**Studies on the usefulness of applying deep learning to interpreting EMG waveforms**

**Introduction**

Electromyography (EMG) is a electrophysiologic test that records electrical activity generated during muscle contraction controlled by nervous system. It is implemented to identify disorders of the nervous system or muscles according to abnormalities in EMG signals that reflect the anatomical and physiological characteristics of the nervous system and muscles. It is used as one of clues for diagnosing neurological or muscle diseases according to the characteristics of motor unit action potentials (MUAPs) among EMG signals. Typical neurogenic MUAPs seen in neurological diseases are known to have large amplitude, long duration, reduced recruitment, and reduced interference patterns. Interference is known to be small.

Electrophysiologic examination, including needle electromyography (EMG), nerve conduction studies, is obtained electrical signals from nerves, muscles and neuromuscular junctions, and is used to identifying any abnormalities in neuromuscular system, helping to estimate the location and extent of lesions. In case of patients with disorders of nerves, muscles, EMG shows waveforms made by damaged muscles or nerves including abnormal spontaneous potentials or action potentials generated by damaged muscles associated with nerve or muscle disorder. The diagnostic usefulness of electromyography for identifying peripheral neuropathy and myopathy has been suggested in previous studies. (1-7)

Previous studies have been reported that the accuracy of electromyography in patients with radiculopathy, the sensitivity was 36-64%, specificity 54-58%, inter-rater reliability was about 60-90% which is different according to type of abnormality.(1, 4, 6, 8-14) To recognize abnormalities of EMG waveform accurately, considerable time and efforts are needed. For diagnosing neuropathy and myopathy, higher accuracy of electromyography is required, and objective interpretation is required by decreasing the discrepancy between raters. As the importance and frequency of EMG, the workload of physicians and time taken for EMG get increased. This hardship might be alleviated by automated algorithms such as deep learning algorithms which assist physicians’ interpretations.

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. Recent development of deep learning technologies leads to gesture recognition based on EMG, assisting interpretation of electrocardiography and electroencephalography. Convolutional neural network, one kind of deep learning techniques, has applied to analyzing time series data and waveform data such as electrocardiography, electroencephalography. 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.(15) To our knowledge, previous studies on analyzing volitional EMG data have rarely been reported.

In order to increase the accuracy of electromyography, we applied deep learning to interpreting the EMG waveforms, and assess the accuracy of machine learning based-EMG interpretation and compare the results done by physicians. We retrospectively reviewed EMG waveforms, which were examined in patients with neuropathy or myopathy or normal, analyzed those by using convolutional neural network built-in Python.

**Methods and materials**

**Data acquisition and preparation**

The data analyzed in this article were from the Seoul National University Hospital database that includes electromyography data of 58 subjects, visited Seoul National University Hospital from Jun, 2015 to Jul, 2020 and divided into 3 datasets of neuropathy, myopathy and normal according to EMG waveforms characteristics which were neurogenic potentials or myopathic potentials or not. The criteria for dividing myopathy, neuropathy, and normal were whether there are one of small amplitude-short duration, high amplitudes-long duration, reduced recruitment, early recruitment, and reduced interference pattern during minimal, moderately and maximally contraction. 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. EMG was performed with a monopolar needle electrode from muscles of subjects body (Viking Quest (Natus, USA)). The filter setting was set at 20 Hz (low-cut) and 10 kHz (high-cut). The results of the last 10 seconds of the EMG 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 form through the MATLAB software (version R2020b) program and edited into a form suitable for analysis. The result of removing noise such as motion artifact that occurred while moving the needle electrode during the EMG test was used for analysis.

The raw EMG data, which was originally sampled at 48 kHz, was downsampled to 10 kHz for reducing computational complexity and sliced into multiple segments with a fixed length (0.4 seconds) which was decided after the experiment. 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.

