**Convolutional neural network for needle-electromyography 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 Office of Hospital Information, Seoul National University Hospital, Seoul, Republic of 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

Seoul National University Hospital, Office of Hospital Information, 101, Daehak-ro, Jongno-gu, Seoul, Republic of Korea

Tel: +82-2-2072-

Fax: +82-2-

E-mail: koreateam23@korea.ac.kr

**Abstract**

**Background**

Electromyography

**Methods and findings**

We used

**Conclusions**

In this study, we found that

**Author summary**

**Why was this study done?**

Diagnosing neuropathy is difficult job

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

We found the

**What do these findings mean?**

The findings of this study

**Introduction**

Needle Electromyography (nEMG) is a type of electromyography, 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 nEMG signals that reflect the anatomical and physiological characteristics of peripheral nerves and muscles. [1-6] Among the nEMG signals, the signal recorded during muscle contraction is called motor unit action potentials (MUAPs); Through this, it is possible to determine whether the subject is has neuropathy or myopathy or not. It has been known that the nEMG signals seen when examining a subject with peripheral neuropathy commonly show characteristics of large amplitudes, long durations, and reduced recruitments, whereas the nEMG signals seen when examining a patient with myopathy show characteristics of small amplitudes, short durations, and early recruitments. These differences in nEMG signals have been reported as important and useful information when diagnosing peripheral neuropathy and myopathy in previous studies. [1, 5-12]

Although nEMG 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 nEMG relies to a lot extent on proficiency of the examiner. Previous studies have reported that sensitivity of nEMG in the diagnosis of neuropathy, myopathy, and normal is 47–83%, specificity is 73–81% and inter-rater reliability is 62–81%. [13-15] Additionally, to recognize abnormalities of nEMG signals accurately, considerable time and efforts are needed. As the prevalence of neuropathy and myopathy continues to increase, the frequency of nEMG for diagnosing it, the time it takes to interpret it, and the workload of the examiner are bound to increase. [16-19] 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. [20, 21] Convolutional neural network, one kind of deep learning techniques, has applied to analyzing time series data and waveform data such as electrocardiography, electroencephalography. [20, 22, 23] As a result of the study of 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. [24] Previous studies that analyzed nEMG signals using machine learning were mostly those that analyzed surface nEMG or needle nEMG signals during resting state. [25-29] To our knowledge, few studies have been reported analyzing nEMG signals during volitional state.

To overcome the limitations of nEMG, we developed a deep learning model named “nEMGNet” after the motifs from VGGNet and ResNet, which are known to show good performance in image analysis, and applied it to classifying the nEMG signal. [30, 31] To evaluate the usefulness of nEMGNet, we compared the classification results of nEMG signals by nEMGNet and 6 physicians with more than 1 year of nEMG reading experience.

For this study, we retrospectively reviewed nEMG waveforms, which were examined in subjects with neuropathy or myopathy or normal, analyzed those by using convolutional neural network.

**Methods**

**Study design and population**

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 Internal Review Board of Seoul National University Hospital (No. 2008-055-1147) and conducted according to the Declaration of Helsinki and its later amendments. Informed consent was not obtained because this study is retrospective analysis. nEMG was performed with a monopolar needle electrode from the subject’s muscles. (Viking Quest, Natus, Middleton, WI). 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. Based on the elbow joint of the upper extremity and the knee joint of the lower extremity, the muscles close to this joint were classified as proximal muscles and distal muscles.

The results of the waveform data of patients stored numerically in the electromyography machine were extracted, and they were made into a waveform through the MATLAB software (version R2020b) program. Among the created waveform data, artifacts occurring 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 some noise in the middle portion was preserved. The raw nEMG data, which was originally sampled at 48 kHz, was downsampled to 10 kHz to reduce computational complexity and sliced in fixed window length of 0.4 seconds units and hop size of 0.1 seconds units that were likely to be the most optimal length for post-experimental analysis. After slicing, total segments were composed of 2700 segments from subjects with myopathies, 3664 segments of subjects with neuropathies, and 1706 segments of subjects without neither neuropathy nor myopathy. Consequently, rest of the numerical data was used for analysis.

