**Deep learning model for needle electromyography**

**electro-diagnosis in comparison with physicians:**

**A retrospective study**

**Ilhan Yooa†, Jaesung Yoob†, Dongmin Kimc, Ina Yound, Hyodong Kima, Michelle Youna, Jun Hee Wone, Cho Woosupe, Youho Myonge, Kim sehoone, Ri Yue, Sung-Min Kimf, Kwangsoo Kimg, Seung-Bo Leeh\* and Keewon Kime\***

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, South Korea

†These Authors contributed equally to this work.

**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) is known as playing an important role in diagnosing the neuromuscular disorders with characteristic abnormalities of waveform. However, it has some limitations in that there are discrepancies of electro-diagnosis result among physicians, and the accuracy of electro-diagnosis of EMG quite relies on the experience of physicians. We investigated whether deep learning approach, a convolutional neural network (CNN) algorithm, can classify EMG waveforms to distinguish between neuropathy, myopathy, and normal as well as show better performance than physician’s electro-diagnosis results.

**Methods and findings**

We classified EMG data (58 patients; 382 muscles) stored in Seoul National University Hospital database from June 2015 to July 2020 among myopathy, neuropathy, normal by using the CNN algorithm.

Based on the classified results of the our method, the accuracy, sensitivity, specificity, positive predictive value and F1 score were 0.811, 0.720, 0.853, 0.725, and 0.718, respectively; mean values of the results electro-diagnosed by physicians were 0.537, 0.527, 0.770, 0.582, and 0.511, respectively. The performance of CNN based method for predicting myopathy, neuropathy, and normal was also evaluated with 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 the CNN based method could be contributable in interpreting EMG of patient with neuropathy or myopathy on behalf of physicians and assisting physician’s decision making in diagnosing patients with suspected neuromuscular disease. Large, prospective cohort studies with more diverse neuromuscular disease are needed in the future.

**Keywords**: Electromyography, Machine learning, Neuromuscular disease, Convolutional neural network**Author summary**

**Why was this study done?**

* Electromyography (EMG) has established as an important electro-physiologic test widely performed when diagnosing patients with neuromuscular disorder.
* Despite the importance of EMG in diagnosis, the accuracy of EMG readings is not yet high enough and there are often discrepancies among examiners, so more objective and accurate reading means are needed.
* Although EMG in volitional state plays important role in diagnosing neuromuscular disease, previous studies have mainly focused on analyzing gestures by using the EMG or classifying the EMG in a resting state. Therefore, previous studies are not sufficient for clinical application.

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

* The EMG data from 58 patients was classified with convolutional neural network (CNN) algorithm among either myopathy, neuropathy, or normal. The performance of CNN algorithm was compared with electro-diagnosed results by neurologist and rehabilitation medicine doctors.
* The accuracy of classifying EMG data by our CNN algorithm was 0.811, and the average accuracy of electro-diagnosis by physicians was 0.537.
* The classification by our CNN algorithm was more accurate and faster than electro-diagnosis by physicians.

**What do these findings mean?**

* Deep learning could electro-diagnose EMG on behalf of a physician and be used to assist in decision making in diagnosing patients with neuromuscular disorder.
* The performance of deep learning could be further enhanced by carrying out large-scale prospective study including patients with more variable specific kind of neuropathy and myopathy.

**Introduction**

Electromyography (EMG) is an electrophysiological test that records electrical activity generated from nerves, muscles, and neuromuscular junctions through a needle inserted into the muscle or surface electrode during resting and volitional state. [1-6] It is used to identify disorders of the peripheral nerves or muscles based on abnormalities in EMG signals that reflect the anatomical and physiological characteristics of peripheral nerves and muscles. [1-6] Among the EMG signals, the signal recorded during muscle contraction is called motor unit action potential (MUAP) and is shown only in a volitional state not in a resting state; it is essential in that physician can distinguish whether the subject has neuropathy or myopathy only by checking the MUAP that cannot be distinguished by EMG in a resting state. [1, 5-12] Additionally, when the neuromuscular disorder is in the early state or in a mild abnormal state, abnormal findings are found only in MUAP during the volitional state not during in the resting state. [13, 14] It has been known that the EMG signals seen when examining a subject with peripheral neuropathy commonly show characteristics of high amplitudes, long durations, and reduced recruitments, whereas the EMG signals seen when examining a patient with myopathy show characteristics of small amplitudes, short durations, and early recruitments. [1, 5-12] Needle EMG (nEMG), one type of EMG, is useful to detecting the action potentials in individual motor units and deep muscles. [15]

Although EMG plays an important role in diagnosing normal, neuropathy and myopathy, it has some limitations in that there are discrepancies among examiners, and the accuracy of EMG relies to a lot extent on proficiency of the examiner. Previous studies have reported that sensitivity of EMG in the diagnosis of neuropathy, myopathy, and normal is 47–83%, specificity is 73–81% and inter-rater reliability is 62–81%. [16-18] Additionally, to recognize abnormalities of EMG signals accurately, considerable time and efforts are needed. As the prevalence of neuropathy and myopathy continues to increase, the frequency of EMG for diagnosing it, the time it takes to interpret it, and the workload of the examiner are bound to increase. [19-22] A new approach may be helpful in clinically diagnosing neuropathy or myopathy through nEMG more efficiently and accurately in a shorter time.

Recently, deep learning has been used to analyzing big data in many field, and it is also applied to clinical data including waveform, time series data. [23, 24] Convolutional neural network, one kind of deep learning techniques, has applied to analyzing time series data and waveform data such as electrocardiography, electroencephalography. [23, 25, 26] Based on a result of the study on reading the results of electrocardiography and electroencephalography using deep learning, the accuracy was similar to or superior to that of medical students or residents, and detect nonobvious abnormalities easily overlooked. [27] In previous studies, machine learning has been mainly used to analyze the spontaneous potential that occurs in the resting state of the nEMG or to analyze the gesture using the surface EMG that are not related to the diagnosis of neuromuscular disorders. [28-32] However, as mentioned above, it is needed to analyze the MUAP of nEMG recorded during the volitional state to differentiate between myopathy, neuropathy and normal, or to identify neuromuscular disorder in early state or mild abnormal state. To the best of our knowledge, few studies have been reported using machine learning to analyze nEMG signals during volitional state.

