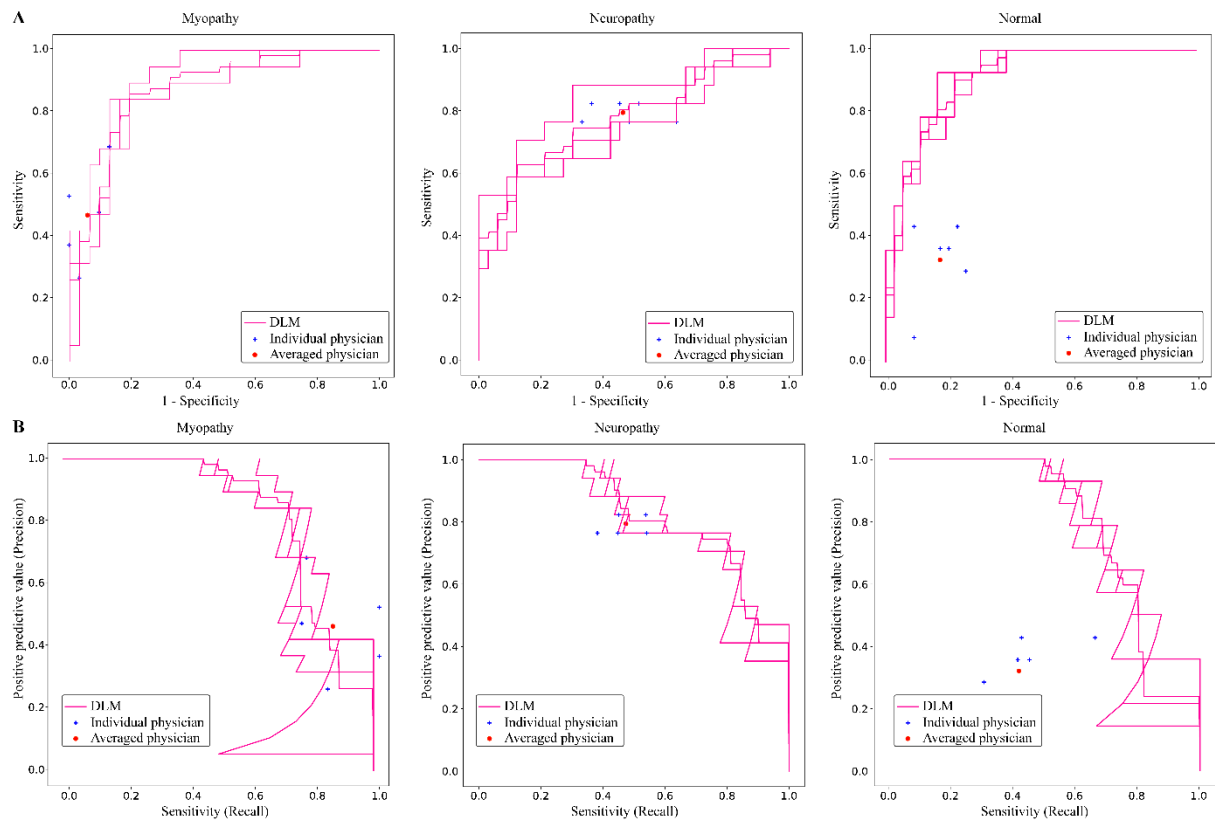|  | Physicians (95% CI) | Deep-learning model (95% CI) |
| --- | --- | --- |
| Accuracy | 0.537 (0.491–0.583)[*] | 0.811 (0.682–0.758)[†] |
| Sensitivity (recall) | 0.527 (0.480–0.573)[*] | 0.720 (0.668–0.762)[†] |
| Specificity | 0.770 (0.747–0.793)[*] | 0.853 (0.836–0.879)[†] |
| Positive predictive value (PPV, Precision) | 0.582 (0.530–0.633)[*] | 0.725 (0.695–0.757)[†] |
| F1 score | 0.511 (0.455–0.568)[*] | 0.715 (0.673–0.757)[†] |

[*] Average results of electrodiagnosis by six physicians

[†] Average results of classification using three deep-learning models

The metrics, except for accuracy, are the average of the binary classification scores for each class from the one-versus-rest method.

95% CI, 95% confidence interval

Per-class ROC curves and precision-recall curves of the deep-learning model were measured and depicted. For comparison, the electrodiagnosis results of individual physicians and averaged physicians were measured and depicted in terms of sensitivity, specificity, and PPV in Fig. 1. The per-class physicians' results for myopathy and neuropathy were placed near the ROC and precision-recall curves of the deep-learning model. The deep-learning model outperformed the classification of normal patients. The AUROC of the deep-learning model classification result per-class were 0.874 (95% confidence interval [CI] 0.858–0.889), 0.781 (95% CI 0.723–0.839), and 0.906 (95% CI 0.899–0.913) for myopathy, neuropathy, and normal patients, respectively.

**Fig 1. Per-class receiver operating characteristic and precision-recall curves of the deep-learning model and six physicians**

DLM, Deep-learning model

Area under the receiver operating characteristic curve (A) and precision-recall curve (B) were measured and depicted by dividing all data into myopathy, neuropathy, and normal.

The individual physician performance is annotated by the blue cross, and the average physician performance is annotated by the red dot.