**Building the convolutional neural network model**

To find the characteristics of the nEMG signal, we used a 1-dimensional convolutional neural network (CNN) named nEMGNet. The structure of nEMGNet includes spatial block-1, which reduces the resolution by half, spatial block-2 which reduces the resolution by quarter, and residual block which solves the problem of poor backward propagation as the layer gets deeper by making a residual connection. (S1 Table)

|  |  |  |
| --- | --- | --- |
| **Spatial reduction block-1**  **(n,k)** | **Spatial reduction block-2**  **(n,k)** | **Residual block**  **(n)** |
| Conv(k)-n,  Stride(1) | Conv(k)-n,  Stride(2) | Conv(5)-n,  Stride(1) |
| BatchNorm | BatchNorm | BatchNorm |
| ReLU | ReLU | ReLU |
| Max-pool(2),  Stride(2) | Max-pool(2),  Stride(2) | Conv(5)-n,  Stride(1) |
|  |  | BatchNorm |

S1 Table. convolutional blocks of nEMGNet. n, the number of channel; k, the number of filter; conv, convolutional layer; BatchNorm, batch normalization; Max-pool, max pooling.

The nEMGNet was tested with 4 versions of nEMGNet-A, nEMGNet-B, nEMGNet-C, and nEMGNet-D with different versions according to the number of residual blocks. (S2 Table)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **nEMGNet-A** | **nEMGNet-B** | **nEMGNet-C** | **nEMGNet-D** |
| Block 1 | SR block-2  (64, 11) | SR block-2  (64, 11) | SR block-2  (64, 11) | SR block-2  (64, 11) |
| Block 2 | SR block-2  (64, 7) | SR block-2  (64, 7) | SR block-2  (64, 7) | SR block-2  (64, 7) |
| Block 3 | SR block-2  (64, 5) | SR block-2  (64, 5) | SR block-2  (64, 5) | SR block-2  (64, 5) |
| Block 4 |  | Residual block  (64) 2 | Residual block  (64) 4 | Residual block  (64) 6 |
| Block 5 | SR block-1  (128, 5) | SR block-1  (128, 5) | SR block-1  (128, 5) | SR block-1  (128, 5) |
| Block 6 |  | Residual block  (128) 2 | Residual block  (128) 4 | Residual block  (128) 6 |
| Block 7 | SR block-1  (256, 5) | SR block-1  (256, 5) | SR block-1  (256, 5) | SR block-1  (256, 5) |
| Block 8 |  | Residual block  (256) 2 | Residual block  (256) 4 | Residual block  (256) 6 |
| Block 9 | SR block-1  (512, 5) | SR block-1  (512, 5) | SR block-1  (512, 5) | SR block-1  (512, 5) |
| Block 10 |  | Residual block  (512) 2 | Residual block  (512) 4 | Residual block  (512) 6 |
| Block 11 | SR block-1  (1024, 5) | SR block-1  (1024, 5) | SR block-1  (1024, 5) | SR block-1  (1024, 5) |
| Block 12 |  | Residual block  (1024) 2 | Residual block  (1024) 4 | Residual block  (1024) 6 |
|  | FC-512 | FC-512 | FC-512 | FC-512 |
|  | FC-256 | FC-256 | FC-256 | FC-256 |
|  | FC-64 | FC-64 | FC-64 | FC-64 |
|  | FC-16 | FC-16 | FC-16 | FC-16 |
|  | FC-10 | FC-10 | FC-10 | FC-10 |
|  | Softmax | Softmax | Softmax | Softmax |

S2 Table. configuration of 4 different versions of nEMGNet. SR block, spatial resolution block; FC, fully connected layer.

A rectified linear unit (ReLU) is applied to the fully connected layer after the convolutional layer. (Figure 1)



Figure 1. nEMGNet structure. Processing composed of convolutional neural network, batch normalization, rectified linear unit (ReLU), max pooling. Light gray box; spatial block-1 and spatial block-2, Dark gray box and bold curved arrow; residual block. 유재성연구원 그림 바꿔서 주면 변경

The initial values of nEMGNet hyperparameter were empirically determined based on values that have been widely used. The learning rate, batch size, and epoch were set to 10-3, 32, and 100. Adam optimizer was used for optimizer, and inversely proportional values were used for the class weight to the number of signal segments for preventing erroneous prediction.