To overcome the diagnostic limitations of EMG and investigate the feasibility of machine learning to electro-physiologically diagnose nEMG signals in volitional state on behalf of a physician and help decision-making in diagnosis of patients, we developed a deep learning model, which are known to show good performance in image analysis. [33, 34] The development of deep learning-based nEMG analysis could lead to the development of faster and more accurate automated nEMG interpretation. We retrospectively reviewed nEMG data examinated in subjects with peripheral neuropathy or myopathy or normal subjects, analyzed those by using convolutional neural network (CNN) algorithm, and compared the classification results of nEMG signals with electro-diagnosis results by 6 physicians.

**Methods**

**Study design and preparation**

In this study, nEMG signal data of 58 subjects who visited Seoul national university hospital from June 2015 to July 2020 were used for analysis by dividing them into peripheral neuropathy, myopathy, and 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 conducted according to the Declaration of Helsinki plus its later amendments. Informed consent was not necessary because this study is retrospective analysis and all nEMG signal data was anonymized before analysis.

nEMG was performed with a Nicolet EDX EMG system and monopolar needle electrode from the subject’s muscles. The filter setting was set at 20 Hz (low-cut) and 10 kHz (high-cut). The results of the last 10 seconds of the nEMG were recorded and used for analysis.

Certified neurologist and rehabilitative physician reviewed the nEMG data and confirmed the diagnosis of all subjects. The nEMG data were extracted from the numerical data stored in the electromyography machine and then transformed as waveform data throuth MATLAB software (version R2020b). Among the created waveform data, artifacts which occurred in the cases including move of the needle electrode or patients moving, among the data at the beginning and at the end were excluded, and all of the noise in the middle portion was preserved. The original nEMG data sampled at 48 kHz was downsampled to 10 kHz to reduce complexity, and partitioned into fixed window lengths of 0.4 second and hop sizes of 0.1 second, which were likely to be the optimal length for analysis. Dataset consisted of different numbers of muscle nEMG data because the number of muscles tested was different for each subject. After slicing, total segments were composed of 2700, 3664, and 1706 segments extracted from subjects with myopathy, neuropathy, and normal subjects, respectively. Based on the elbow joint of the upper extremity and the knee joint of the lower extremity, muscles close to the trunk were classified as proximal and distant muscles as distal muscles.

**Classification by physicians**

After de-identifying the number of patient identification as a random number, the transformed nEMG waveform data similar to the actual one displayed on the screen of the nEMG machine were stored in web-based labeling platform so that residents belonging to different institutions could participate; then, it was provided to 2 residents of neurology and 4 residents of rehabilitation medicine, so that the nEMG results were electro-diagnosed as one of myopathy, neuropathy, and normal. (S1 Fig.)

The six physicians electro-diagnosed EMG signal data without any clinical information such as symptom or age of the subject. When the physician pressed the randomly assigned number of the subject, the EMG waveform was simultaneously played with sound and showed both real-time waveform data and waveform data stacked for 500 microseconds; physicians were allowed to be able to change the amplitude of wave not just 100, 200, and 500 microvolts, but also 1 and 2 millivolts. Physicians first annotated the muscles, and then diagnosed the subjects by considering the results of the muscles annotation. After the physician completes diagnosis, the electro-diagnosis results were stored within the platform as well as aggregated, and compared with the confirmed diagnosis.

**Classification by CNN algorithm**

CNN comprised of 7 spatial reduction blocks and 5 residual blocks with 1 and 2 convolutional layers, respectively. (S2 Fig.) Spatial block and residual block consisted of convolutional layers, batch normalization, rectified linear unit (ReLU), residual connection, and max pooling. Hyper-parameters were determined empirically. Learning rate, batch size, and epoch was set to 10-3, 32, and 100, respectively.

Current CNN was used to sequentially classify the subject in 2 stages; first, it received the nEMG numerical data of each muscle tested for each subject as an input and elicited one of myopathy, neuropathy, and normal as an output. Then, the final output was presented as one of myopathy, neuropathy, and normal by considering all the probability values belonging to myopathy, neuropathy, and normal of the tested muscles to the subject. Classification was conducted by inputting only nEMG data without muscle location information, and additionally performed by inputting nEMG data and muscle location information together.

**Assessing the performance of CNN algorithm**

The performance of CNN algorithm was evaluated with the accuracy, F1 score, area under receiver operating characteristic curve (AUROC), positive predictive value (PPV; precision), sensitivity (recall) and specificity. Since the number of subjects was small, the accuracy of this algorithm was calculated by cross entropy with 5-fold cross-validation. Based on the results of accuracy, and F1 score as well as PPV, sensitivity, and specificity, we compared the result classified by CNN algorithm, which not considering muscle location information not only with averaged result by physicians but also with CNN algorithm considering muscle location information. The degree of agreement was also calculated between physicians and that between physicians and CNN algorithm.

**Statistical analysis**

Statistical analyses were performed using R statistical software (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria) and Python 3. The differences among the groups for categorical variables were assessed using the Fisher’s exact test or Pearson’s χ2 test and those for continuous variables were assessed using the Kruskal–Wallis test or one-way analysis of variance test. Data are expressed as means ± standard deviation for continuous variables and number (%) for categorical variables. A *p* value less than 0.05 was regarded as statistically significant. For assessment of algorithm, an ROC (receiver operating characteristic) analysis was used with one versus other method, sensitivity plus specificity were measured with binary decision for each label, and PPV plus recall were calculated and are depicted with PPV-recall curve. Inter-rater reliability was analyzed and is expressed with value of Fleiss kappa.

**Results**

The data of the subjects used for the analysis were 20 normal subjects and 19 subjects with neuropathy with whom the diagnosis was radiculopathy, motor axonal polyneuropathy, and motor neuron disease, etc., myopathy was 19 subjects with whom the diagnosis was muscular dystrophy and inflammatory myopathy. The number of nEMG data used for analysis was 124, 161, and 97 for myopathy, neuropathy, and normal, respectively, and the rest of the results are shown in Table 1.