The overall prediction pattern was identified from the confusion matrix of physicians and the deep-learning model (Fig 2). The correctly predicted ratios for myopathy, neuropathy, and normal patients by the deep-learning model were 80.70%±4.96%, 64.71%±4.80%, and 69.05%±12.14%, respectively. The correctly electro-diagnosed ratios for myopathy, neuropathy, and normal patients by the physicians were 49.49%±13.04%, 79.41%±2.94%, and 32.14%±12.20%, respectively.
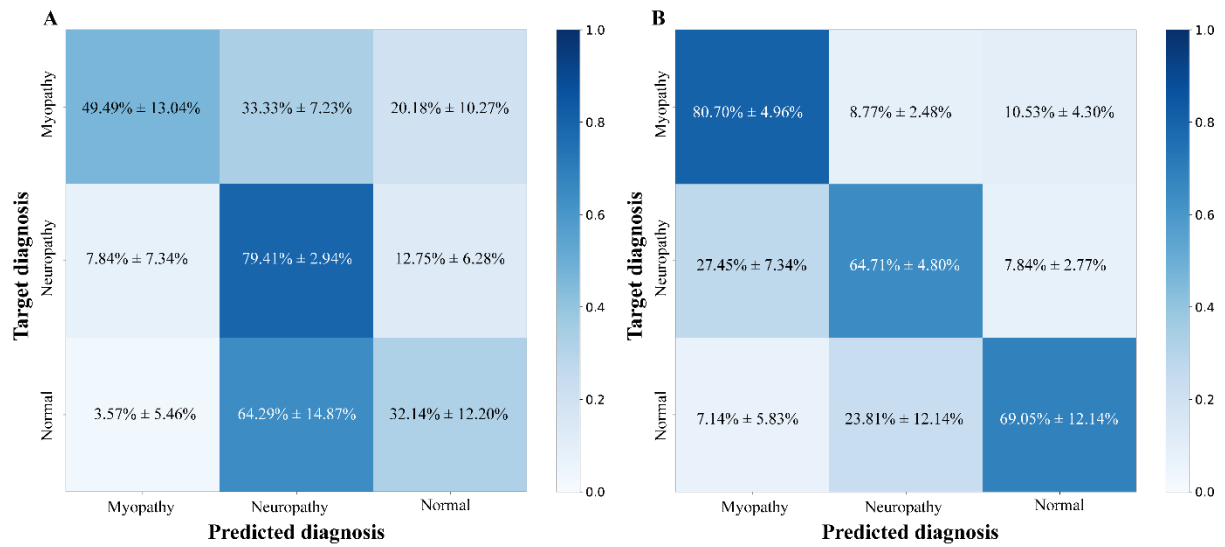
**Fig 2. Confusion matrices of (A) physician electrodiagnosis results and (B) the deep-learning model prediction results**

## Performance of the machine learning model considering muscle location information

The performance of the deep-learning model did not change significantly when muscle location information was added. The accuracy, sensitivity, specificity, PPC, and F1 score of the deep-learning model considering muscle location information were 0.700±0.053, 0.698±0.048, 0.848±0.026, 0.703±0.058, and 0.694±0.049, respectively (S2 Table). The AUROC of the same deep-learning model were 0.841 (95% CI 0.811–0.871), 0.736 (95% CI 0.731–0.741), and 0.792 (95% CI 0.719–0.864) for myopathy, neuropathy, and normal patients, respectively (S3 Fig). The correctly predicted ratios for myopathy, neuropathy, and normal patients were 80.70%±8.95%, 54.90%±2.77%, and 73.81%±3.37%, respectively (S4 Fig). The prediction accuracies of CNN with muscle location information and those of CNN without muscle location information were as follows: the former and latter results for myopathy, neuropathy, and normal patients were 80.70%±8.95%, 54.90%±2.77%, and 73.81%±3.37% and 80.70%±4.96%, 64.71%±4.80%, and 69.05%±12.14%, respectively (S4 Fig).

## Mis-predicted cases of the deep-learning model

To analyze the reason for the misclassification of the deep-learning model, we reviewed misclassified

nEMG signals. Examples of these are shown in Figure S5.

**Learned features of the deep-learning model**

The learned features of the deep-learning model were identified from the signals created by applying feature visualization. The generated signals were similar to the typical characteristics of neuropathy, myopathy, and normal patients. The waveform most likely to be predicted as myopathy and neuropathy showed small amplitudes and short durations (Fig 3A,B) and high amplitudes and long durations, respectively (Fig 3C,D). Thus, we validated that the deep-learning model made predictions based on relevant features rather than on artifacts.
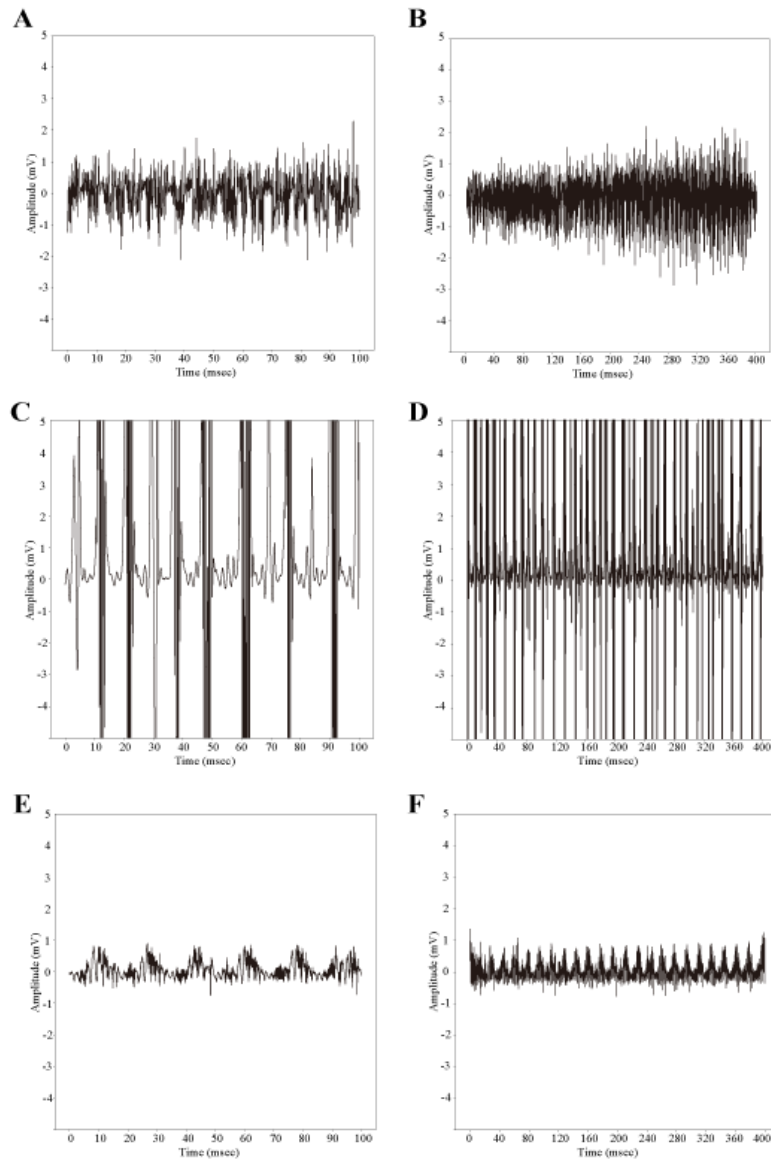
**Fig 3. Feature visualization results of the deep-learning model**