**Building machine learning models**

One dimensional convolutional neural network was used to identifying EMG signal characteristics. nEMGNet

근전도 신호의 특징을 찾기위해 기존에 이미지의 분류에 효과를 입증받은 VGGNet과 ResNet에서 기반한 1차원 CNN을 사용했고 이를 nEMGNet으로 명명했습니다. nEMGNet의 구조는 3종류의 block으로 구성됐고 spatial reduction block-1은 spatial resolution을 반으로 줄이고 spatial reduction block-2는 spatial resolution을 4로 줄이고 residual block은 복잡한 특징을 찾아내기 위한 deep layer를 위한 안정적인 training을 가능하게 한다. Residual block의 개수를 달리해서 4가지의 nEMGNet을 구축했습니다. Convolutional layer이후의 fully connected layer에는 rectified linear unit (ReLU)를 적용했습니다. 분석에 사용한 EMG signal의 길이가 달라서 Divide and Vote (DiVote) pipeline을 적용하였습니다. DiVote는 각 EMG signal을 homogeneous shape의 segment로 나눴고 feature extractor를 통해 3개의 signal segment prediction score로 변환했습니다. Signal segment prediction score는 soft voting에 의해서 muscle signal prediction score로 aggregate해서 muscle별로 EMG 결과를 도출했고 muscle signal prediction score를 subject별로 aggregate해서 subject feature를 도출했습니다. Subject feature를 도출할 때 muscle을 proximal과 distal muscle로 구분한 추가정보를 넣고 분석한 것 (all)과 추가정보 없이 분석한 것 (proximal/distal) 2가지를 실험했습니다.

머신러닝 모델의 하이퍼파라미터의 설정은 경험적으로 결정했는데 learning rate는 10-3, batch size는 32, epoch는 100으로 설정하였습니다. Optimizer는 Adam optimizer를 사용했고 optimization을 위해 cross entropy를 사용했으며 class weight는 signal segment의 개수갯 반대의 방향으로 적용해서 예측에 있어서의 bias를 방지하고자 했다.

머신러닝은 convolutional neural network를 이용하여 근전도 파형을 분석하였음. 48Hz의 frequency로 분석하였고 0.2초 정도의 시간간격으로 데이터를 분할하여 학습시키고 validation및 적용을 하였음. 이후 환자의 최종임상 진단 중 myopathy, neuropathy, normal어디에 해당하는지의 결과와 머신러닝을 이용해 분석한 결과를 비교하여 정확도를 구했음 총 58명의 데이터로 상대적으로 수가 적어서 5-fold cross validation을 하여 머신러닝 알고리즘의 정확도를 평가했습니다. nEMGNet의 성능을 비교하기 위해 DiVote pipeline으로부터 segment를 추출해서 classification의 결과를 비교했습니다?

**Visualization and evaluation of result**

Feature visualization은 학습한 머신러닝 모델의 특징을 보여주기 위해 사용하는 technique인데 nEMGNet의 특징을 보여주는데 적용했습니다. 1500번의 학습이후 learning rate는 10-2로 조정했고 gradient descent를 적용했습니다.

nEMGNet의 평가는 2단계를 거쳐 했는데 neuropathy, myopathy, normal을 잘 구분해 내는지에 대해 평가를 했습니다. 우선 각 근육별 EMG signal segment의 분류 성능에 대해 평가를 했고 각각의 대상자의 진단을 하는 성능에 대해 평가했습니다. 평가를 위해 accuracy, precision, recall, F1, area under receiver operating characteristic curve (AUROC), mathew’s correlation coefficient (MCC)를 통해 하였고 다음과 같은 공식으로 계산했습니다.

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 labes.

3 class 분류 성능의 정확도는 accuracy와 MCC를 통해서 평가했습니다.

**Comparison the result of nEMGNet with physicians**

nEMGNet의 분석결과의 유용성을 검증하기 위해 현재 근전도 검사를 하고 있는 신경과와 재활의학과 소속의 레지던트 6명에게 전체 EMG data를 random shuffling한 뒤 근전도 파형데이터를 제외한 대상자의 정보는 제거하여 각 physician에게 파형 데이터를 제공한 뒤 각 physician이 분석한 결과를 nEMGNet이 예측한 결과와 비교하였습니다. 레지던트의 분석결과는 실제 환자에게 검사할 때 근전도 장치에서 파형을 보고 분석하는 것과 유사하게 하는 화면을 제공하여 분석하였습니다. (Suppl. Figure 1) 분석할 때에는 division당 10msec로 했고 파형의 amplitude는 200μV, 500μV, 1mV로 바꿔서 분석할 수 있게 하였고 500msec동안 지나간 파형을 보고 neuropathy, myopathy, normal중에 labeling을 하게 했습니다.