Table 1. Demographic characteristics of subjects and their needle electromyography data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Myopathy | Neuropathy | Normal | p-value |
| Number of Subjects | 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 |
| Proportion of nEMG 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 |  |

The performance of classification by CNN algorithm was compared with results of electro-diagnosis by physicians using the accuracy, sensitivity, specificity, PPV, and F1 score (excluding the missing values of 8 subjects and 10 muscles). Those values of the CNN algorithm were 0.811, 0.720, 0.853, 0.725, and 0.718, respectively; the counterparts of physicians were 0.537, 0.527, 0.770, 0.582, and 0.511, respectively. (Table 2) The inter-rater reliability, expressed in Fleiss κ, between electro-diagnosis results by physicians were 0.26; that between electro-diagnosis results by physicians and classification results by CNN algorithm were 0.26. (Table 2)

Table 2. The results of electro-diagnosis by physicians and those of classification by convolutional neural network algorithm

|  |  |  |
| --- | --- | --- |
|  | Results | |
|  | Physicians | Convolutional neural network |
| Accuracy | 0.537\* | 0.811† |
| Sensitivity (recall) | 0.527\* | 0.720† |
| Specificity | 0.770\* | 0.853† |
| Precision | 0.582\* | 0.725† |
| F1 score | 0.511\* | 0.718† |
| Inter-rater reliability (Fleiss κ) |  | |
| Overall | 0.26‡ | 0.26§ |
| Myopathy | 0.36‡ | 0.40§ |
| Neuropathy | 0.26‡ | 0.25§ |
| Normal | 0.20‡ | 0.17§ |

\* Average value of results of electro-diagnosis by 6 physicians.

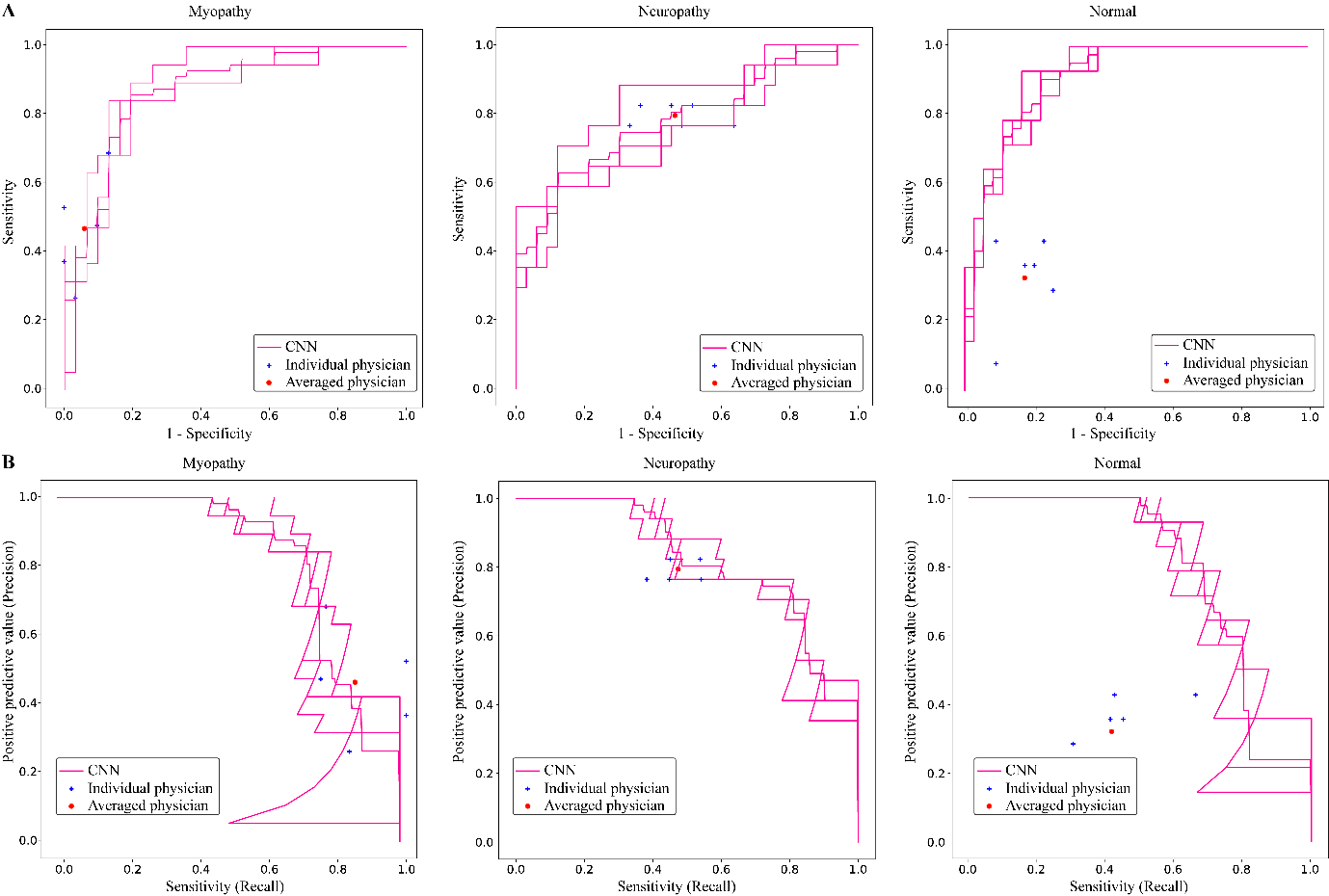
† Weighted average value of classification results considering different number of subjects for each myopathy, neuropathy, and normal group.

‡ Fleiss κ value between the results by 6 physicians.

§ Fleiss κ value between the results by 6 physicians and those of the convolutional neural network algorithm.

ROC curves and precision-recall curves were calculated and depicted according to myopathy, neuropathy, and normal; for comparison, individual physician performance and averaged physician performance were calculated and depicted on the same figure. Based on the results showing on figure, CNN algorithm is likely to exceed the averaged physicians’ performance for all group. (Fig. 1) The AUROCs calculated by dividing all data by myopathy, neuropathy, and normal 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.

Figure 1. ROC and precision-recall curves.