A and B, Myopathy; C and D, Neuropathy; E and F, Normal.

Note that A, C, and E were plotted with 10 ms of x-axis interval and B, D, and F were plotted with 40

ms of x-axis interval to show the recruitment and interference pattern of the overall waveforms.

# Discussion

The aim of the present study was to validate the diagnostic performance of machine learning using nEMG signals in a volitional state, and whether machine learning could support physicians' decisions for more accurate and efficient diagnosis. For this purpose, deep-learning was used to classify the nEMG signals as myopathy, neuropathy, and normal, and the classified results were compared with the electrodiagnosis results of six physicians. The results demonstrated that our deep-learning model outperformed the electrodiagnosis results of physicians in all performance scores, with an accuracy of $0.720\pm0.040$ for the deep-learning model and $0.537\pm0.069$ for physicians.

There have been reports that machine learning shows good performance when applied to analyzing EMG signals [31-35]. However, previous studies regarding EMG have mainly focused on abnormal spontaneous activities in the resting state, and the signals were preprocessed to form two-dimensional features, which may increase computational complexity [33, 34]. It is well known that EMG evaluation is important in the diagnosis of the neuromuscular disease; and EMG signals during a volitional state play an essential role in differentiating myopathy, neuropathy, normal patients as well as identifying whether the neuromuscular disease is in the initial state or the minimally involved state [1-6, 8, 13, 14, 41]. To the best of our knowledge, only few studies have analyzed EMG data during volitional states using deep-learning. We analyzed the nEMG signal in the volitional state, which is important for the electrodiagnosis of neuropathy and myopathy, and confirmed that it showed better performance than physicians.

To diagnose a patient, the nEMG results of all tested muscles should be considered together. However, not all nEMG signals from muscles show abnormalities, and the number of muscles examined may differ slightly between patients; therefore, this variation needs to be considered for the deep-learning model. To solve this variation problem, we constructed feature vectors for patients from the muscle signal prediction probabilities and utilized an additional classifier to determine the classification result of the patient. This method allows the deep-learning model to not only consider all the signals measured from different muscles from different patients in a consistent format, but also

make predictions based on the signal characteristics, and not on the prevalence of nEMG signals in certain muscle types.

Usually, there is a typical pattern of muscle involvement in which peripheral neuropathy mainly shows abnormalities in the distal muscles, whereas myopathy mainly shows abnormalities in the proximal muscles [13]. Although muscle location information is meaningful for differentiating neuropathy from myopathy, there was no significant change in the performance of the deep-learning model when muscle location information was used. This may be due to the following two reasons; first, some types of myopathy may show abnormalities in both proximal and distal muscles. For instance, in muscular dystrophy, the distribution of the affected muscles depends on the disease process. Both the proximal and distal muscles are affected in statin-induced myopathy and critical illness myopathy. Myotonic dystrophy type 1,2 or distal myopathy may affect distal muscles more frequently than proximal muscles [13, 42, 43]. Second, the small number of patients may have been insufficient to generate significant features within the muscle location information.

Among the signals misclassified by the deep-learning model, several examples were inspected (S5 Fig). Signals that contained parts with high amplitudes were misclassified as neuropathy (S5A,D Fig), and signals that contained parts with small amplitudes were misclassified as myopathy (S5B,C Fig). From the feature visualization results of the trained deep-learning model, it was determined that the deep-learning model makes predictions based on the amplitude, duration, recruitment, and interference pattern of typical myopathy, neuropathy, and normal patients. Thus, the amplitudes of mis-predicted nEMG signals may dominate the recruitment and interference patterns which results in incorrect prediction by the deep-learning model.

Interestingly, the diagnostic accuracy of the physicians was 0.537, which was lower than expected. This low accuracy can be attributed to the two following reasons. First, our data included 38 patients out of 58 with neuromuscular disease, which is much higher than the real prevalence of approximately 200 per 100,000 people [44]. Second, the electrodiagnosis of physicians during this study was different from the real-world diagnosis process. Physicians considered nEMG signals as well as

additional information about the patient such as demographics or symptoms, which were absent in the labeling platform.

This study has several limitations. First, it used retrospective data from a single institution. When additional data from other institutions are available, external validation can be performed to further verify model performance. Second, a larger amount of nEMG data needs to be examined to demonstrate the stable performance of deep-learning in nEMG classification. The 58 patients from this study are not sufficient to demonstrate the usefulness of this deep-learning model and a larger cohort may elicit useful information from the muscle location as well. Third, the patient diagnosis labels were only divided into neuropathy, myopathy, and normal patients. There are diverse subtypes of neuromuscular diseases such as chronic inclusion body myositis, ongoing-state dermatomyositis, and late-stage muscular dystrophy, which co-exhibit MUAPs of myopathy and neuropathy, with short and long durations. Additional nEMG data for more specific neuromuscular diseases could improve the usefulness of deep-learning for nEMG clinical assistance. Finally, resting-state nEMG data may be added for better prediction. In some neuromuscular diseases, such as Pompe disease, EMG abnormalities may be revealed only in the resting state of the paraspinal muscles rather than the limb muscles [13, 45]. Future prospective studies with resting and volitional state EMG data could further improve the applicability of deep-learning for EMG electrodiagnosis.

