Computer Methods and Programs in Biomedicine 226 (2022) 107079

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability

Jaesung Yoo[a,1], Ilhan Yoo[b,1], Ina Youn[c], Sung-Min Kim[d], Ri Yu[e], Kwangsoo Kim[f], Keewon Kim[g,*], Seung-Bo Lee[h,*]

[a] School of Electrical Engineering, Korea University, Seoul, Republic of Korea
[b] Department of Neurology, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul, Republic of Korea
[c] Department of Computer Science, New York University, NY, USA
[d] Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea
[e] Department of Software and Computer Engineering, Department of Artificial Intelligence, Ajou University
[f] Transdisciplinary Department of Medicine and Advanced Technology, Seoul National University Hospital, Seoul, Republic of Korea
[g] Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[h] Department of Medical Informatics: Keimyung University School of Medicine, Daegu, Republic of Korea

ARTICLE INFO

Article history:
Received 24 January 2022
Revised 25 July 2022
Accepted 20 August 2022

Keywords:
Needle Electromyography
Neuromuscular Disorder
Electrophysiologic Diagnosis
Deep Learning
Convolutional Neural Network
Feature Visualization

ABSTRACT

Background and Objective: Neuromuscular disorders are diseases that damage our ability to control body movements. Needle electromyography (nEMG) is often used to diagnose neuromuscular disorders, which is an electrophysiological test measuring electric signals generated from a muscle using an invasive needle. Characteristics of nEMG signals are manually analyzed by an electromyographer to diagnose the types of neuromuscular disorders, and this process is highly dependent on the subjective experience of the electromyographer. Contemporary computer-aided methods utilized deep learning image classification models to classify nEMG signals which are not optimized for classifying signals. Additionally, model explainability was not addressed which is crucial in medical applications. This study aims to improve prediction accuracy, inference time, and explain model predictions in nEMG neuromuscular disorder classification.

Methods: This study introduces the nEMGNet, a one-dimensional convolutional neural network with residual connections designed to extract features from raw signals with higher accuracy and faster speed compared to image classification models from previous works. Next, the divide-and-vote (DiVote) algorithm was designed to integrate each subject's heterogeneous nEMG signal data structures and to utilize muscle subtype information for higher accuracy. Finally, feature visualization was used to identify the causality of nEMGNet diagnosis predictions, to ensure that nEMGNet made predictions on valid features, not artifacts.

Results: The proposed method was tested using 376 nEMG signals measured from 57 subjects between June 2015 to July 2020 in Seoul National University Hospital. The results from the three-class classification task demonstrated that nEMGNet's prediction accuracy of nEMG signal segments was 62.35%, and the subject diagnosis prediction accuracy of nEMGNet and the DiVote algorithm was 83.69 %, over 5-fold cross-validation. nEMGNet outperformed all models from previous works on nEMG diagnosis classification, and heuristic analysis of feature visualization results indicate that nEMGNet learned relevant nEMG signal characteristics.

Conclusions: This study introduced nEMGNet and DiVote algorithm which demonstrated fast and accurate performance in predicting neuromuscular disorders based on nEMG signals. The proposed method may be applied in medicine to support real-time electrophysiologic diagnosis.

* Corresponding authors at: K. Kim, Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul, Republic of Korea. SB. Lee, Department of Medical Informatics, Keimyung University School of Medicine, Daegu, Republic of Korea.
E-mail addresses: kksoo716@gmail.com (K. Kim), koreateam23@gmail.com (S.-B. Lee). [1] Authors contributed equally.

https://doi.org/10.1016/j.cmpb.2022.107079

1. Introduction

The ability to control our movements voluntarily is one of the basic functions of humans and is crucial to having a life of purpose. However, such an invaluable biological function may break down due to diabetes [1], chemotherapy [2], or other unknown reasons [3,4], resulting in neuromuscular disorders [5,6]. In the electrophysiologic diagnosis of neuromuscular disorders, needle electromyography (nEMG) has been widely used, which is an electrophysiological test that records electrical activity generated from nerves, muscles, and neuromuscular junctions by inserting a needle into a muscle at rest or during muscle contraction [7–11]. A skilled electromyographer diagnoses the types of neuromuscular disorders of the subject based on the abnormalities in measured nEMG signals [7,12]. Despite the effectiveness of nEMG in diagnosing the subtypes of neuromuscular disorders [13,14], the currently used subjective method is highly dependent on the experience of the electromyographer, making it vulnerable to errors, as is evident from the inter-rater reliability of 61-81% [15].

Among the various machine learning models, deep learning [16] has shown outstanding performance by leveraging the power of large data in nonlinear tasks that are difficult to analyze mathematically [17–19]. While deep learning has also been increasingly applied in medicine [20–23], its application in nEMG neuromuscular disorder classification remains limited. Previous works on deep learning neuromuscular disorder classification using electromyography signals often utilized surface electromyography data [24–26], and existing works that used nEMG signals were primarily focused on utilizing conventional machine learning models and hand-crafted feature extraction methods [27–31]. There have been two studies on deep learning nEMG neuromuscular disorder classification, both of which utilized image classification models to classify the nEMG signals. Nodera et. al. [32] generated mel-spectrograms from nEMG signals and used pretrained image classification models to classify signals into six types of diagnosis labels. Nam et. al. [33] used plotted nEMG signal images and utilized a pretrained image classification model to classify signals into three types of diagnosis labels.

While previous works have demonstrated the potential of deep learning application in nEMG diagnosis classification, additional improvements are required to apply deep learning in real nEMG electrophysiologic diagnosis. First, comparable performance is not guaranteed when using image classification models in signal classification tasks since the models were designed for natural image classification [34]. Second, previous works have focused on classifying the signals instead of the individual subjects. However, signals of various lengths are measured from different types and numbers of muscles for each subject in nEMG electrodiagnosis. An appropriate measure to integrate the heterogeneity in signal length, muscle sources, and number of signals is necessary to predict the diagnosis of subjects. Lastly, it is critical to identify how a machine learning model makes predictions, especially in medical applications [35–38]. A machine learning model must be investigated to ensure the model is making predictions based on relevant features and not artifacts [39,40], a topic not addressed by previous works.

In this study, we propose nEMGNet, a one-dimensional convolutional neural network (1D-CNN) model that extracts features from raw nEMG signals with improved accuracy compared to models from previous works [32,33]. The motivation for nEMGNet was to build a domain-specific deep learning model rather than to use models designed for other tasks. This study also introduces the divide-and-vote (DiVote) algorithm to predict the diagnosis of a subject by integrating the heterogeneous muscle signals into a homogeneous form. By combining nEMGNet and DiVote algorithm, diagnoses of subjects were predicted from a heterogeneous data structure, which allows for practical implementation in electrophysiologic diagnosis. After the performance of nEMGNet and the DiVote algorithm was verified, we identified how nEMGNet made predictions by applying feature visualization [41,42]. Feature visualization is a powerful explainable deep learning technique that shows the features learned by a deep learning model.