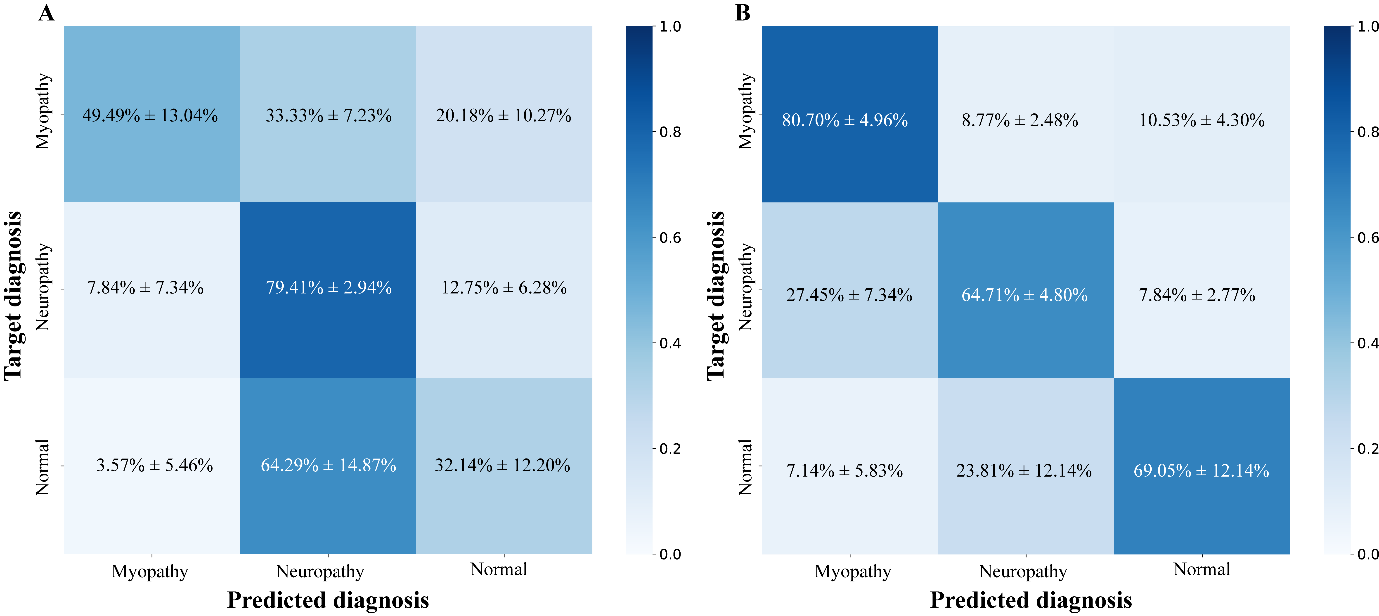
CNN = Convolutional neural network

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

Individual physician performance is annotated by the blue colored cross and averaged physician performance is annotated by the red dot.

The prediction accuracies of this CNN algorithm and physicians were calculated by dividing by group after excluding missing values (n=8 for subjects, n=10 for muscle signals), and those are depicted as confusion matrices divided into each group. The prediction accuracy of CNN algorithm and physicians divided by group are as follows; the former and latter results for myopathy, neuropathy, and normal were 80.70%±4.96%, 64.71%±4.80%, 69.05%±12.14% and 46.49%±13.04%, 79.41%±2.94%, 32.14%±12.20%, respectively. (Fig. 2)

Figure 2. Confusion matrices showing the accuracy of prediction by current convolutional neural network algorithm and that by physicians



The accuracies of this convolutional neural network and those averaged of 6 physicians were calculated except missing values (n=8 subjects)

A. Averaged accuracy of physician electro-diagnosis divided into each diagnosis

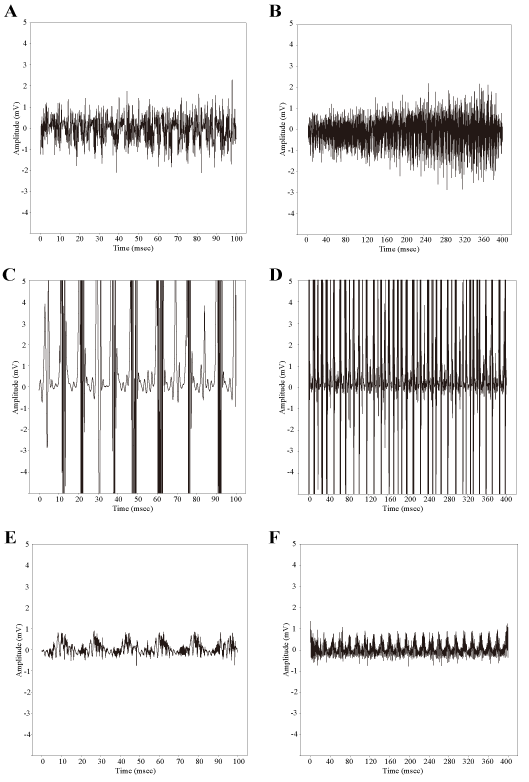
B. Accuracy of current convolutional neural network algorithm divided into each diagnosis

The performance of CNN algorithm with and without muscle location information was compared based on accuracy, sensitivity, specificity, PPV, and F1 score (excluding the missing values of 8 subjects and 10 muscles). Those values of the CNN algorithm that did not consider muscle location information were 0.811, 0.720, 0.853, 0.725, and 0.718, respectively; the counterparts of those that consider muscle location information were 0.798, 0.700, 0.845, 0.704, and 0.695, respectively. (S1 Table) The AUROCs for myopathy, neuropathy, and normal of CNN with muscle location information as well as those without muscle location information 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) as well as 0.874 (95% CI 0.858–0.889), 0.781 (95% CI 0.723–0.839), and 0.847 (95% CI 0.836–0.858), respectively. (S3 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 were 80.70%±8.95%, 54.90%±2.77%, 73.81%±3.37%, and 80.70%±4.96%, 64.71%±4.80%, 69.05%±12.14%, respectively. (S4 Fig.)

The cases which CNN algorithm predicted wrongly were analyzed, and among them, there were cases that CNN algorithm predicted incorrectly but more than two-thirds of the physicians predicted correctly as well as cases that both the physicians and CNN algorithm predicted incorrectly. Among the former, cases in which normal was mispredicted as neuropathy accounted for the most (50%), and those in which normal was mispredicted as myopathy as well as those in which neuropathy was mispredicted as myopathy accounted for 25%, respectively; among the latter, those in which myopathy was mispredicted as normal or neuropathy plus those in which neuropathy was mispredicted as myopathy or normal accounted for 42.86%, respectively, and those where neuropathy was mispredicted as myopathy or normal accounted for 14.28%. The examples of aforementioned waveform are shown in S5 Figure.

The waveform created based on the learned features through CNN algorithm was similar to the typical waveforms of myopathy, neuropathy, and normal. The waveform of myopathy showed small amplitude as well as short duration and counterpart of neuropathy showed high amplitude as well as long duration. (Fig. 3)

Figure 3. Waveforms created by convolutional neural network algorithm after learning



A and B, myopathy; C and D, neuropathy; E and F, normal.