Few studies have analyzed nEMG signals in a volitional state using deep-learning. The results of this study demonstrated that deep-learning could analyze nEMG signals in a short time and with high accuracy, and that our relatively simple model has the potential to be embedded in an nEMG device. Embedding a fast, accurate, and simple machine-learning model into an nEMG machine could allow clinical assistance within the device without sharing personal medical information. Such an application could not only reduce the burden on physicians, but also lead to a widely applicable, low-cost clinical decision system for use in small medical institutions.

In summary, our study presented significant potential for deep-learning to contribute to the automatic computer-aided electrodiagnosis system of patients with neuromuscular diseases.

# Supporting information
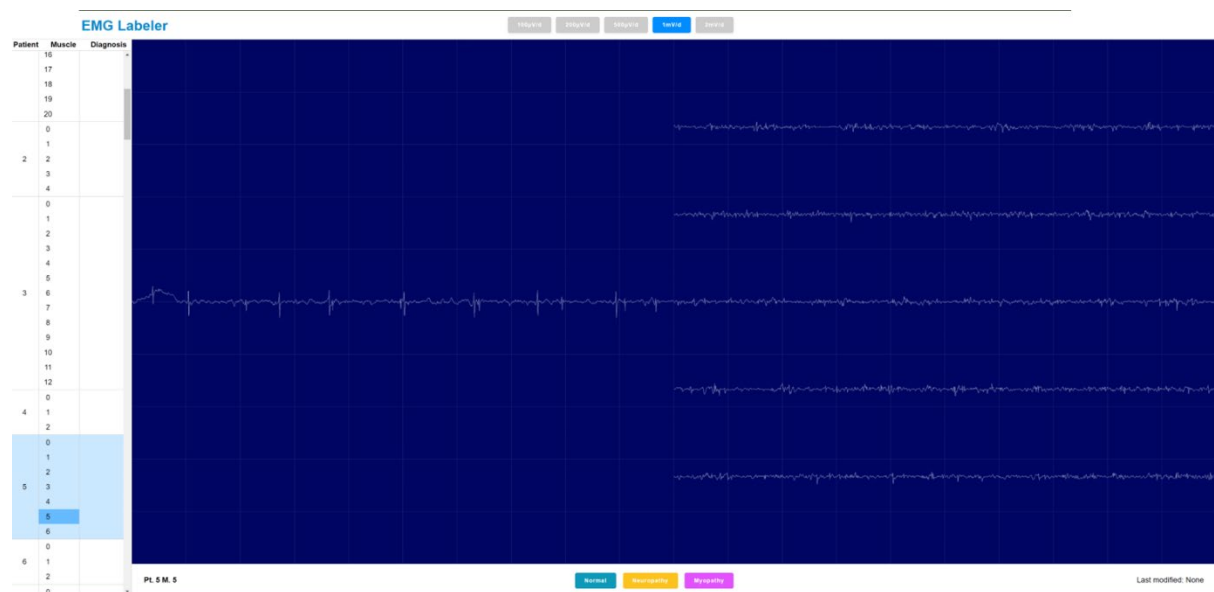


**S1 Fig. Electromyography labeler platform**

A web-based labeling platform named 'EMG Labeler'. nEMG signals were presented in a similar way to the signals from the EMG device

**S1 Table. Degree of agreement between physicians and the deep-learning model***

| | Physicians | Deep-learning model |
|---|---|---|
| Overall | 0.260[†] | 0.256[‡] |
| Myopathy | 0.360[†] | 0.400[‡] |
| Neuropathy | 0.256[†] | 0.247[‡] |
| Normal | 0.204[†] | 0.165[‡] |

* Data are expressed as Fleiss κ.

† Fleiss κ value between the results of the six physicians

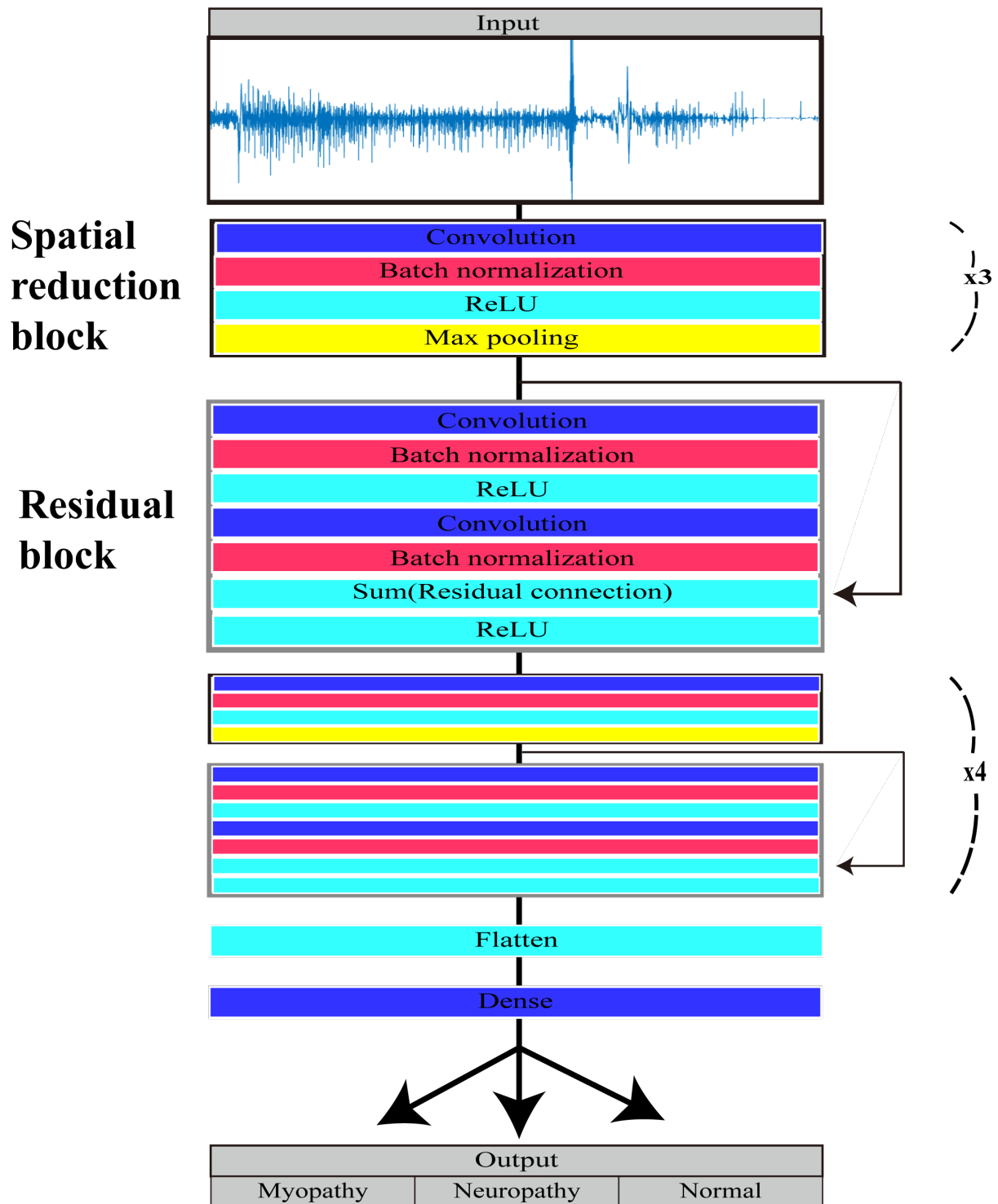‡ Fleiss κ value between the results of six physicians and the deep-learning model