Table 1
Summary of needle electromyography dataset

	Myopathy	Neuropathy	Normal	Total
Number of subjects	19	19	19	57
Number of signals	122	160	94	376
Proximal muscle signals	64	63	17	144
Distal muscle signals	58	97	77	232
Total signal length (s)	312.84	422.78	203.50	939.12

2. Materials and methods

2.1. Data

The nEMG dataset included information from 57 subjects who visited Seoul National University Hospital from June 2015 to July 2020. Each subject was diagnosed with one of three neuromuscular disorder types by a certified electromyographer. Types of neuromuscular disorders were categorized into myopathy (M), neuropathy (N), and normal (NL). The signal characteristics of each neuromuscular disorder type are presented in Fig. 1. Single-channel nEMG signals were sampled at 48 kHz. Each subject contained a different number of signals acquired from different muscles, and each signal from each muscle was heterogeneous in length. The shortest signal was 0.41 seconds long, and the longest signal was 4.00 seconds long. Each signal contained the muscle location information it was recorded from, either a proximal muscle (P) or distal muscle (D). A summary of the nEMG dataset is shown in Table 1. This study was approved by the Institutional Review Board (IRB) of Seoul National University Hospital (No. 2008-055-1147) and was conducted in accordance with the Declaration of Helsinki and its later amendments. Informed consent was not obtained as the study was retrospective.

2.2. Preprocessing

Raw nEMG signals were used as input to avoid loss of information content [43] through transformations and reduce the time spent in preprocessing. The signals were downsampled to 10 kHz to reduce computational complexity. Downsampled frequency was selected after inspecting the morphology of downsampled signals under different frequencies. Each waveform was sliced into multiple segments by slicing the waveform with fixed window length T and hop size d. T=0.4 seconds, d=0.1 seconds was chosen empirically after experimentation. The resulting signal segment was 4000 samples in length. Each signal segment was labeled with the diagnosis label of the corresponding subject. There were 2700 myopathic signal segments, 3664 neuropathic signal segments, and 1706 normal signal segments after preprocessing.

2.3. nEMGNet model

A 1D-CNN model was designed to capture signal characteristics from raw nEMG signals. The nEMGNet design was inspired by VGGNet [44] and ResNet [45], which are neural network models that have demonstrated outstanding performance in image classification tasks. nEMGNet comprises three types of blocks that are analogous to the building blocks of VGGNet and Resnet (Table 2).

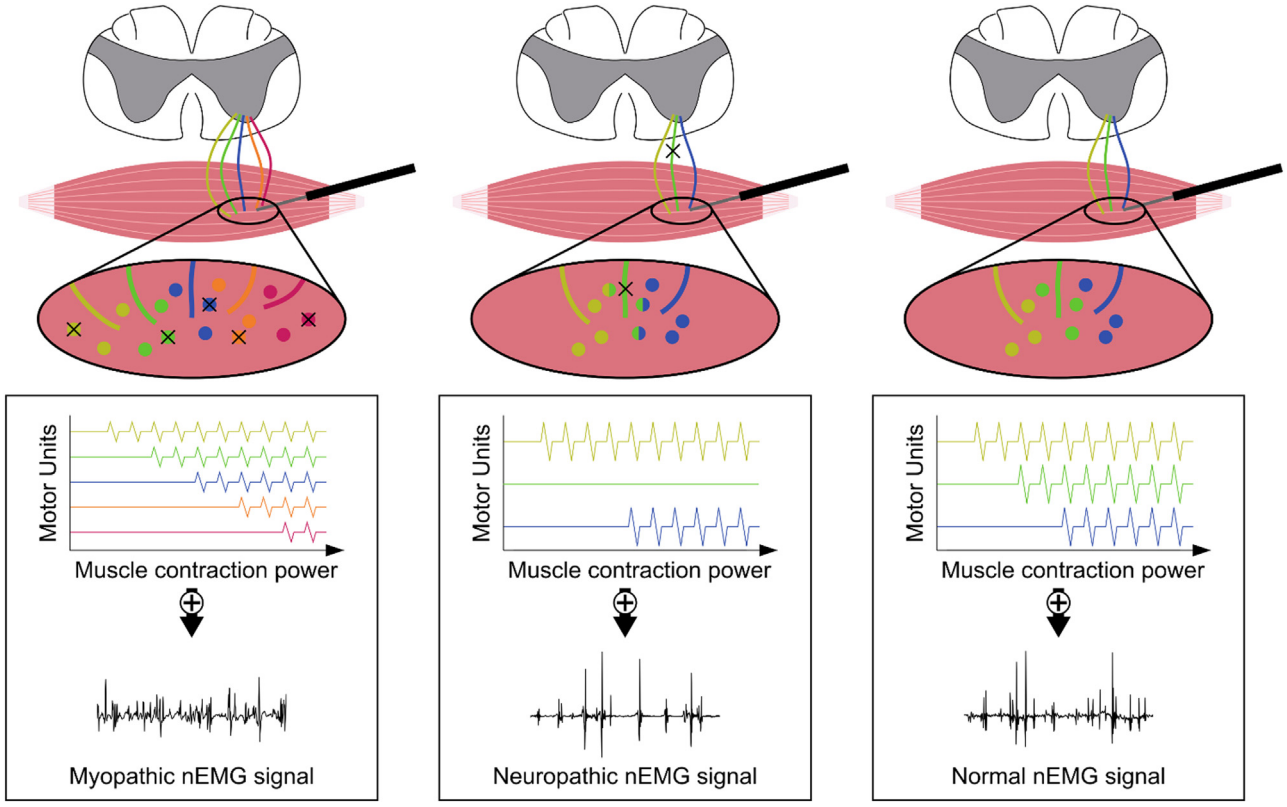

Fig. 1. Needle electromyography signals of neuromuscular disorder type. Myopathy refers to the damage in the muscle fibers, which results in motor unit action potentials with small amplitudes and short durations. Neuropathy refers to the damage in the peripheral nerves, which shows motor unit action potentials with large amplitudes and long durations. nEMG signals from normal muscles show motor unit action potentials with medium amplitudes and durations than myopathy or neuropathy.

Table 2
Convolutional blocks of nEMGNet. Spatial reduction block-1 reduces the spatial resolution to 50%, whereas spatial reduction block-2 reduces the spatial resolution to 25%. Residual blocks have residual connections which enable stable training in deep layers. In each block's title, 'n' and 'k' inside the parentheses refer to the number of output channels and the size of the kernel, respectively.

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

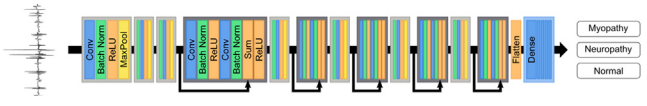

Fig. 2. nEMGNet model architecture. Blocks in light grey represent spatial reduction blocks, and blocks in dark grey represent residual blocks. The number of residual blocks between spatial reduction blocks varies with different versions of nEMGNet.