Note that A, C, E were plotted with 10 milli second of x-axis interval and B, D, F were plotted with 40milli second of y-axis interval to show the amplitude and duration of individual waves as well as the recruitment and interference pattern of overall waveform.

**Discussion**

The aim of the present study was to evaluate the accuracy of classifying the nEMG waveform data using machine learning, and to demonstrate if it can support physician’s decision to enable more accurate and efficient diagnosis. For that purpose, deep learning was applied to classifying the nEMG waveforms, assessed the performance; additionally, the classified results were compared with electrophysiological diagnosis by physicians. Based on the classified results by our CNN algorithm, the accuracy was superior to accuracy of the physician’s diagnosis.

Previously, there have been reports that machine learning showed good performance when applied to image analysis, surface EMG, and needle nEMG. [28-34] Previous studies that analyzed nEMG data as 2 dimensional data using machine learning were studies to analyze gestures using surface nEMG or signals during resting state using needle nEMG. [28-32] It has been well known that needle nEMG is useful in the diagnosis of neuromuscular disorders; and has been also suggested that nEMG signals during a resting state plus a volitional state should be considered in order to distinguish myopathy, neuropathy, normal as well as to diagnose neuromuscular disorders in the initial state or minimally involved state. [1-6, 8, 13, 14, 35]. However, there have been few studies that classify nEMG data during a volitional state using deep learning.

We analyzed nEMG data in volitional state, which is important for electrophysiological diagnosis of neuropathy and myopathy, and confirmed that it showed better performance than physicians. Moreover, in order to minimize data loss, nEMG data, which is one-dimensional numerical data, was used after minimal noise was removed from the beginning and end of the nEMG data. To confirm the clinical applicability of CNN algorithm, the diagnostic accuracy of physicians was measured and compared with that of CNN algorithm. Finally, we found that the accuracy and time-taken of diagnosing neuropathy, myopathy, and normal were 0.811 and 40 seconds in using only nEMG data by CNN algorithm, which is better and much shorter than that of physicians.

In order to finally diagnose a patient, the nEMG results of all tested muscles should be considered altogether. However, the number of muscles tested may be slightly different for each patient, and abnormalities may not be found in all muscles, but only in some. To consider these points, a newly devised method was used to determine final diagnosis of subject; it was that the data of each muscle were analyzed individually, the probability value for each label was divided as a result of the analysis by the number of muscles and averaged, and finally the label with the highest probability value was determined as the diagnosis. Additionally, considering that peripheral neuropathy mainly shows abnormalities in the distal part muscle, whereas, myopathy mainly shows abnormalities in the proximal part muscles, additional information about muscle location, which means whether the muscles are located close to the trunk or not, was added CNN algorithm, and contributed to improved accuracy.

Interestingly, the diagnostic accuracy of physicians was lower than expected at 54%, which is thought to be due to 2 main reasons; first, in the data used in this study, the proportion of peripheral neuropathy and myopathy is out of distribution, which is much higher than the prevalence in population. Secondly, it is thought that the pre-test probability of diagnosing only with the nEMG data without clinical information such as the patient’s age and symptoms, as in clinical practice, may have worked. When the muscle location information was added, the accuracy and AUROC for diagnosing the neuropathy improved slightly, but performance of diagnosing the myopathy and normal were not improved significantly, which is thought to be due to following reasons; first, the normal group contained significantly more distal muscles than proximal ones. A bias may have occurred since the MUAP recorded in the EMG are affected by the muscle fiber density, the distance between needle tip and the muscle fibers as well as muscle fiber diameter.[36] Secondly, myopathy usually involves more frequently in proximal muscle than in distal ones, but in some cases this may not. For instance, in muscular dystrophy, the distribution of the affected muscle is somewhat different depending on the disease process; statin induced myopathy and critical illness myopathy may involve both proximal and distal muscles, and myotonic dystrophy type 1,2 or distal myopathy may involve more frequently in distal muscles than in proximal muscles.[13, 37, 38] The characteristic of misclassified waveform is that when parts with high amplitude are partially mixed, it is misclassified as neuropathy, and when parts with small amplitudes are partially mixed, it tends to be misclassified as myopathy; this is thought to be because the recruitment and interference patterns are relatively less reflected than the amplitude or duration of the nEMG waveform. Additionally, the waveforms generated based on the learning results of CNN algorithm showed not only amplitude, duration but also recruitment, interference patterns seen in nEMG waveform of typical myopathy, neuropathy, and normal subjects.

However, current study also has some limitations. First, this study deal with retrospective data from only one center study. Secondly, study number is not enough to demonstrate perfect usefulness of deep learning on nEMG classification. Finally, we focused only on data crudely divided neuromuscular disorder into neuropathy, myopathy, and normal. In the case of myopathy, the characteristic waveform shows short duration and small amplitude of MUAP. However, several diseases, which neurogenic process are accompanied, including chronic inclusion body myositis, ongoing-state dermatomyositis, and late-stage muscular dystrophy, co-exhibit MUAP of short duration and long duration; which involve only axial muscles such as Pompe disease, EMG abnormalities may be revealed only in the resting state of the paraspinal muscles rather than the limb muscles. [13, 39] Future prospective study with data from multicenter consisting of more variable disease and taking altogether of resting and volitional state will further improve the performance of deep learning to nEMG classification.

Until now, few studies on analyzing nEMG data of volitional state by deep learning have been documented. Our study suggests that machine learning with short analysis time and relatively simple algorithm has the potential to be embedded in nEMG machines, which will lead to reducing errors in nEMG electro-diagnosis and the workload of physicians; and also contribute to preventing the leakage of personal medial information that could occur if nEMG data is uploaded to database separate from the EMG machine. Eventually, it will shed lights on automatically electro-diagnosing patient with neuropathy or myopathy via machine learning.

In summary, it is concluded that deep learning may play a significant role in the electro-diagnosis of patients with neuromuscular disorder.