The number of muscles tested with nEMG is different for each subject, and among the tested muscles, abnormal and normal nEMG can coexist. To overcome these limitations, we applied a method called the DiVote (Divide and Vote) algorithm. DiVote divided each nEMG signal into segments of homogeneous length and converted it into 3 signal segment prediction scores through a feature extractor. The muscle signal prediction score was calculated by aggregating the signal segment prediction score and aggregated to derive the subject prediction score through soft voting. When deriving the subject prediction score, two different method were tried. The first method is to calculate and aggregate prediction scores by classifying them by neuropathy, myopathy, and normal without information on the location of the muscle. The second method is to classify whether it is proximal or distal according to the location of the muscle, and calculate and aggregate prediction scores of each neuropathy, myopathy, and normal. After 1500 training, in the process for extracting features of neuropathy, myopathy, and normal the learning rate was adjusted to 10-2 and gradient descent was applied.

There two major methods of aggregated probabilities corresponding to neuropathy, myopathy, and normal of the segment from nEMG signal data for each muscle, hard voting and soft voting. The former is to select one item with the highest probability for each segment, count the number of selected items and finally select the one that occupies the largest number among them. On the other hand, the latter is to select items with highest probabilities after averaging each probability from each segments. In the case of hard voting, where only values with high probability are selected for each segment, data loss might occur, so soft voting was used to reduce data loss. The most probable diagnosis obtained through soft voting in the segment is the muscle prediction score, and the subject prediction score is the soft voting of all muscles by subject after soft voting in the segment.

The classifier to distinguish neuropathy, myopathy, and normal was obtained in two ways and the classified results were compared. First, it was determined through logistic regression analysis with the subject features of each group as the independent variables and the diagnosis as the dependent variable; Second, it was determined by additionally considering muscle locations as an independent variables.

**Visualization and Evaluation of the results**

nEMG signal was classified into 2 stages through nEMGNet. First, the nEMG signals of individual muscles were classified regardless of the subject, and the individually classified nEMG signals of each muscle were collected for each subject, considering together, and then the subjects were classified. And the performance of nEMGNet was evaluated with the following indicators such as at each step. The accuracy, precision, recall, F1, area under receiver operating characteristic curve (AUROC), and Mathew’s correlation coefficient (MCC) were used as indicators for evaluation, and it was calculated using the following formula;

Accuracy=(TP+TN)/(TP+TN+FP+FN), Precision=TP/(TP+FP)

Recall=TP/(TP+FN)

F1=2×Precision×RecallPrecision+Recall

MCC=klm/(CkkClm-CklCmk)k(lCkl)(l'k'≠kCk'l')k(lClk)(l'k'≠kCl'k')

TP, number of true positive; TN, number of true negative; FP, number of false positive; FN, number of false negative, C; confusion matrix from n-class classification result with columns of true labels and rows of predicted labels. Ckk means that actual label is a ‘k’ label and predicted result by nEMGNet is also ‘k’ label.

We used feature visualization to identify the characteristics of each classified waveform learned through nEMGNet. The accuracy of nEMGNet was calculated by cross entropy, and since the number of subjects was small, the actual diagnosis and the predicted diagnosis for each subject were compared with 5-fold cross-validation.

In order to evaluate the applicability of nEMGNet to clinical practice, the nEMG signal numerical data were transformed to waveform data, which similar to the actual test screen shown on the screen of the nEMG machine, and the waveform that provided to the 6 residents of the Department of Neurology and Rehabilitation medicine who currently conduct and interpret nEMG. 데이터를 labeling하면 어떤 과정으로 저장해서 데이터 얻었는지 간략하게 서술(김동민연구원 내용주면 추가) (S1 Figure) The classified results by 6 residents were compared with results by nEMGNet. The degree of agreement between physicians and nEMGNet and accuracy were obtained.