Spatial reduction block-1 (SR block-1) reduces the spatial resolution to 50%, and spatial reduction block-2 (SR block-2) reduces the spatial resolution to 25%. Residual block has residual connections which enable stable training in deep layers [46]. Various versions of nEMGNet with different numbers of residual blocks were tested since the residual blocks can be repeated an arbitrary number of times. The configuration of nEMGNet versions is described in Table 3. The rectified linear unit (ReLU) activation function was applied to each hidden layer in the fully connected layers. The softmax function was applied to the final output layer. Signal segments and diagnosis labels were used to train nEMGNet. Proximal and distal muscle information was not used to train nEMGNet.

2.4. Divide-and-Vote (DiVote) algorithm

In nEMG electrophysiologic diagnosis, each subject contains a different number of signals from different muscles, and the length of each muscle signal is heterogeneous. This study proposes the "Divide-and-Vote (DiVote)" algorithm (Fig. 3) to predict the diagnosis of a subject based on a given heterogeneous data structure. During the inference stage, each muscle signal is divided into signal segments of homogeneous shape, which are equivalent to the preprocessed signal segments in this study. Each signal segment is then converted into three-class signal segment prediction score by a feature extractor, which contains the probabilities for each diagnosis label. nEMGNet is used as a feature extractor in this study.

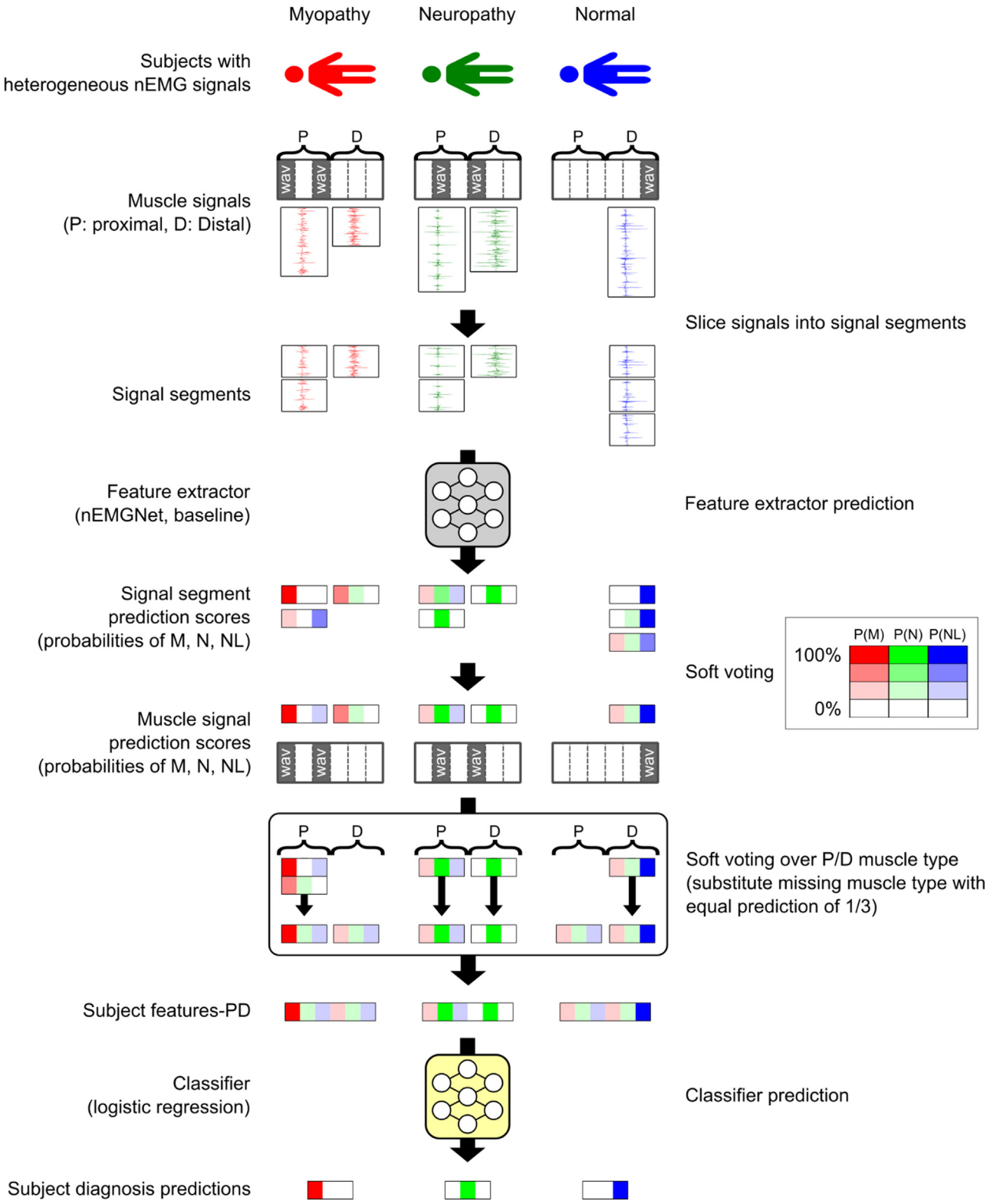

Fig. 3. DiVote algorithm. The positioning and colors in the signal segment prediction scores, muscle signal prediction scores, subject features, and subject diagnosis predictions correspond to myopathy, neuropathy, and normal. The color intensity indicates the magnitude of the value. Subject features-PD is created by soft voting muscle signal prediction scores over proximal and distal muscle separately. Only subject features-PD is described, and subject features-all is omitted.

Table 3
Configuration of nEMGNet. Different numbers of residual blocks are experimented with. The numbers inside the parentheses of the blocks refer to the hyperparameters defined in Table 2.

	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-3	FC-3	FC-3	FC-3
	Softmax	Softmax	Softmax	Softmax

The signal segment prediction scores from the same muscle signal are soft voted to generate the muscle signal's prediction score. The muscle signal prediction score contains prediction probabilities for each of the three diagnosis labels. Since each subject possesses different numbers and types of muscle signals, they also possess different numbers and types of muscle signal prediction scores. This heterogeneity is integrated by soft voting the muscle signal prediction scores which generates subject features assigned to each subject. Two types of subject features were tested: soft voting over all muscle types, named "subject feature-all," and soft voting over proximal and distal muscles and concatenating the results, named "subject feature-PD." Therefore, subject feature-all is a three-dimensional vector, whereas subject features-PD is a six-dimensional vector. When deriving subject feature-PD for a subject with no muscle signal prediction score in a specific muscle type, an equal prediction score of 1/3 was substituted (Fig. 3, myopathy and normal subjects). Finally, a logistic regression classifier was used to predict the final three-class diagnosis label of the subject based on the subject features. The feature extractor was trained from signal segments, and the logistic regression classifier was subsequently trained from subject features generated by the trained feature extractor.