**Supporting information**

S1 Figure. An example of nEMG waveform data shown on EMG Labeler



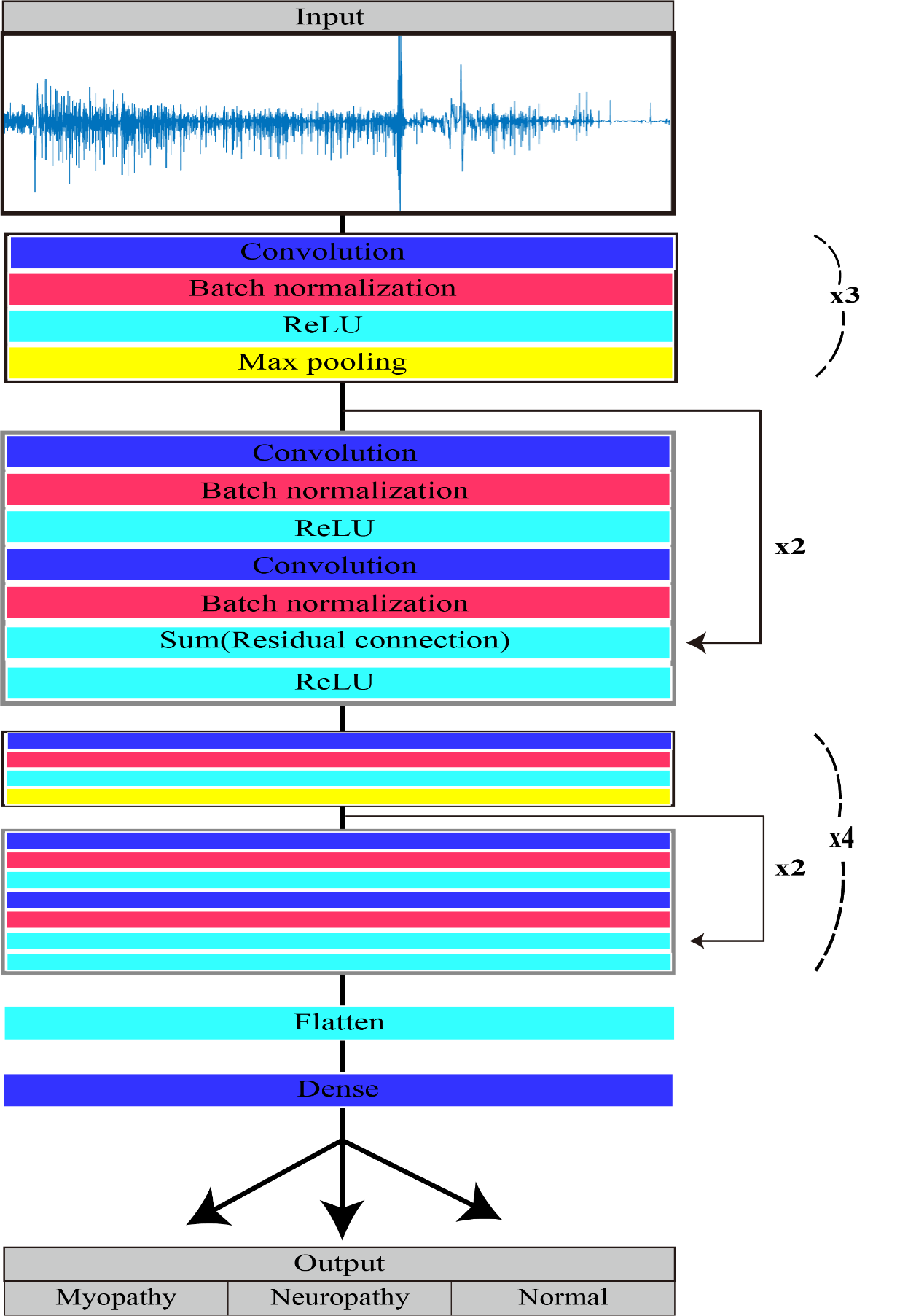
A web-based labeling platform named ‘EMG Labeler’ for reproducing nEMG numerical data as a waveform and displaying it to the physicians and storing diagnosis by the physician.

S1 Table. Comparison of performance difference of the convolutional neural network depending on whether muscle location information was considered or not

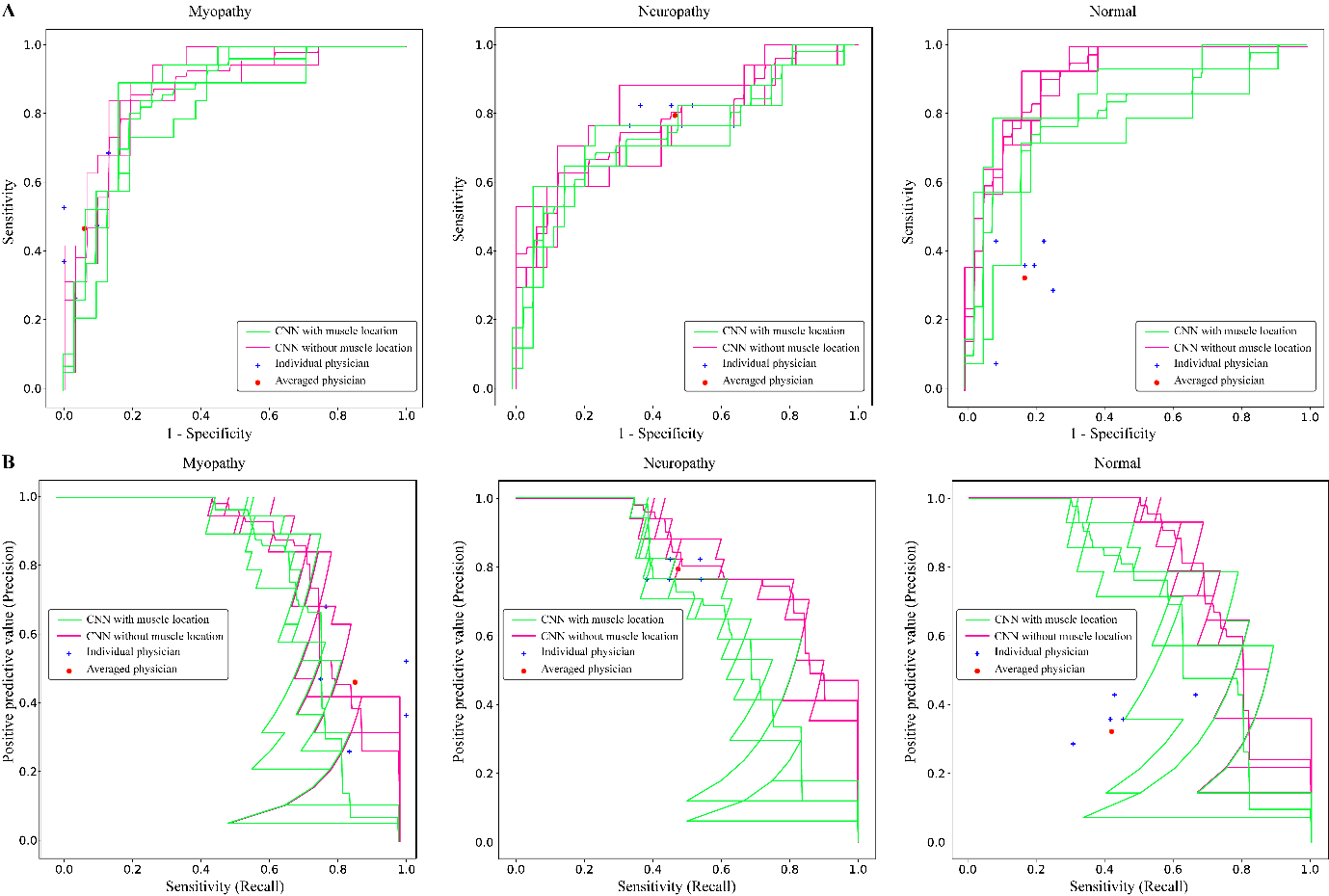
|  |  |  |
| --- | --- | --- |
|  | Results\* | |
|  | Without information | With information |
| Accuracy | 0.811 | 0.798 |
| Sensitivity (recall) | 0.720 | 0.700 |
| Specificity | 0.853 | 0.845 |
| PPV (precision) | 0.725 | 0.704 |
| F1 score | 0.718 | 0.695 |