S1 Figure. Example of provided nEMG waveform data

**Statistical analysis**

Statistical analyses were performed using R statistical software (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria). Inter-rater reliability was analyzed and presented with value of Fleiss kappa. The differences among the groups for categorical variables were assessed using the Fisher’s exact or Pearson’s χ2 tests and those for continuous variables were assessed using the Kruskal–Wallis tests or one-way analysis of variance tests. 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. 파이썬으로 정확도 등 구했다. (김동민연구원 내용주면 추가)

**Results**

The data of the subjects used for the analysis were 20 subjects with normal and 19 subjects with neuropathy with whom the diagnosis was radiculopathy, motor axonal polyneuropathy, motor neuron disease, etc., myopathy was 19 subjects with whom the diagnosis was muscular dystrophy and inflammatory myopathy. The number of nEMG signal data used for analysis was 125, 161, and 97, respectively, length was 204.31 seconds, 423.12 seconds, and 204.31 seconds. (Table 1)

Table 1. Demographic characteristics of subjects

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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) | 6.58±3.85 | 4.85±1.93 | 8.47±4.59 | 0.011 |
| Total signal length (sec) | 313.54 | 423.12 | 204.31 |  |

The results of filtering the entire nEMG segment through a convolutional block were plotted through a method called uniform manifold approximation and projection (UMAP). While the data had been passed through convolutional block, the dimension gradually decreased, neuropathy and myopathy were well distinguished, and the normal was located between the two groups. (S2 Figure) 

S2 Figure. Dimension reduction of signal segments after passing through convolutional layers of nEMGNet. (a) Initial state, state after passing through (b) 2nd block, (c) 4th block, (d) 6th block, (e) 9th block, (f) 12th block. z1 and z2, reduced dimensions; M, myopathy; N, neuropathy; NL, normal. 유재성연구원 그림 바꿔서 주면 변경

The classified results were depicted as a heatmap and a 3-dimensional plot. (Figure 3) The predicted result with the largest value among the muscle signal prediction scores through the DiVote pipeline is expressed in 3 different color, and the highter the probability, the darker the color. In the muscle signal prediction score, the predicted result with the largest value among the subject prediction scores that passed through the DiVote pipeline again was denoted by N for neuropathy, M for myopathy, and NL for normal. (Figure 2A) The subject prediction score was depicted as a 3-dimensional plot with the probability of being classified as myopathy, neuropathy, and normal as each axis. (Figure 2B) The classifier measured using logistic regression and argmax function was added as a decision boundary that distinguishes myopathy, neuropathy, and normal. (Figure 2B) Neuropathy and myopathy were distinguished relatively well, however, normal was directed toward the center and not well differentiated. (Figure 2C) The decision boundary was shifted using the classifier measured by adding the location information of the muscles divided into proximal or distal muscles, and as a result, the normal was better distinguished than the classifier measured without location information of the muscles. (Figure 3D) 

Figure 2. The results of subject classification through DiVote pipeline and decision boundary (A) Heatmap of the most probable diagnosis among muscle signal prediction scores. Each square box represents the most probable diagnosis value in color after aggregating the signal segment prediction scores predicted by nEMGNet. (B) The subject prediction scores as dots in a 3-dimensional plot. (C) Decision boundary calculated through simple argmax function. (D) Shifted decision boundary after adding the information on location of muscles. M, myopathy; N, neuropathy; NL, normal; P, proximal muscle; D, distal muscle. 유재성연구원 그림 바꿔서 주면 변경

The accuracy of total prediction over all, myopathy, neuropathy, and normal segments without processing of DiVote pipeline was 62.35%, 71.58%, 63.2%, and 52.26%. As a result of applying the DiVote pipeline, the accuracy improved to 76%–81%, and the accuracy was further improved to 76%–83% when the location of muscle was also considered. As a result of comparing the accuracy of 4 versions with different number of residual blocks among nEMGNets, nEMGNet-B, which included 2 residual blocks between spatial blocks, showed the best accuracy, while nEMGNet-A, which included no residual block, showed the poorest accuracy. (Table 2)

|  |  |  |  |
| --- | --- | --- | --- |
| nEMGNet subtype | Accuracy (%) | | |
| No classifier | Classifier using from subject features without muscle location | Classifier using from subject features with muscle location |
| A | 67.17±10.75 | 76.06±4.90 | 76.57±10.23 |
| B | 73.64±7.27 | 81.92±4.83 | 83.69±5.28 |
| C | 69.95±7.77 | 81.26±6.35 | 81.87±6.80 |
| D | 75.35±6.93 | 81.26±6.35 | 80.81±5.31 |

Table 2. The accuracy of subject classification according to nEMGNet version with and without DiVote (Divide and Vote) pipeline processing and additional information of muscle location. All values are expressed as mean ± standard deviation.