**Statistical analysis**

Statistical analyses were performed using R statistical software (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria). The *p*-value less than 0.05 was considered statistically significant. Analyses for calculating between the inter-rater and machine learning reliability were performed by chi-square test and result was reported by Fleiss κ and Cohen κ.

**Results**

신경병은 radiculopathy, axonal neuropathy, motor neuron disease 의 환자를 대상으로 하였고 근육병은 muscular dystrophy, inflammatory myopathy 등의 환자를 대상으로 하였음. (Table 1)

Table 1. Demographic characteristics

Figure 1. Examples of electromyography waveform

Signal segments를 convolutional block을 통해 filter한 결과를 uniform manifold approximation and projection이라고 하는 방법을 통해 plot으로 나타냈고 figure 3 depicts the results of the nEMGNet-B which showed best accuracy among several nEMGNet experimented.



Figure 3 Signal segments filtered by convolutional layers of nEMGNet. Filtered signal after each convolutional block of the nEMGNet is plotted by reducing the dimensionality to 2 dimensions using Uniform Manifold Approximation and Projection (UMAP). (a) Initial signal, (b) 2nd block, (c) 4th block, (d) 6th block, (e) 9th block, (f) 12th block. z1 and z2 corresponds to reduced dimensions. M stands for myopathy, N stands for neuropathy, and NL stands for Normal signal segments.

가장 좋은 performance를 보인 nEMGNet-B이 각 EMG signal segment별로 classification하는 정확도를 class로 나눠서 실제 class와 예측한 결과 class로 confusion matrix로 나타냈습니다. (Figure 4)

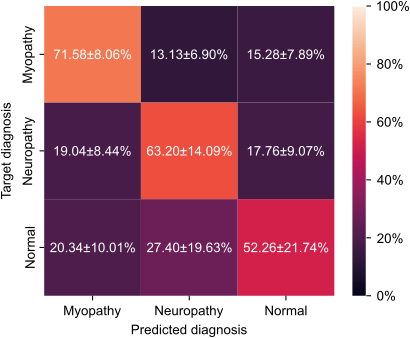


Figure 4 Confusion matrix of prediction by machine learning (nEMGNet-B)

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%. Figure 5 depicts the process of DiVote pipeline, especially, figure 5A는 각 subject별로 그리고 proximal, distal muscle별로 nEMGNet이 예측한 결과를 색을 달리해서 나타낸 것입니다. Figure 5B는 각 subject별로 근육별 예측결과를 종합해서 진단한 결과를 3차원의 plot에 나타낸 것인데 정상 subject는 neuropathy, myopathy에 비해 상대적으로 가운데에 치우쳐 있는 것을 확인했습니다. Figure 5B의 결과를 토대로 normal, neuropathy, myopathy를 나누는 기준이 되는 boundary를 argmax function을 이용해서 만들었습니다. Figure 5D는 정상인 대상자가 가운데에 애매하게 위치하는 것을 neuropathy나 myopathy와 조금 더 확연하게 구분되게 나타나게 하기위해 subject feature를 가지고 5B에서 더 훈련한 결과 boundary가 5C와 다르게 바뀐 것을 보여줍니다.



Figure 5 Process of DiVote pipeline. (a) Heatmap of muscle signals prediction scores for subjects from test set. Each square in the heatmap is computed by majority voting the signal segment prediction results, predicted by nEMGNet. (b) Subject features (all) plotted on 3d plane. Each point in the plot corresponds to features of each subject from train set, which are aggregated through majority voting the signal (muscle) features over all muscle types. Only existing muscle signals features are averaged per subject to create a single point. (c) Decision boundary of simple argmax function. (d) Decision boundary of classifier trained with subject features (b). M stands for myopathy, N stands for neuropathy, and NL stands for Normal signals and subjects.