\* Weighted average value of classification results considering different number of subjects for each myopathy, neuropathy, and normal group.

S2 Figure. Structure of the CNN algorithm. 7 spatial reduction blocks and 5 residual blocks with 1 and 2 convolutional layers, respectively



S3 Figure. ROC and precision-recall curves of the convolutional neural network algorithm depending on whether the muscle location information was considered or not

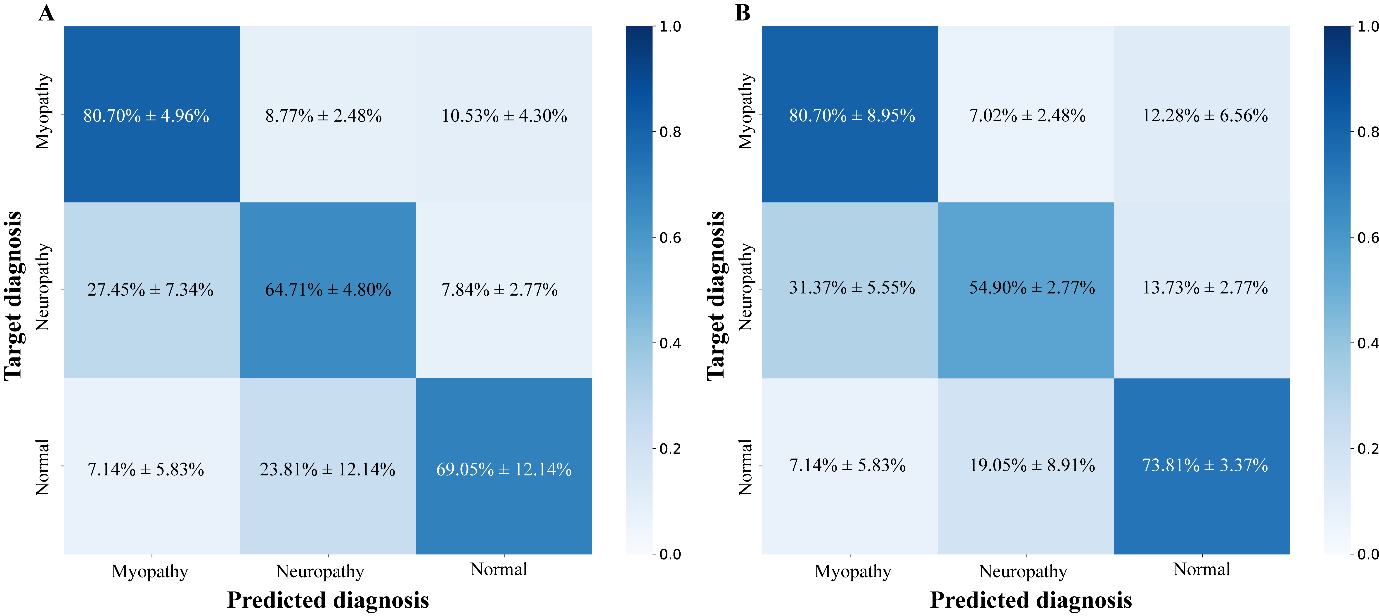


CNN = convolutional neural network

The area under receiver operating characteristic curve (A) and the precision-recall curve (B) was calculated and depicted by dividing all data into myopathy, neuropathy, and normal.

ROC and precision-recall curves of the convolutional neural network algorithm depending on whether the muscle location information was considered (green lines, CNN with muscle location) or not (pink lines, CNN without muscle location).