**S2 Table. Classification results by the deep-learning model with and without muscle location information**

| | Performance results (95% CI) | |
| --- | --- | --- |
| | Without information | With information |
| Accuracy | 0.720 (0.682–0.758) | 0.700 (0.650–0.750) |
| Sensitivity (recall) | 0.715 (0.668–0.762) | 0.698 (0.652–0.744) |
| Specificity | 0.858 (0.836–0.879) | 0.848 (0.824–0.873) |
| Positive predictive value (PPV, precision) | 0.726 (0.695–0.757) | 0.703 (0.648–0.758) |
| F1 score | 0.715 (0-.673–0.757) | 0.694 (0.647–0.741) |

CI, Confidence interval

The average metrics of the classification results of the three deep-learning models are presented. The metrics, except for accuracy, are the average of the binary classification cores for each class from the one-versus-rest method.

## Input



**Spatial reduction block**

| Convolution |
| Batch normalization |
| ReLU |
| Max pooling |

x3

**Residual block**

| Convolution |
| Batch normalization |
| ReLU |
| Convolution |
| Batch normalization |
| Sum(Residual connection) |
| ReLU |

x4

| Flatten |

| Dense |

| Output | | |
| Myopathy | Neuropathy | Normal |

**S2 Fig. Structure of the deep-learning model**

There were seven spatial reduction blocks and five residual blocks with one and two convolutional layers, respectively
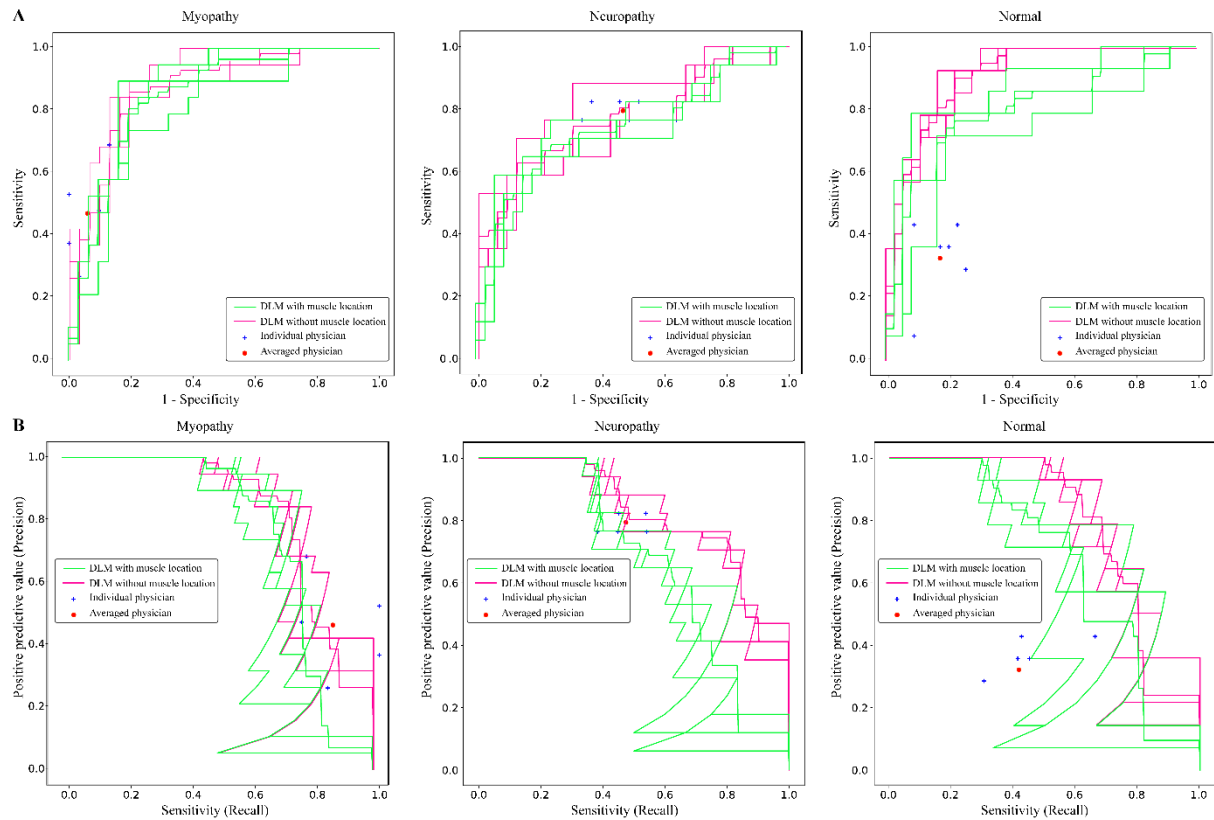
**S3 Fig. Per-class receiver operating characteristics and precision-recall curves of the machine learning model with and without the muscle location information**
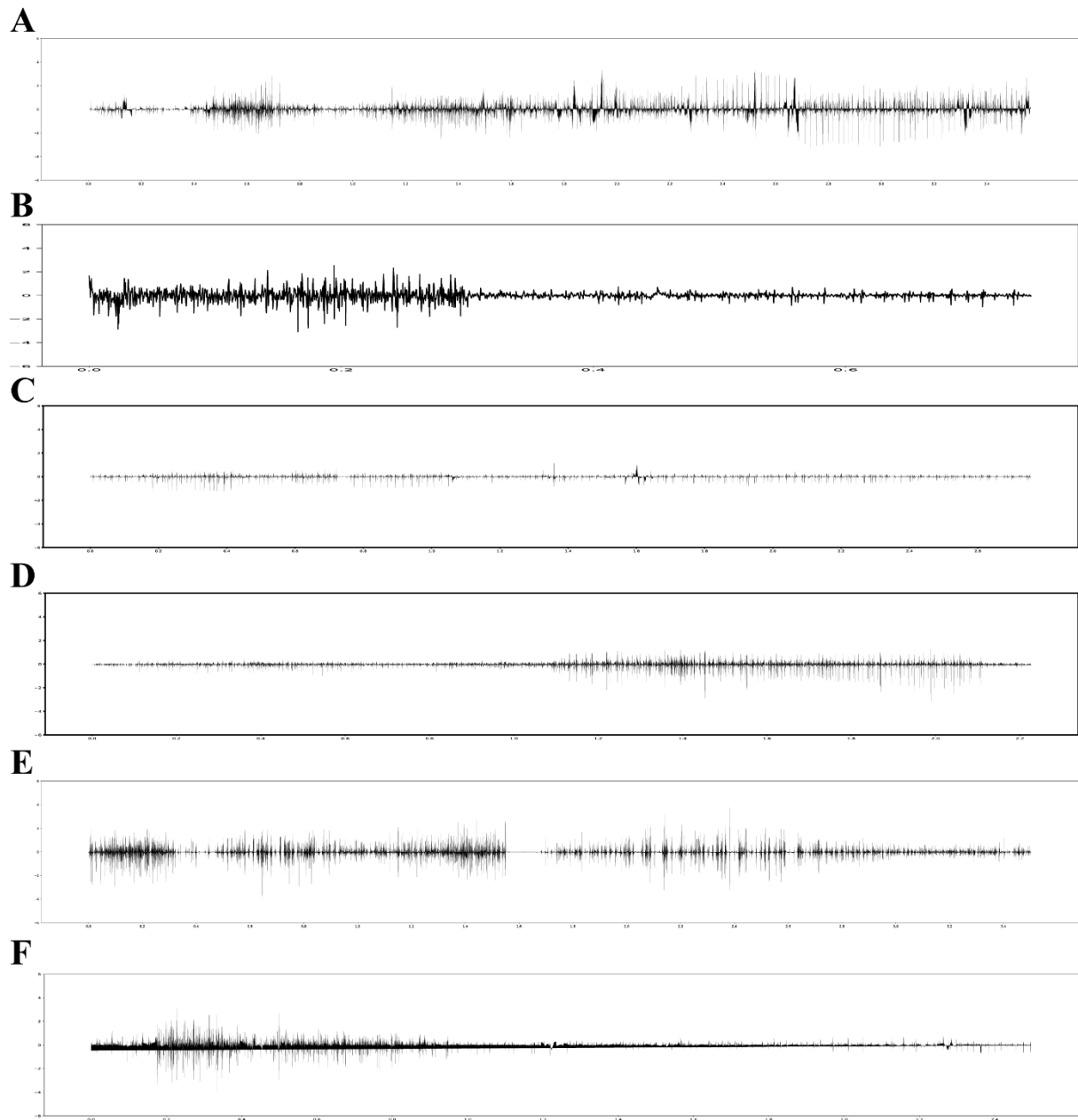
DLM, Deep-learning model

The areas under the receiver operating characteristic (ROC) curves (A) and precision-recall curves (B) were measured and depicted by dividing all data into myopathy, neuropathy, and normal.

The individual physician performance is annotated by the blue cross, and the average physician performance is annotated by the red dot.