각 subject를 classification하는데 추가정보 없이 분류했을 때에는 약 67-75%였는데 subject feature를 모두 넣었을 때는 약 76-81%로 조금 더 향상됐고 근육의 위치를 나타내는 proximal/distal에 대해서 추가로 정정보 넣었을 때 정확도가 더 향상돼서 약 76-83%의 정확도를 보였습니다. Subject feature type에 무관하게 추가 정보를 주었을 때 performance가 더 향상되는 결과를 보였고 nEMGNet별로 비교하면 SR block-2내에서 2개의 residual block을 포함하는 nEMGNet-B가 가장 좋은 performance를 보였고 이와 대조적으로 residual block을 포함하지 않는 nEMGNet-A가 가장 떨어지는 performance를 보였습니다. (Table 4)

|  |  |  |  |
| --- | --- | --- | --- |
| nEMGNet | No classifier  (argmax) | Subject features  (all) | Subject features  (proximal/distal) |
| A | 67.1710.75% | 76.064.90% | 76.5710.23% |
| B | 73.647.27% | 81.924.83% | **83.695.28%** |
| C | 69.957.77% | 81.266.35% | 81.876.80% |
| D | 75.356.93% | 81.266.35% | 80.815.31% |

Table 4 Subject diagnosis accuracy of nEMGNet and classifier pipeline

분석한 모든 nEMGNet의 performance를 evaluation metrics를 기준으로 기존에 발표된 다른 연구결과와 비교한 것입니다. (Table 5)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Reference | Feature extractor | Evaluation Metrics (%) | | | | | |
| Accuracy | F1 | Precision | Recall | AUROC | MCC |
| Proposed | SignalNet-A | 76.57 | 74.67 | 79.17 | 76.57 | 89.43 | 68.33 |
|  | SignalNet-B | **83.69** | **83.59** | **87.96** | **83.69** | **91.45** | **77.70** |
|  | SignalNet-C | 81.87 | 81.61 | 85.74 | 81.87 | 91.21 | 74.66 |
|  | SignalNet-D | 80.81 | 80.56 | 86.65 | 80.81 | 90.53 | 74.16 |
| Nam et al. [28] | Inception-v4 | 57.47 | 51.38 | 56.53 | 57.47 | 78.39 | 42.03 |
| Nodera et al. [27] | ResNet50 | 73.84 | 72.59 | 81.94 | 73.84 | 81.27 | 64.76 |
|  | ResNet152 | 75.61 | 74.90 | 81.11 | 75.61 | 85.30 | 65.50 |
|  | VGG16 | 68.23 | 65.47 | 66.86 | 68.23 | 79.85 | 55.42 |
|  | VGG19 | 72.32 | 69.66 | 74.71 | 72.32 | 81.10 | 62.91 |
|  | Inception-v3 | 71.92 | 70.64 | 79.29 | 71.92 | 83.48 | 61.60 |

흥미롭게도 nEMGNet을 통해 훈련된 근전도 데이터를 토대로 neuropathy, myopathy, normal의 파형 데이터의 특징을 시각화해서 본 결과 실제 대상자의 근전도 파형과 유사했는데 myopathy는 small amplitude, short duration, early recruitment의 특징을 보였고 neuropathy는 high amplitude의 long duration, reduced recruitment의 특징을 보였습니다. (Figure 7)