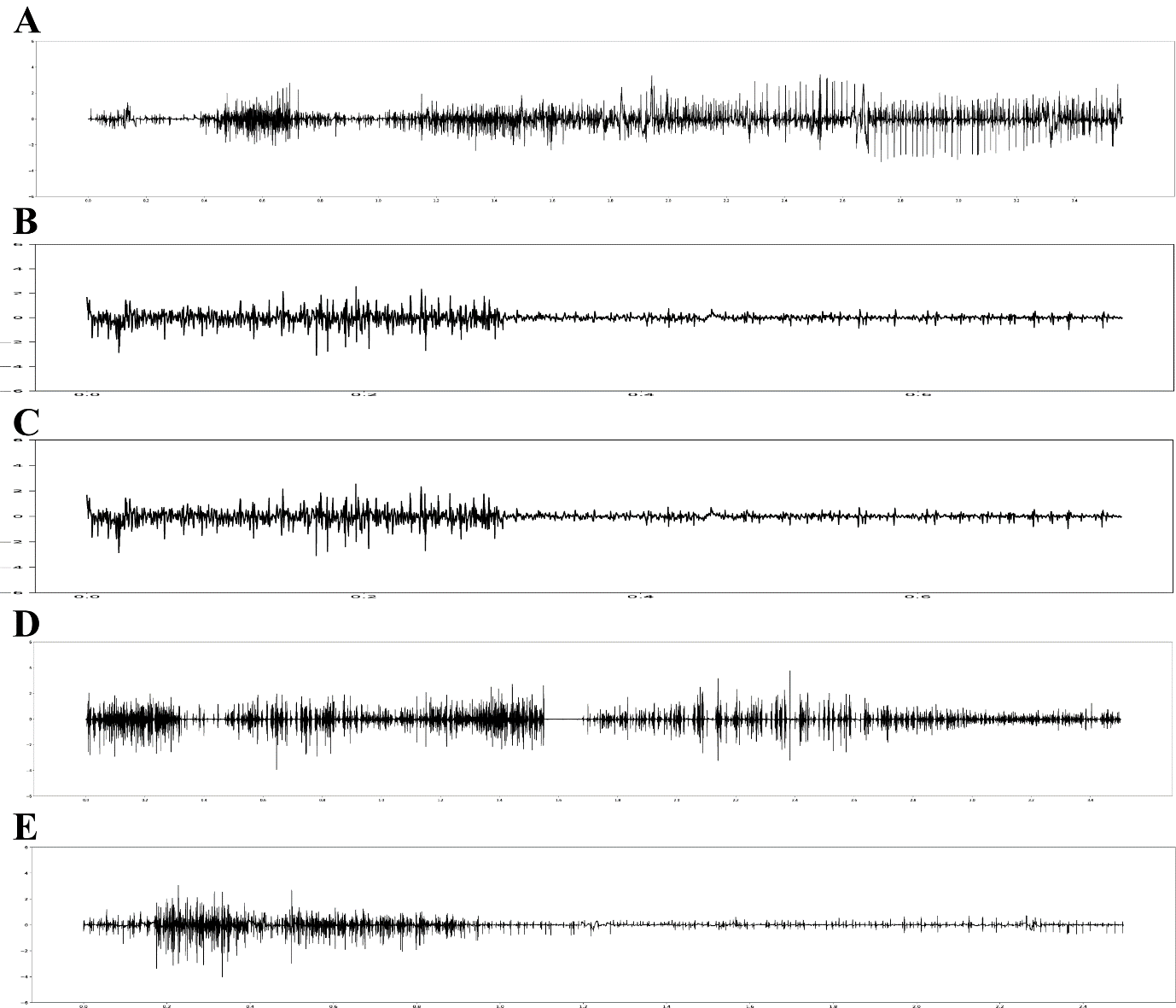
S4 Figure. Confusion matrices showing the accuracy of prediction by the convolutional neural network algorithm depending on whether the muscle location information was considered or not



The accuracies of convolutional neural network were calculated except missing values. (n=8 subjects)

Accuracy of convolutional neural network algorithm divided into each diagnosis depending on whether the muscle location information was considered (B) or not (A).

S5 Figure. Examples of electromyography waveforms mispredicted by convolutional neural network



A: the waveform that incorrectly predicted normal as neuropathy

B: the waveform that incorrectly predicted normal as myopathy

C: the waveform that incorrectly predicted neuropathy as myopathy

D: the waveform that incorrectly predicted myopathy as neuropathy

E: the waveform that incorrectly predicted myopathy as normal

**Author Contributions**

**Conceptualization**: Ilhan Yoo, Sung-Min Kim, Keewon Kim.

**Data curation**: Ilhan Yoo, Ina Youn, Hyodong Kim, Michelle Youn, Jun Hee Won, Cho Woosup, Youho Myong, Kim sehoon, Keewon Kim.

**Formal analysis**: Jaesung Yoo, Seung-Bo Lee.

**Funding acquisition**: no.

**Investigation**: Ilhan Yoo, Jaesung Yoo, Seung-Bo Lee, Keewon Kim.

**Methodology**: Kwangsoo Kim, Ri Yu, Seung-Bo Lee, Jaesung Yoo.

**Project administration**: Keewon Kim.

**Resources**: Keewon Kim.

**Software**: Jaesung Yoo, Dongmin Kim.

**Supervision**: Keewon Kim, Seung-Bo Lee.

**Validation**: Ilhan Yoo, Jaesung Yoo, Keewon Kim, Seung-Bo Lee.

**Visualization**: Ilhan Yoo, Jaesung Yoo.

**Writing – original draft**: Ilhan Yoo.

**Writing – review & editing**: Jaesung Yoo, Keewon Kim, Seung-Bo Lee,

**Abbreviations**

needle electromyography, nEMG; convolutional neural network, CNN; confidence interval, CI; electromyography, EMG; rectified linear unit, ReLU; area under receiver operating characteristic curve, AUROC; positive predictive value, PPV; 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. Journal of Neurology, Neurosurgery & Psychiatry. 2005;76(suppl 2):ii32-ii5. 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. Gerardo Gutiérrez Gutiérrez CBLFNAMM. Use of Electromyography in the Diagnosis of Inflammatory Myopathies. Reumatología Clínica (English Edition). 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-. 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. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 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. Nature Communications. 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). 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. Frontiers in Physics. 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 Advanced Information Technology, Electronic and Automation Control Conference (IAEAC); 2018 12-14 Oct. 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. 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. Nature Communications. 2020;11(1):1760. doi: 10.1038/s41467-020-15432-4.

28. 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>.

29. 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.

30. Nam S, Sohn MK, Kim HA, Kong H-J, Jung I-Y. 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.

31. 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 20190223. doi: 10.1016/j.clinph.2019.01.024. PubMed PMID: 30870796.

32. Wei W, Dai Q, Wong Y, Hu Y, Kankanhalli M, Geng W. Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning. IEEE Trans Biomed Eng. 2019;66(10):2964-73. doi: 10.1109/TBME.2019.2899222.

33. He K, Zhang X, Ren S, Sun J, editors. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 27-30 June 2016.

34. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 14091556. 2014.

35. 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.

36. Daube JR. AAEM minimonograph #11: Needle examination in clinical electromyography. Muscle Nerve. 1991;14(8):685-700. doi: 10.1002/mus.880140802. PubMed PMID: 1890993.

37. 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.

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

39. 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 20110513. doi: 10.1016/j.clinph.2011.04.016. PubMed PMID: 21570905.