2.5. Experimental setup

Evaluation results from 5-fold cross-validation were reported as the final performance of the model. Each fold was repeated three times with different nEMGNet weight initializations since the performance of deep learning model is dependent on weight initialization [47]. The training lasted 100 epochs using the Adam optimizer [48] with a learning rate of 1e-3 and a batch size of 32. The model was evaluated using the test set every 50 updates to find the weights configuration that produced the best accuracy and prevent overfitting to the train set. Cross entropy was used as the objective function, and class weights were applied inversely proportional to the number of signal segments per diagnosis label to prevent bias in prediction. Hyperparameters were selected empirically, and baseline experiments were conducted using nEMG classification models from previous works to compare the predictive performance of nEMGNet. Baseline nEMG classification models were used as the feature extractor from the DiVote algorithm. In the baseline experiment using a model from Nam et al. [33], the downsampled signal segments were plotted as images and were used to train the Inception-v4 [49] image classification model. When using baseline models from Nodera et al. [32], mel-spectrograms were generated from the same signal segments and were used to train the ResNet-50, ResNet-152 [45], VGG16, VGG19 [44], and Inception-v3 [50] image classification models. All experiments were performed on NVIDIA V100 GPUs.

2.6. Feature visualization

Feature visualization is an explainable deep learning technique that extracts the features a neural network has learned to accomplish a task [41,42]. Feature visualization was applied to a trained nEMGNet to visualize the features learned by the nEMGNet. Three types of augmented signal segments were generated by optimizing random noise with respect to each nEMGNet output node, which corresponds to each diagnosis label. Initial signal segments were generated from $N(0, 1)$. Random jitter was applied by 12.5% of signal segment length to prevent over-optimization of the signal segments. The learning rate was set to 1e-2, gradient descent was applied for 1500 updates, and the Adam [48] optimizer was used to optimize the signal segments.

2.7. Evaluation metrics

The evaluation was performed in two stages. Signal segment classification was evaluated first, followed by subject diagnosis classification. Both evaluations were three-class classification tasks. Accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUROC), and Mathew's correlation coefficient (MCC) [51] were computed. Performance metrics were computed using Eqs. (1)-(5).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1 = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

$$MCC = \frac{\sum_k \sum_l \sum_m (C_{kk} C_{lm} - C_{kl} C_{mk})}{\sqrt{\sum_k \left(\sum_l C_{kl}\right)\left(\sum_{l'} \sum_{k' \neq k} C_{k'l'}\right)} \sqrt{\sum_k \left(\sum_l C_{lk}\right)\left(\sum_{l'} \sum_{k' \neq k} C_{l'k'}\right)}} \tag{5}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. C denotes the confusion matrix from n-class classification results, with columns indicating true labels and rows indicating predicted labels. Note that accuracy and MCC are computed as three-class classification metrics, whereas all other methods are weighted averages of metrics from the one-versus-rest classification method [52], weighted according to the number of samples per class.

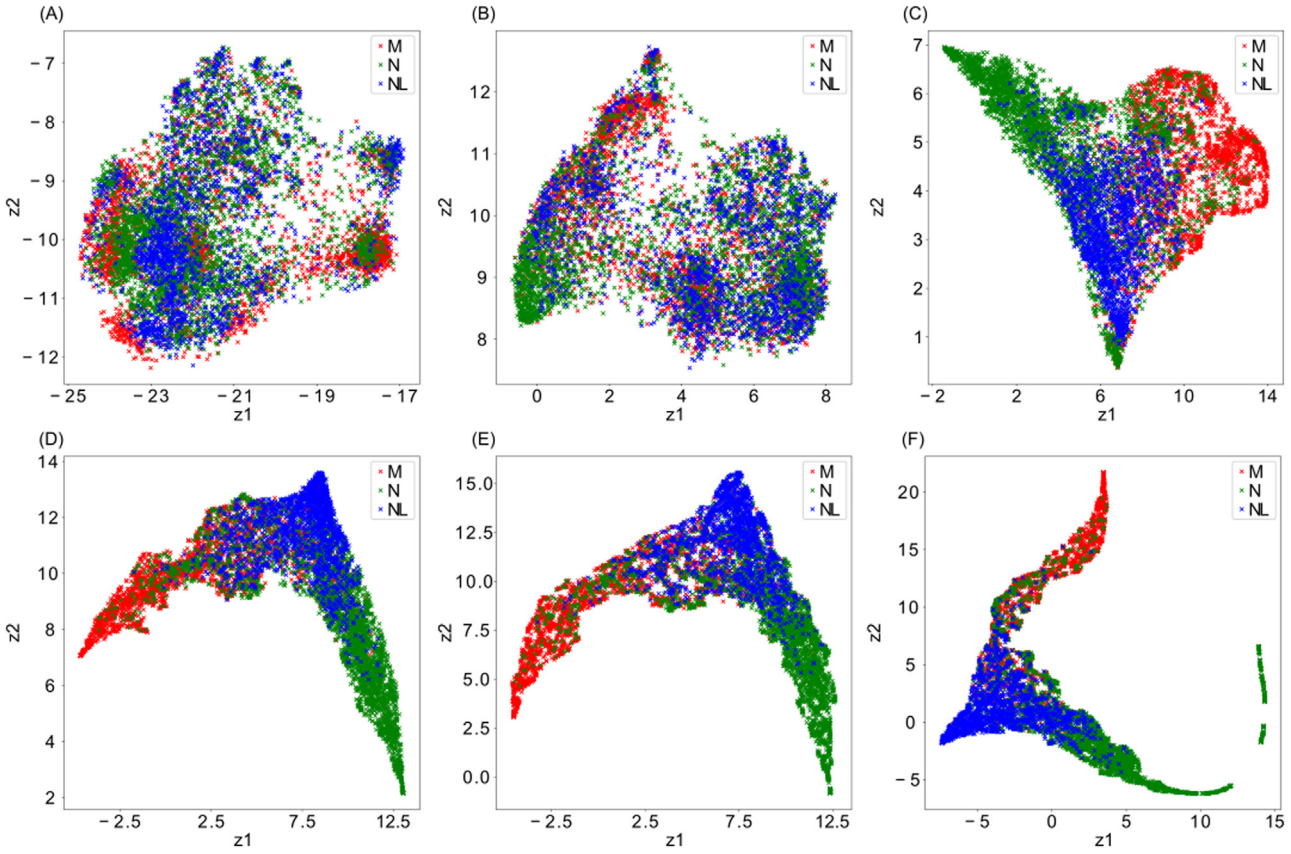

Fig. 4. Signal segments filtered through convolutional layers of nEMGNet. Filtered signals after passing through different convolutional blocks of the nEMGNet are plotted by reducing the dimensionality to two dimensions using Uniform Manifold Approximation and Projection. (a) Initial signal, (b) 2nd block, (c) 4th block, (d) 6th block, (e) 9th block, and (f) 12th block. z1 and z2 correspond to reduced dimensions. M stands for myopathy, N stands for neuropathy, and NL stands for normal signal segments.

3. Results

3.1. EMG segment classification

The signal segments filtered after different convolutional blocks were plotted to visualize the filtering effect of nEMGNet. Uniform manifold approximation and projection [53] was used to reduce the dimension of the filtered signal segments to two dimensions. The best performing nEMGNet-B among the different nEMGNet versions was used to plot Fig. 4. Tensors after fully connected layers were not used to plot Fig. 4 since the fully connected layers are not convolutional operations, and thus, they do not filter the signal segments. The signal segments from each label were farther apart in the reduced dimension as they were filtered through deeper layers. Additionally, the filtered signals of normal signal segments lied between myopathic and neuropathic filtered signal segments.