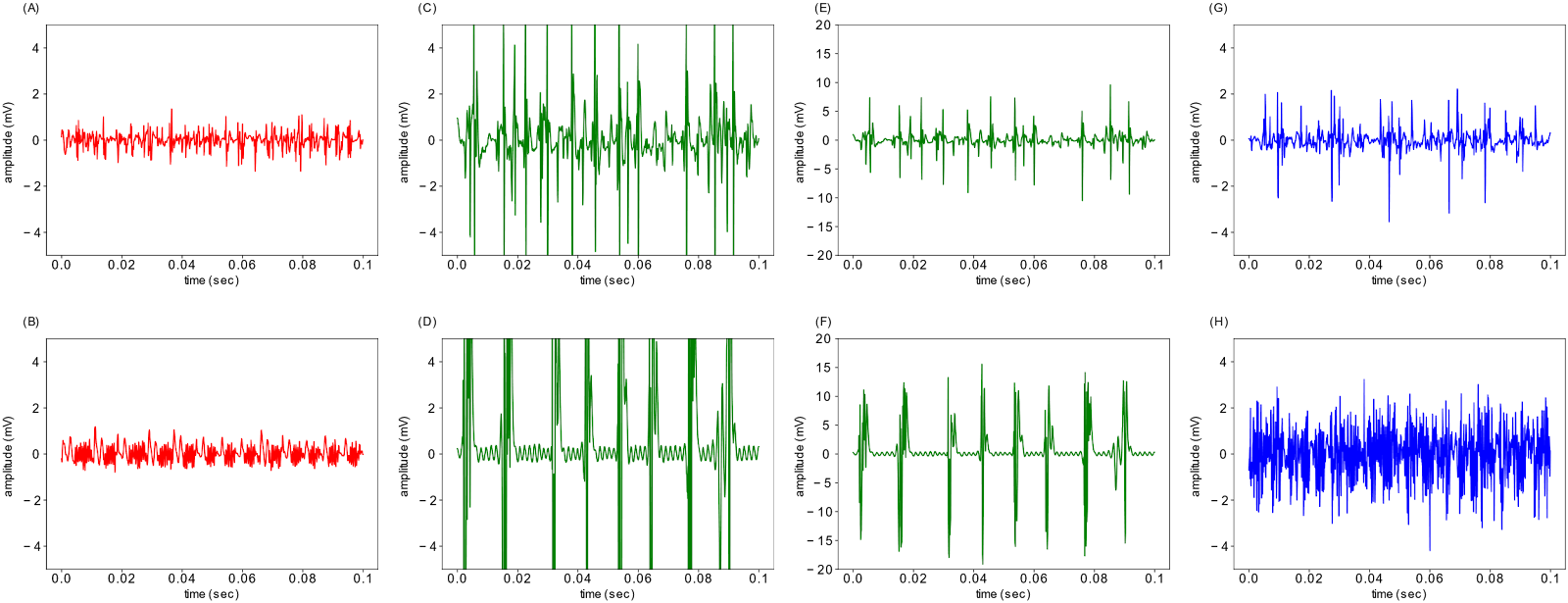


Figure 7 Learned features of nEMGNet. Top row are real signals and bottom row are generated signals through feature visualization. nEMGNet from the 1st fold was used to plot the figure. (a) Real myopathic signal. (b) Generated myopathic signal. (c) Real neuropathic signal. (d) Generated neuropathic signal. (e) Real neuropathic signal with 20mV y-axis limit. (f) Generated neuropathic signal with 20mV y-axis limit. (g) Real normal signal, (h) Generated normal signal. Note that (a), (b), (c), (d), (g), (h), is plotted with 5mV y-axis limit for better comparison between signals of different labels. (e), (f) is plotted with 20mV y-axis limit to show the overall shape of the neuropathic signal. (c)&(e) and (d)&(f) are identical signals.

nEMGNet과 phsycian의 classification 결과를 비교하기 위해 physician에게 EMG signal data를 가지고 근전도 기계와 거의 유사하게 파형을 재생하는 방식으로 보여주고 physician별로 각각 근육별로, subject별로 classification하게한 뒤 결과를 저장하고 소요된 시간을 측정했습니다. nEMGNet의 예측결과와 6명의 physician이 classification한 결과를 비교한 결과 physician들의 sensitivity는 median, IQR%, specificity는 median,IQR%였고 nEMGNet의 예측결과의 sensitivity는 ~%, specificity는 ~%였습니다. 그리고 physician간의 inter-rater reliability (Fleiss κ)는 ~였고 nEMGNet과 각 physician간의 inter-rater reliability는 ~였습니다. 그리고 classification에 소요된 시간(mean±SD)은 근육 1개당 소요된 시간은 physician이 ~초±~초였고 nEMGNet은 ~초±~초, subject 1명당 소요된 시간은 physician이 ~초±~초, nEMGNet이 ~초±~초로 nEMGNet이 훨씬 적은 시간에 classification을 완료하는 것으로 나타났습니다. (Table 6)

|  |  |  |
| --- | --- | --- |
|  | Physician | nEMGNet |
| Sensitivity (median, IQR) |  |  |
| Specificity (median, IQR) |  |  |
| Inter-rater reliability (Fleiss kappa)  Between physician  Between Physician and nEMGNet |  |  |
|  | |
| Elapsed time (sec) |  |  |

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

**Conclusions**

We applied deep learning to interpreting the EMG waveforms, and assess the accuracy of machine learning based-EMG interpretation and compare the results done by physicians. 결과는 사람이 분석한 것보다 더 좋은 정확도를 보였고 시간도 절약되는 장점을 보였습니다.