In the process of classifying with nEMGNet-B with muscle location information added, the weight values of myopathy, neuropathy, and normal were obtained by dividing them according to proximal and distal muscles. Five-fold cross validation was performed, and each fold was repeated 3 times to obtain each weight value from a total of 15 classifiers, and then the average value of all weights was measured. (S1 Table) For example, if the proximal muscle is classified as myopathy just before the final classification, the weight value multiplied during the final classification as myopathy were 1.56±0.96. For final classification as myopathy, the weight value of myopathy (proximal muscle, 1.56±0.96; distal muscle, 1.32±0.76) was the largest, followed by normal (proximal muscle, -1.63±0.99; distal muscle, -1.10±0.87) and neuropathy (proximal muscle, -1.76±0.94; distal muscle, -1.37±0.60), and among proximal and distal muscles, the weight value of proximal muscle myopathy (1.56±0.96) was larger than that of distal counterpart (1.32±0.76). For final classification as neuropathy, the order of the weight values was neuropathy (proximal muscle, 1.76±0.78; distal muscle, 1.29±0.83), myopathy (proximal muscle, -1.08±0.58; distal muscle, -1.06±0.80), and normal muscle (proximal muscle, -1.67±1.11; distal muscle, -2.12±1.28), and the weight value of proximal muscle neuropathy (1.76±0.78) was larger than that of distal counterpart (1.29±0.83). The results of other weight values are shown in S1 table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classified results | Input | | | | | |
| Proximal muscles | | | Distal muscles | | |
| M | N | NL | M | N | NL |
| M | 1.56±0.96 | -1.76±0.94 | -1.63±0.99 | 1.32±0.76 | -1.37±0.60 | -1.10±0.87 |
| N | -1.08±0.58 | 1.76±0.78 | -1.67±1.11 | -1.06±0.80 | 1.29±0.83 | -2.12±1.28 |
| NL | -1.53±1.18 | 0.17±0.98 | 1.53±1.05 | -0.78±0.92 | -1.31±0.97 | 1.93±1.13 |

S1 Table. Total results of weight values. All values were averaged over 15 weight values and are expressed in mean±standard deviation.

M, myopathy; N, neuropathy; NL, normal.

The accuracy of nEMGNet-B is depicted as confusion matrices by muscle prediction results and subject prediction results for each diagnosis. (Figure 3)



Figure 3. Confusion matrices of prediction by nEMGNet-B. Left; Accuracy of predicted results for each nEMG waveform, Right; Accuracy of predicted results by considering all nEMG of each patient together. Missing value는 제외했다는 내용 추가

Based on the results of training through nEMGNet, the characteristics of the waveform of myopathy, neuropathy, and normal were similar to characteristics of the actual nEMG waveforms. The waveform of myopathy showed small amplitude as well as short duration and counterpart of neuropathy showed high amplitude as well as long duration. (Figure 5)



S3 Figure. Trained waveforms of myopathy, neuropathy, and normal based on nEMGNet.

(A), (C) and (E), waveform based on learned features by nEMGNet; (B), (D) and (F), actual waveform

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

Note that (A), (B), (E), (F) were plotted with 5mV y-axis limit and (C), (D) was plotted with 20mV y-axis limit to show the overall shape of the nEMG signal from neuropathy subjects which has characteristic of high amplitude.

The prediction results of muscle and subject classification by not only nEMGNet but also 6 physicians were compared. The mean accuracies of the former were 53.66% and 52.01%; those of the latter were 68.67% and 81.03%, respectively. The inter-rater reliabilities for classifying each muscle nEMG and subject nEMG between physicians were 0.258 and 0.260 expressed in Fleiss κ; the inter-rater reliability between physicians and nEMGNet were 0.249 and 0.256, respectively. (Table 3)

|  |  |  |
| --- | --- | --- |
| Muscle classification results  Subject classification results | Physician | nEMGNet |
| Accuracy (%) | 0.54\*  0.52\* | 0.69  0.81 |
| F1 | 0.45\*  0.41\* | 0.51  0.61 |
| MCC | 0.30\*  0.31\* | 0.53  0.72 |
| Inter-rater reliability (Fleiss kappa) | 0.26†  0.26† | 0.25‡  0.26‡ |

Table 3. The results of classification by physicians (mean result of 6 physicians) and nEMGNet. Result was shown with sensitivity, specificity, inter-rater reliability.