The confusion matrix of signal segment classification accuracy was plotted to demonstrate the classification performance of nEMGNet for each diagnosis label (Fig. 5). Results from the best performing nEMGNet-B were used to plot the confusion matrix. Confusion matrices were normalized over all folds and random repetitions. Total signal segment prediction accuracy over all classes was 62.35± 4.60%. Myopathy and neuropathy signal segments were classified with prediction accuracies of 71.58% and 63.20%, respectively, whereas normal signal segments were relatively more misclassified with a prediction accuracy of 52.26%.

The DiVote algorithm, which is run after signal segment prediction, is shown in Fig. 6. The train and test set from the first fold and nEMGNet-B were used to plot the figure. Fig. 6A shows that each subject possessed different types and numbers of muscle signals. In addition, signals from myopathic or neuropathic subjects tended to be predicted as their respective labels, whereas signals from normal subjects tended to be ambiguous, which is notable from the colors of the heatmap (Fig. 6A). Note that subject features-PD cannot be plotted similarly to Fig. 6B since subject features-PD is six-dimensional and cannot be plotted in three-dimensional space. Due to ambiguous prediction results of muscle

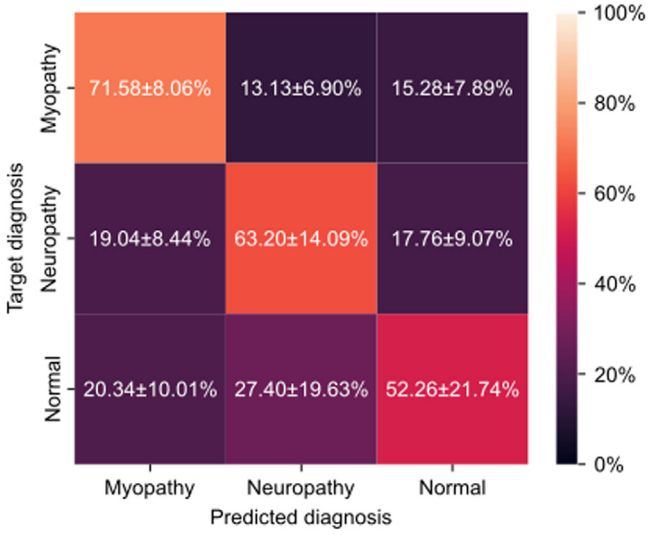

Fig. 5. Confusion matrix of signal segment classification accuracy. The values indicate mean± standard deviation over all folds and random repetitions.

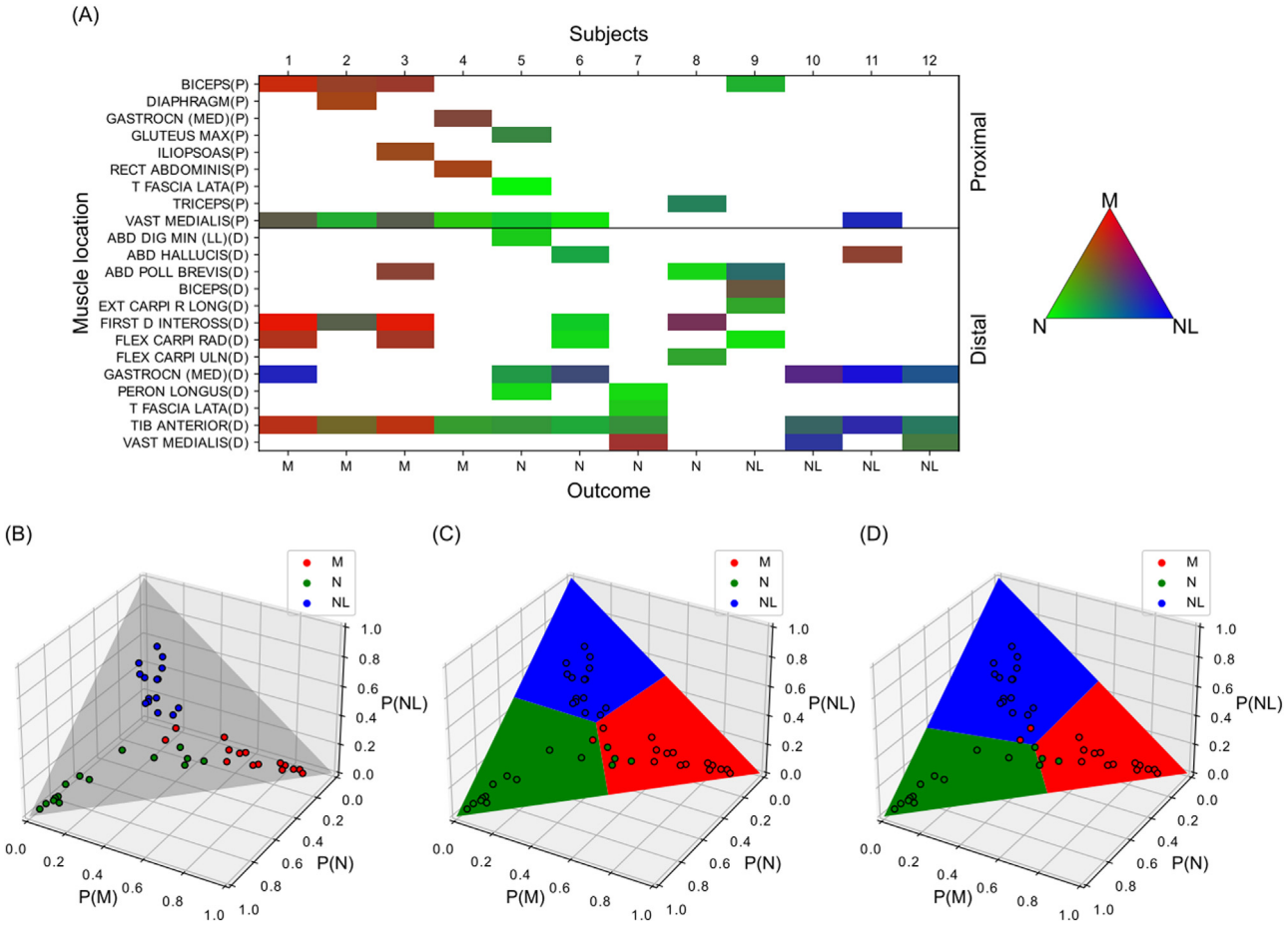

Fig. 6. Process of DiVote algorithm. (a) Heatmap of muscle signal prediction scores for subjects from the test set. Each square in the heatmap was computed through soft voting the signal segment prediction scores within a muscle signal, predicted by nEMGNet. White blocks represent missing signals. The red, green, and blue colors represent the probability scores for myopathy, neuropathy, and normal. (b) Subject features-all plotted on a P(M)+P(N)+P(NL)=1 plane, where P(M), P(N), P(NL) stands for muscle signal prediction scores for myopathy, neuropathy, and normal, respectively. Each point in the plot corresponds to each subject feature from the train set, which was aggregated through soft voting the muscle signal prediction scores over all muscle types. The same scatterplot is described in (c) and (d). (c) Decision boundary of simple argmax function. (d) Decision boundary of the classifier trained with subject features of (b). M stands for myopathy, N stands for neuropathy, and NL stands for normal signals and subjects.

signals from normal subjects, subject features of normal subjects were relatively biased toward myopathic and neuropathic features (Fig. 6B). An elementary approach in deriving the final subject diagnosis prediction scores involved using the subject features from Fig. 6B as the final subject prediction scores, for which the decision boundary is presented in Fig. 6C. A classifier was trained using the subject features from Fig. 6B to mitigate the bias of subject features from normal subjects, thus moving the decision boundary in feature space (Fig. 6D).