ROC and precision-recall curves of the convolutional neural network (CNN) algorithm depending on whether the muscle location information was considered (green lines, CNN with muscle location) or not (pink lines, CNN without muscle location).
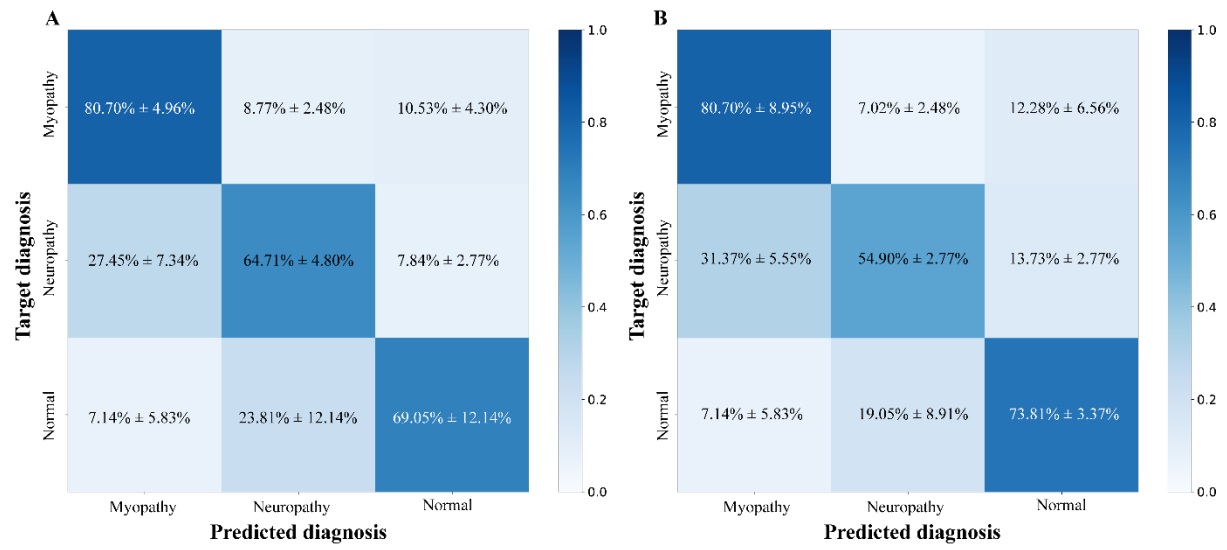
**S4 Fig. Confusion matrices of the machine learning model (A) without muscle location information and (B) with muscle location information**

**S5 Fig. Examples of electromyographic signals mispredicted by the deep-learning model**

Electromyographic signal mis-predicted: (A) normal to neuropathy; (B) normal to myopathy; (C) neuropathy to myopathy; (D) neuropathy to normal; (E) myopathy to neuropathy; (F) myopathy to normal.

# Author Contributions

## Abbreviations

nEMG, needle electromyography; CNN, convolutional neural network; CI, confidence interval; EMG, electromyography; ReLU, rectified linear unit; AUROC, area under the receiver operating characteristic curve; PPV, positive predictive value; ROC, receiver operating characteristic

# References

1. Daube JR, Rubin DI. Needle electromyography. Muscle Nerve. 2009;39(2):244–70. Epub 2009/01/16. doi: 10.1002/mus.21180. PubMed PMID: 19145648.
2. Kimura J. Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice: Oxford University Press. 2013.
3. Mills KR. The basics of electromyography. J Neurol Neurosurg Psychiatry 2005;76(suppl 2):ii32–5. doi: 10.1136/jnnp.2005.069211.
4. Oh SJ. Clinical Electromyography: Nerve Conduction Studies: Lippincott Williams & Wilkins. 2003.
5. Rubin DI. Needle electromyography: Basic concepts. Handb Clin Neurol 2019;160:243–56. Epub 2019/07/07 doi: 10.1016/B978-0-444-64032-1.00016-3. PubMed PMID: 31277852.
6. Whittaker RG. The fundamentals of electromyography. Pract Neurol 2012;12(3):187–94. Epub 2012/06/05 doi: 10.1136/practneurol-2011-000198. PubMed PMID: 22661353.
7. Aminoff MJ, Goodin DS, Parry GJ, Barbaro NM, Weinstein PR, Rosenblum ML. Electrophysiologic evaluation of lumbosacral radiculopathies: Electromyography, late responses, and somatosensory evoked potentials. Neurology. 1985;35(10):1514–8. Epub 1985/10/01 doi: 10.1212/wnl.35.10.1514. PubMed PMID: 2993952.
8. Bromberg MB. The motor unit and quantitative electromyography. Muscle Nerve. 2020;61(2):131–42. Epub 2019/10/04 doi: 10.1002/mus.26718. PubMed PMID: 31579956.
9. Gutiérrez Gutiérrez G, Barbosa López C, Navacerrada F, Miralles Martínez A. Use of electromyography in the diagnosis of inflammatory myopathies. Reumatol Clin (Engl Ed) 2012;8(4):195–200. doi: 10.1016/j.reumae.2011.10.004.
10. Leblhuber F, Reisecker F, Boehm-Jurkovic H, Witzmann A, Deisenhammer E. Diagnostic value of different electrophysiologic tests in cervical disk prolapse. Neurology. 1988;38(12):1879–81. doi: 10.1212/wnl.38.12.1879.
11. Sawada K, Horii M, Imoto D, Ozaki K, Toyama S, Saitoh E, et al. Usefulness of electromyography to predict future muscle weakness in clinically unaffected muscles of polio survivors. PM R 2020;12(7):692–8. Epub 2019/11/09 doi: 10.1002/pmrj.12281. PubMed PMID: 31702870.
12. Tonzola RF, Ackil AA, Shahani BT, Young RR. Usefulness of electrophysiological studies in the diagnosis of lumbosacral root disease. Ann Neurol 1981;9(3):305–8. Epub 1981/03/01 doi: 10.1002/ana.410090317. PubMed PMID: 6261675.
13. Paganoni S, Amato A. Electrodiagnostic evaluation of myopathies. Phys Med Rehabil Clin N Am 2013;24(1):193–207. Epub 2012/10/16. doi: 10.1016/j.pmr.2012.08.017. PubMed PMID: 23177039.
14. Meekins GD, So Y, Quan D. American Association of Neuromuscular & Electrodiagnostic Medicine evidenced-based review: Use of surface electromyography in the diagnosis and study of neuromuscular disorders. Muscle Nerve 2008;38(4):1219–24. doi: 10.1002/mus.21055. PubMed PMID: 18816611.
15. Merletti R, Farina D. Analysis of intramuscular electromyogram signals. Philos Trans A Math Phys Eng Sci 2009;367(1887):357–68. doi: doi:10.1098/rsta.2008.0235.
16. Haig AJ, Tong HC, Yamakawa KS, Quint DJ, Hoff JT, Chiodo A, et al. The sensitivity and specificity of electrodiagnostic testing for the clinical syndrome of lumbar spinal stenosis. Spine (Phila Pa 1976). 2005;30(23):2667–76. Epub 2005/12/02 doi: 10.1097/01.brs.0000188400.11490.5f. PubMed PMID: 16319753.
17. Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve 2006;34(2):238–41. Epub 2006/04/13 doi: 10.1002/mus.20554. PubMed PMID: 16609977.
18. Nirkko AC, Rösler KM, Hess CW. Sensitivity and specificity of needle electromyography: A prospective study comparing automated interference pattern analysis with single motor unit potential analysis. Electroencephalogr Clin Neurophysiol 1995;97(1):1–10. Epub 1995/02/01 doi: 10.1016/0924-980x(94)00248-6. PubMed PMID: 7533715.
19. Arthur KC, Calvo A, Price TR, Geiger JT, Chiò A, Traynor BJ. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun 2016;7(1):12408. doi: 10.1038/ncomms12408.
20. Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis: An update of recent literature. Curr Opin Neurol 2019;32(5):771–6. doi: 10.1097/WCO.0000000000000730. PubMed PMID: 31361627.
21. Parker MJS, Oldroyd A, Roberts ME, Ollier WE, New RP, Cooper RG, et al. Increasing incidence of adult idiopathic inflammatory myopathies in the City of Salford, UK: A 10-year epidemiological study. Rheumatol Adv Pract 2018;2(2):rky035. doi: 10.1093/rap/rky035.
22. Rose L, McKim D, Leasa D, Nonoyama M, Tandon A, Bai YQ, et al. Trends in incidence, prevalence, and mortality of neuromuscular disease in Ontario, Canada: A population-based retrospective cohort study (2003–2014). PLOS ONE 2019;14(3):e0210574. doi: 10.1371/journal.pone.0210574.