\* Mean value of 6 physicians’ results.

† Fleiss κ value between nEMGNet’s result and physicians’ results

† Fleiss κ value between physicians’ results

**Discussion**

The aim of this study was to evaluate the accuracy of detecting the presence of peripheral neuropathy or myopathy by analyzing nEMG waveform data using machine learning, and to confirm its applicability in clinical practice. For that purpose, we applied deep learning named nEMGNet to interpreting the nEMG waveforms, and assess the performance and compare the accuracy with classification by 6 physicians. As a result of analysis with nEMGNet, the time required was shorter, and the accuracy was superior to accuracy of the physician’s analysis.

When nEMG data is classified as machine learning, there are some limitations in that the number of muscles tested for each subject and time length of nEMG data for each muscle is different, and the some nEMG data of all muscles showed abnormalities. DiVote pipeline was used to overcome these limitations and contributed to the improved accuracy. Additionally, considering that peripheral neuropathy mainly shows abnormalities in the distal part muscle, whereas, myopathy mainly shows abnormalities in the proximal part muscles, addtional information on muscle location, which means whether the muscles are located close to the trunk or not, was added to the nEMGNet and this contributed to improved accuracy. Neuropathy, myopathy를 구분하는데 특정 그룹에 proximal muscle이나 distal muscle이 많이 포함되어 있으면 세부적인 정보는 무시하고 포함된 근육의 위치정보로만 진단하는 bias를 방지하고자 다른 label은 개수가 부족하더라도 0.3으로 반영했다.(유재성연구원 내용주면 추가)

Previously, there have been reports that machine learning showed good performance when applied to image analysis, surface nEMG, and needle nEMG. [25-31] 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. [25-29] For the purpose of diagnosing neuromuscular disorders, needle nEMG is useful rather than surface nEMG, and not only the signal during resting state but also the signal of during volitional state should be considered among needle nEMG signal. [1-6, 8, 32, 33]. After minimal noise were removed at the beginning and the end of the nEMG data, the rest nEMG data as the 1-dimensional numerical data during the volitional state for minimizing the data loss that may occur while using the 2-dimensional data as in the previous studies. To confirm the clinical applicability of nEMGNet, the diagnostic accuracy of physicians was measured and compared with that of nEMGNet. Finally, we found that the accuracy and time-taken of diagnosing neuropathy, myopathy, and normal were 83.69% and 40 seconds in using only nEMG data by nEMGNet, which is better than that of the machine learning model found in previous studies or physicians.

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. Additionally, when myopathy and neuropathy were classified with nEMGNet, the weight values given to proximal and distal muscles were greater in both cases. In the case of myopathy, the weight value of proximal muscle is greater than that of distal counterpart, which is usually consistent with the more common proximal involvement in myopathy. On the other hand, in the case of neuropathy, the weight value of proximal muscle is greater than that of distal counterpart, which is slightly different from the previously reported result that distal involvement is more common in neuropathy.

Out study also has some limitations. First, this study deal with retrospective data from only 1 center study. Secondly, study number is not enough to demonstrate perfect usefulness of deep learning on nEMG classification. Finally, we focused only on dividing nEMG signal into neuropathy, myopathy, and normal. However, more specialized diagnosis could be identified with more concise machine learning algorithms. Future study with much more data from multicenter will show potential of applying machine learning to nEMG interpretation.

Until now, few studies on analyzing nEMG data of volitional state by deep learning have been documented. Our study suggest that machine learning has the possibilities to be embedded in nEMG machines, reducing errors in nEMG interpretation and the workload of physicians, and potentially preventing personal medial information leakage that can arise when nEMG data is uploaded online for nEMG analysis, so shed lights on diagnosis patient of suspected neuropathy or myopathy by machine learning which might help with nEMG signal classification. Decision support의 장점도 있다.

**Contributor and guarantor information**

YIH and KKW conceptualized this work.

**Supporting information**

S1 Fig.

**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. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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>.

23. 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.

24. 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.

25. 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>.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

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

32. Haig AJ, Gelblum JB, Rechtien JJ, Gitter AJ. Technology assessment: the use of surface nEMG 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.

33. 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.