3.2. Subject diagnosis prediction

Subject diagnosis prediction accuracies of different nEMGNet versions and various prediction methods within the DiVote algorithm were compared (Table 4). The simple averaging setting used the average of muscle signal prediction scores as the final subject prediction score, which is equivalent to using subject features-all as the final subject prediction score. As there was no additional classifier, the decision boundary of simple averaging in feature space was equal to the standard argmax function (Fig. 6C). Using an additional classifier always resulted in better performance regardless of the nEMGNet version and subject feature type. Using subject features-PD resulted in better performance than using subject features-all for most nEMGNet models. nEMGNet-B, which contains two residual blocks between SR blocks, presented the best performance among the different nEMGNet versions. nEMGNet-A, which has no residual blocks, presented the worst performance. Increasing the number of residual blocks did not consistently increase model performance.

The performance of nEMGNet was compared to models from previous works [32,33] by using baseline models as the feature extractor in the DiVote algorithm. Table 5 describes subject diagnosis prediction results of nEMGNet and baseline models. nEMGNet-B presented the best performance in all metrics among the various nEMGNet versions, and nEMGNet-B outperformed all models from previous works in all evaluation metrics. nEMGNet-B showed an accuracy improvement of 8.08% compared to the best performing baseline model (ResNet-152). Models from Nodera et al. [32], which used mel spectrograms, presented higher scores in all metrics than the model from Nam et al. [33], which used plotted images. Among different models from Nodera et al. [32], ResNet-152 presented the best performance in terms of classification accuracy.

The time spent in the complete subject diagnosis prediction pipeline and the number of parameters for each model were measured for nEMGNet-B and baseline models (Table 6). Test sets from each fold and random repetition were used to measure inference time, and the average number of signal segments, muscle signals, and subjects in test sets were 1614, 75.2, and 11.4, respectively. As

Table 4
Subject diagnosis prediction accuracy of nEMGNet and DiVote algorithm. Different types of nEMGNet and subject features within the DiVote algorithm are compared. Subject features-all refers to those created by soft voting the muscle signal prediction scores over all muscle types. Subject features-PD refers to subject features created by soft voting the same scores over proximal and distal muscle types and concatenating the results. Best performance is described in bold font. Results are expressed in mean±standard deviation over all folds and random repetitions.

nEMGNet	Accuracy (%)		
	Simple averaging (No classifier)	Subject features-all	Subject features-PD
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 5
Subject diagnosis prediction result of nEMGNet and baseline models. Baseline models were used as feature extractors within the DiVote algorithm to compare the performance. Subject features-PD was generated from the DiVote algorithm. Metrics are mean values over all folds and random repetitions. Best evaluation metrics are described in bold font.

Reference	Feature extractor	Evaluation Metrics (%)					
		Accuracy	F1	Precision	Recall	AUROC	MCC
Current study	nEMGNet-A	76.57	74.67	79.17	76.57	89.43	68.33
	nEMGNet-B	**83.69**	**83.59**	**87.96**	**83.69**	**91.45**	**77.70**
	nEMGNet-C	81.87	81.61	85.74	81.87	91.21	74.66
	nEMGNet-D	80.81	80.56	86.65	80.81	90.53	74.16
Nam et al. [33]	Inception-v4	57.47	51.38	56.53	57.47	78.39	42.03
Nodera et al. [32]	ResNet-50	73.84	72.59	81.94	73.84	81.27	64.76
	ResNet-152	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

Table 6
Elapsed times for subject diagnosis prediction and model parameter size of nEMGNet and baseline models. nEMGNet-B and baseline feature extractors are compared. The time spent to predict subject diagnosis of the test set is measured. Elapsed times are the mean values over all folds and random repetitions. Elapsed time for classifier prediction is not described since the classifier is identical across all methods, and the time measured was less than 0.01 seconds. Total elapsed time refers to the time required to preprocess the signal segments, load the model parameters, extract subject features using the feature extractor, and compute the subject diagnosis prediction scores. The best metrics are shown in bold font.

Reference	Feature extractor	Elapsed time (s)				Number of parameters
		Preprocessing	Load model parameters	Feature extractor prediction	Total	
Proposed	nEMGNet-B	**0.01**	**0.43**	**0.54**	**0.99**	33,238,686
Nam et al. [33]	Inception-v4	68.92	1.17	5.32	75.42	41,210,744
Nodera et al. [32]	ResNet-50	19.08	0.61	1.21	20.90	23,567,352
	ResNet-152	19.20	1.45	2.25	22.92	58,301,534
	VGG16	19.61	2.63	1.60	23.87	134,272,835
	VGG19	19.53	2.71	1.78	24.04	139,582,531
	Inception-v3	19.42	2.60	1.38	23.40	**21,826,241**

there was no preprocessing for nEMGNet, the 0.01 second is measured from loading the signal segments. Additionally, preprocessing for baseline models from Nodera et al. [32] is identical and the slight variations in preprocessing time are empirical variances. The nEMGNet requires significantly less time for preprocessing than other baseline models because no preprocessing was applied. The time consumed in loading the model parameters corresponds to the number of parameters but is not directly proportional. The time spent in feature extractor prediction was the lowest for the nEMGNet, with 0.54 seconds. The total time required from preprocessing to subject diagnosis prediction was also the fastest for the nEMGNet, with 0.99 seconds. nEMGNet presented a 96.7% reduction in total inference time compared to the best performing baseline model (ResNet-152), from 22.92 seconds to 0.99 seconds. The number of parameters includes the weights and biases of layers and the running mean and variance of batch normalization layers. nEMGNet-B was smaller in model size compared to the ResNet-152, which was the best performing model among baseline models. Inception-v3 was the smallest in terms of model size.

3.3. Learned features of trained nEMGNet

Feature visualization results of nEMGNet are presented in Fig. 7. The bottom row of Fig. 7 contains the signal segments generated through feature visualization, which the nEMGNet best perceives as myopathy, neuropathy, and normal. The maximum absolute signal amplitude was 1.35mV, 1.27mV, 10.50mV, 19.10mV, 3.54mV, and 4.19mV from Fig. 7A, B, C (E), D (F), G, and H, respectively. The amplitudes of the signal segments generated through feature visualization resembles those of each corresponding real signal segment.