23.	Alfaras M, Soriano MC, Ortín S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Front Phys 2019;7(103). doi: 10.3389/fphy.2019.00103.

24.	Lu X, Wu Y, Yan R, Cao S, Wang K, Mou S, et al., editors. Pulse waveform analysis for pregnancy diagnosis based on machine learning. 2018 IEEE 3rd Adv. Inf. Technol. Electron Autom Control conference IAEAC; 2018.

25.	Gemein LAW, Schirrmeister RT, Chrabąszcz P, Wilson D, Boedecker J, Schulze-Bonhage A, et al. Machine-learning-based diagnostics of EEG pathology. Neuroimage. 2020;220:117021. doi: https://doi.org/10.1016/j.neuroimage.2020.117021.

26.	Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: A systematic review. J Neural Eng 2019;16(5):051001. doi: 10.1088/1741-2552/ab260c.

27.	Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. PLOS Med 2018;15(11):e1002699. doi: 10.1371/journal.pmed.1002699.

28.	Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25(1):65–9. doi: 10.1038/s41591-018-0268-3.

29.	Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Med 2018;15(11):e1002686. doi: 10.1371/journal.pmed.1002686.

30.	Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020;11(1):1760. doi: 10.1038/s41467-020-15432-4.

31.	Akef Khowailed I, Abotabl A. Neural muscle activation detection: A deep learning approach using surface electromyography. J Biomech 2019;95:109322 doi: https://doi.org/10.1016/j.jbiomech.2019.109322.

32.	Atzori M, Cognolato M, Müller H. Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. Front Neurorobot 2016;10:9–. doi: 10.3389/fnbot.2016.00009. PubMed PMID: 27656140.

33.	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):131–8. Epub 2019/04/30. doi: 10.4258/hir.2019.25.2.131. PubMed PMID: 31131148.

34.	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;130(5):617–23. Epub: 617–23. doi: 10.1016/j.clinph.2019.01.024. PubMed PMID: 30870796.

35.	Wei W, Dai Q, Wong Y, Hu Y, Kankanhalli M, Geng W. Surface-electromyography-based gesture recognition by multi-view deep learning. IEEE Trans Bio Med Eng 2019;66(10):2964–73. doi: 10.1109/TBME.2019.2899222.

36.	Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.

37.	He K, Zhang X, Ren S, Sun J, editors. 2016. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016(27–30, June).

38.	Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition: ArXiv 14091556. 2014.

39.	Prechelt L. 1998. Early stopping – But when? In: Neural Networks: Tricks of the Trade Orr GB, Müller K-R, editors: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 55–69.

40.	Olah C, Mordvintsev A, Schubert L. Feature visualization. Distill. 2017;2(11):e7. doi: 10.23915/distill.00007.

41.	Haig AJ, Gelblum JB, Rechtien JJ, Gitter AJ. Technology assessment: The use of surface EMG in the diagnosis and treatment of nerve and muscle disorders. Muscle Nerve. 1996;19(3):392–5. doi: 10.1002/(SICI)1097-4598(199603)19:3<392::AID-MUS21>3.0.CO;2-T. PubMed PMID: 8606710.

42.	Logigian EL, Ciafaloni E, Quinn LC, Dilek N, Pandya S, Moxley RT, 3rd, et al. Severity, type, and distribution of myotonic discharges are different in type 1 and type 2 myotonic dystrophy. Muscle Nerve. 2007;35(4):479–85. doi: 10.1002/mus.20722. PubMed PMID: 17230537.

43.	Dimachkie MM, Barohn RJ. Distal myopathies. Neurol Clin 2014;32(3):817–42, x. Epub 2014/05/15 doi: 10.1016/j.ncl.2014.04.004. PubMed PMID: 25037092.

44.	Carey IM, Banchoff E, Nirmalananthan N, Harris T, Dewilde S, Chaudhry UAR, et al. Prevalence and incidence of neuromuscular conditions in the UK between 2000 and 2019: A retrospective study using primary care data. PLOS One 2021;16(12):e0261983. doi: 10.1371/journal.pone.0261983.

45.	Hobson-Webb LD, Dearmey S, Kishnani PS. The clinical and electrodiagnostic characteristics of Pompe

disease with post-enzyme replacement therapy findings. Clin Neurophysiol 2011;122(11):2312–7. Epub :2312–7. doi: 10.1016/j.clinph.2011.04.016. PubMed PMID: 21570905.