근육별로 파형을 분석한 결과에 비해 subject별로 분석을 했을 때 정확도가 더 향상되는 결과를 보였고 여기에 proximal muscle 인지 distal muscle인지와 같은 추가 정보를 넣었을 때 정확도가 더 향상되는 결과를 보였습니다. 이것을 실제 physician이 분석하는

The diagnostic usefulness of electromyography for identifying peripheral neuropathy and myopathy has been suggested in previous studies.

Recently, deep learning has been successfully applied to assisting diagnosis of medical diseases in so many ways. The diagnostic usefulness of electromyography for identifying peripheral neuropathy and myopathy has been suggested in previous studies.

We applied convolutional neural network to classification of volitional EMG waveforms via making numerical EMG data into waveform, editing and re-transforming into waveform data. Result of analysis showed that EMG waveform was well classified by deep learning algorithm. Our deep learning model might reduce error rate of EMG interpretation and physicians workload. This study suggest that the models built on deep learning-based EMG waveform interpretation might be complementary for physicians’ interpretation and promising.

Deep learning has shown good performance in many medical data including waveform and time series data such as electrocardiography and electroencephalography. In the analysis of waveforms of EMG, deep learning algorithm showed favorable performance compared with analyzed by physician and residents.

Until now, a few studies on analyzing EMG waveforms by deep learning have been documented.

In most cases, the accuracy and reliability of EMG interpretation by physician has been reported about 36-64% of the sensitivity, 54-58% of the specificity, 60-90% of inter-rater reliability. Interpreting the EMG might be difficult task for experienced physicians. From 6 different residents, the sensitivity was ~%, the specificity was ~%, and inter-rater concordance was ~%. With applying the deep learning algorithms, those were ~%, ~%, and ~%.

분석한 여러 nEMGNet의 performance를 기존에 발표된 다른 연구결과와 비교하면 better outcome을 보였고 그 중에서도 nEMGNet-B가 가장 좋은 performance를 보였습니다. (Table 5)

**References**

1. 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.

2. Bromberg MB. The motor unit and quantitative electromyography. Muscle Nerve. 2020;61(2):131-42.

3. Daube JR, Rubin DI. Needle electromyography. Muscle Nerve. 2009;39(2):244-70.

4. 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-.

5. Rubin DI. Needle electromyography: Basic concepts. Handb Clin Neurol. 2019;160:243-56.

6. 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.

7. Whittaker RG. The fundamentals of electromyography. Pract Neurol. 2012;12(3):187-94.

8. Haldeman S, Shouka M, Robboy S. Computed tomography, electrodiagnostic and clinical findings in chronic workers' compensation patients with back and leg pain. Spine (Phila Pa 1976). 1988;13(3):345-50.

9. Khatri BO, Baruah J, McQuillen MP. Correlation of electromyography with computed tomography in evaluation of lower back pain. Arch Neurol. 1984;41(6):594-7.

10. Kuruoglu R, Oh SJ, Thompson B. Clinical and electromyographic correlations of lumbosacral radiculopathy. Muscle Nerve. 1994;17(2):250-1.

11. Nardin RA, Patel MR, Gudas TF, Rutkove SB, Raynor EM. Electromyography and magnetic resonance imaging in the evaluation of radiculopathy. Muscle Nerve. 1999;22(2):151-5.

12. 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.

13. Wu ZA, Tsai CP, Yang DA, Chu FL, Chang T. Electrophysiologic study and computerized tomography in diagnosis of lumbosacral radiculopathy. Zhonghua Yi Xue Za Zhi. 1987;39(2):119-25.

14. Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve. 2006;34(2):238-41.

15. 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.