4. Discussion

In clinical trials of electrophysiologic diagnosis, the electromyographer records nEMG signals from various muscles and considers the signal characteristics along with muscle locations. Significant information regarding the patient's pathology is nested within the signals and muscle types, which must be extracted to successfully

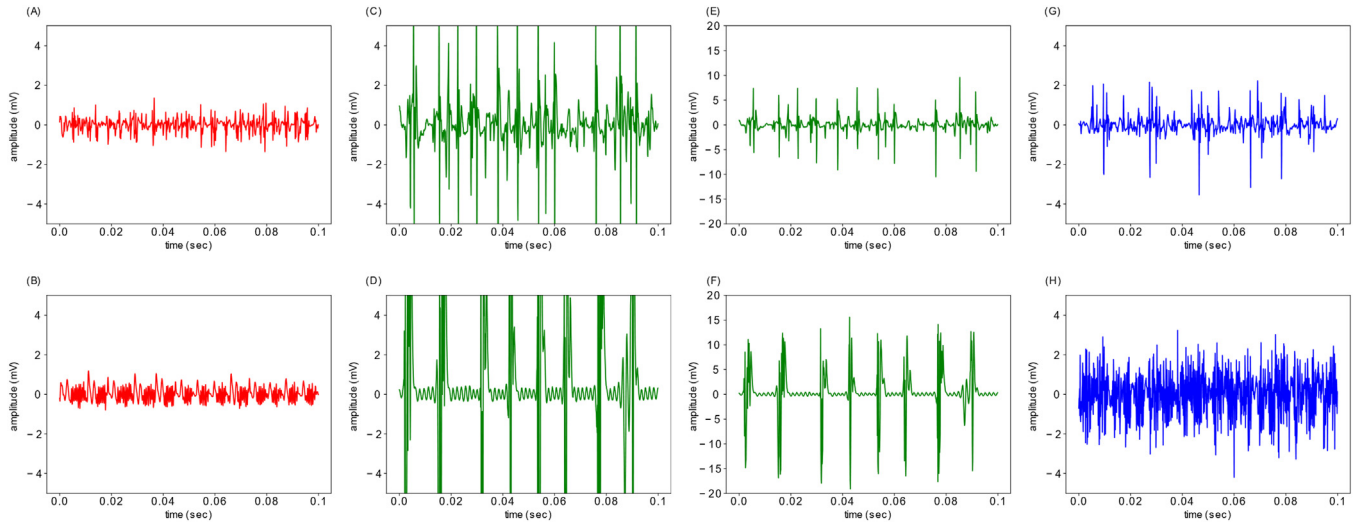

Fig. 7. Learned features of nEMGNet. The top row contains real signals, and the bottom row contains signals generated through feature visualization. nEMGNet-B from the 1st fold was used to plot the figure. (a) Real myopathic signal, (b) generated myopathic signal, (c) real neuropathic signal, (d) generated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

predict the diagnosis label of the given subject. This study proposed nEMGNet, a one-dimensional CNN suitable for nEMG signal classification, to extract features from raw signals. Four types of nEMGNet architectures were evaluated, and the results demonstrated that nEMGNet could learn signal characteristics of each diagnosis label. Comparing nEMGNet with models from previous works showed that nEMGNet outperformed all previous models. Additionally, the DiVote algorithm was introduced to mitigate data heterogeneity in subjects and leverage relevant information within proximal and distal muscle types. The DiVote algorithm could handle data heterogeneity and adjust bias in subject features, which improved subject diagnosis prediction performance compared to the case with no additional classifier. Lastly, the learned features of nEMGNet were visualized by applying feature visualization to a trained nEMGNet, in which the features resembled the signal segments from each diagnosis label.

4.1. Attributes of nEMGNet

Throughout the DiVote algorithm, values corresponding to normal subjects such as the signal segment classification result (Fig. 5), muscle signal prediction scores (Fig. 6A), and soft voted subject features (Fig. 6B) were inclined to have impartial prediction results. Likewise, it was observed from filtered signal segments (Fig. 4F) that normal signals lie between myopathic and neuropathic signals in latent space, and that the latent features are not distinctly clustered for each diagnosis label. This is attributed to normal signal segments having similar latent features with myopathic and neuropathic signals. Some signals labeled as myopathy or neuropathy exhibit signal characteristics of normal muscles since not all muscles from myopathic or neuropathic subjects show typical pathologic signal characteristics. Consequently, signal segments labeled as myopathy or neuropathy contain signal characteristics of normal subjects, while signals from normal subjects do not contain any features which indicate myopathy or neuropathy.

This imbalance in feature distribution among different labels caused the nEMGNet to classify normal signals more ambiguously than other labels, as it was trained to classify normal features as myopathy and neuropathy as well. Note that this phenomenon was also identified from the signal segment classification result of baseline models (Supplementary Figure 1-6). Additionally, the signal amplitude is the largest for neuropathic signals, medium for normal signals, and smallest for myopathic signals, which could also induce intermediate prediction scores of normal signals. Furthermore, the signal segment classification performance was evaluated using the test set, which contains noisily labeled signal segments, lowering the prediction accuracy of signal segments. These overlapping labels in terms of signal segment features produced a signal segment prediction accuracy of 52.26% for normal signal segments (Fig. 5), which is the lowest among the three diagnosis labels.

Nevertheless, the loss function of nEMGNet converged to a value close to zero during training, indicating that the model was fully capable of learning features from the data. Moreover, filtering the signal segment through deeper layers of nEMGNet resulted in distant groups in latent space (Fig. 4), implying that nEMGNet had learned to distinguish signals of different diagnosis labels. Therefore, the relatively ambiguous signal segment accuracy of normal labels was explained by the signal segment data with ambiguous labels consisting of the train and test set, not by the capacity of nEMGNet to learn relevant features. The signal segment prediction results were still useful as long as they were sufficiently divergent to generate distinct subject features, since this study primarily aimed to classify the subjects instead of the signal segments.

Using 1D-CNN presents several advantages over image classification models in signal processing applications, as indicated by the baseline experiment results. The 1D-CNN computes tensors and gradients in a one-dimensional direction, whereas the image classification models operate in two dimensions. This significantly reduces training and inference times under similar number of model parameters (Table 6). In addition, less runtime memory is consumed during training and inference since gradients and hidden layer activation tensors are smaller [54]. The number of parameters within a deep learning model is also smaller in 1D-CNN under identical number of layers and kernel size (Table 6). Furthermore, optimizing the neural architecture in image classification models is difficult because the search space is larger due to the greater number of hyperparameters, such as the width and height of the convolution kernel.

Additionally, the higher performance presented by nEMGNet over the image classification models could be attributed to its proximity to raw signals. When a preprocessing operation is ap-

plied to a signal, the information content is bound to be equal or smaller due to data processing inequality [43]. Preprocessing operations must be carefully applied to prevent the loss of important features within the data. However, conventional signal features such as the mel-spectrogram or discrete wavelet transform require manual selection of preprocessing hyperparameters. In contrast, nEMGNet makes predictions based on raw signals. A convolution operation with a linear filter is equivalent to frequency filtering [55,56], and training linear convolutional layers optimizes a frequency filter to extract features from a given signal. The nEMGNet can similarly be interpreted as a nonlinear filter based on 1D-CNN, which optimizes its filters to directly extract the features from raw signals. The 1D-CNN is also implemented for other signal applications in medicine [57,58], which implies the potential of leveraging 1D-CNN in signal processing applications.

4.2. Significance of DiVote algorithm

There are several points of significance in terms of the DiVote algorithm. First, the DiVote algorithm is a robust pipeline for subjects with heterogeneous data structures. A pipeline that converts the heterogeneous data into a homogeneous form is required when subjects possess different types and numbers of signals which also vary in length. The division, feature extraction, and soft voting process successfully extract and integrate meaningful features from varying data structures in a homogeneous manner.

Second, the DiVote algorithm can classify subjects with inconsistent features. The features and labels are not well labeled in terms of signal segments because all muscle signals from the same subject are labeled as the diagnosis of the subject, as explained earlier. Inaccurately labeled signal segments limit the nEMGNet from learning acute features, reducing the prediction accuracy of signal segments. This limitation was mitigated by using the DiVote algorithm since a classifier makes subject diagnosis predictions based on subject features. While the signal segment prediction scores and the subject features are impartial, the subject features for each diagnosis label are distinguishable in feature space (Fig. 6C). The additional classifier from the DiVote algorithm shifts the decision boundary of the subject features to better classify the subject diagnosis labels (Fig. 6D). Therefore, the DiVote algorithm successfully classifies the given subjects despite the ambiguous features of muscle signals.

Furthermore, generating the subject features from the DiVote algorithm can prevent undesired inductive bias in the deep learning model. Metadata such as the proximal or distal muscle information could be used as a feature vector to train the deep learning model [59,60], but the model may learn irrelevant features from the metadata. For instance, if the muscle type information was used as a feature vector in nEMGNet, the model may predict proximal muscles as non-normal regardless of the signal segment characteristics, because of the small number of proximal muscle signals in normal subjects (Table 1). Applying metadata in the form of subject features may enable the feature extractor to learn the desired features and leverage the information from the metadata, thus improving the task performance.

In medical applications, multiple measurements are often recorded for a subject in different time and space, generating heterogeneous data structures. For instance, longitudinal data are created when a subject visits a medical institution at different points in time, or various nEMG measurements are recorded from different muscles in the electrophysiologic diagnosis of neuromuscular disorders. While homogeneous data structure is preferable in developing machine learning models, data with heterogeneous structure is inevitable. In such medical applications where the data structure is heterogeneous within a given subject [61,62], the DiVote algorithm may leverage all data from the subject and integrate the heterogeneity to predict the diagnosis of the subject. In addition, the metadata may be used to generate new types of subject features, which could improve task performance.

4.3. Learned features of nEMGNet

Fig. 7 shows that the learned signal features of nEMGNet have a similar magnitude as the signal segments of corresponding diagnosis labels. In addition, the signal characteristics of each diagnosis label were similar to the feature visualization results. Myopathic signals are characterized by small amplitude and higher frequency [11], similar to the augmented signal segment in Fig. 7B. Neuropathic signals are characterized by large amplitude and lower frequency [11], which was similar to the augmented signal segment in Fig. 7D and F. Learned features of normal signals were slightly different from real normal signals because the latent features for normal signals were not as concise as those of myopathic or neuropathic signals. Meticulous details of real signals were not simulated in feature visualization results, indicating that very fine details were unnecessary to distinguish the normal signals from myopathic or neuropathic signals. Heuristic analysis conducted by electromyographers suggested that the signals generated for myopathy and neuropathy exhibit typical pathologic characteristics. These results demonstrated that nEMGNet had learned the desired signal characteristics of each diagnosis type, addressing the causality of the deep learning model [63–65].

4.4. Limitation and future works

There are a few limitations of this study. First, only 57 subjects were analyzed, which was not sufficient to validate the evaluation results. As deep learning model requires abundant high quality data to train up to its full potential [66–68], larger clinical nEMG dataset may improve the subject diagnosis prediction performance of the nEMGNet over the 83.69% accuracy measured in this study. Additionally, while the small difference between the two learning curves of the nEMGNet indicate that the model was able to learn generalizable features [69] (Supplementary Figure 7), k-fold cross-validation was performed due to the small subject size which constricted greater generalizability of the model. Therefore, additional nEMG subjects must be collected to better train the nEMGnet and further validate the study. Second, this study only focused on distinguishing crude categories of neuromuscular disorders. However, several subtypes of myopathies include inflammatory myositis, myotonic dystrophy, muscular dystrophies, and congenital myopathies. While these subtypes share common major myopathic motor unit action potential (MUAP) features, they also exhibit subtle differences in MUAP [11]. Similarly, there are various subtypes of neuropathies, including compressive neuropathies, Guillain-Barre syndrome, radiculopathies, and amyotrophic lateral sclerosis, which share common neurogenic MUAP features but exhibit subtle differences [11]. In the future, the proposed method may be able to perform a detailed diagnosis based on nEMG patterns if a larger nEMG dataset including various subtype labels is available.

5. Conclusion

This study introduced nEMGNet, a 1D-CNN model to extract features from raw nEMG signals, and the DiVote algorithm to mediate the heterogeneous data structure in differentiation between myopathy, neuropathy, and normal subjects. The proposed method was verified using single-channel needle electromyography signals of 57 subjects from Seoul National University Hospital. The nEMGNet and DiVote algorithm achieved a subject diagnosis prediction accuracy of 83.69%, outperforming all other base-

line nEMG deep learning models. The subject diagnosis prediction performance was increased by using an additional classifier in the DiVote algorithm, which shifts the decision boundary to mitigate bias in subject features. In addition, leveraging muscle type metadata improved subject diagnosis prediction performance. The typical signal segment features that the model predicts as each diagnosis type were identified using feature visualization, and it was verified that the model was making predictions based on valid features. This study contributed to the application of deep learning in nEMG electrophysiologic diagnosis by improving feature extraction performance, introducing a method to mediate the heterogeneous data structure of subjects, and suggesting the causality of the deep learning model. In future works, the proposed method may be utilized in real electrophysiologic diagnosis of neuromuscular disorders when nEMG data from sufficient subjects are acquired to verify the performance stability further.

Conflict of interest statement

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

Declaration of competing interest

The authors declare no known competing financial interests or personal relationships to report.

CRediT authorship contribution statement

Jaesung Yoo: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Ilhan Yoo:** Conceptualization, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ina Youn:** Data curation, Writing – review & editing. **Sung-Min Kim:** Resources, Data curation. **Ri Yu:** Writing – review & editing. **Kwangsoo Kim:** Resources, Writing – review & editing. **Keewon Kim:** Conceptualization, Resources, Writing – review & editing, Supervision. **Seung-Bo Lee:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Acknowledgments

KK and SL are co-corresponding authors. This work was supported by the Technology Innovation Program (20016225, Development and dissemination of Standard Reference Data), funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). This work was also supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 202011B23). The sponsor had no involvement in the study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2022.107079.

References

[1] G. Said, Diabetic neuropathy—a review, Nature Clin. Practice Neurol. 3 (2007) 331–340, doi:10.1038/ncpneuro0504.
[2] E.H. Bae, M.K. Greenwald, A.G. Schwartz, Chemotherapy-induced peripheral neuropathy: mechanisms and therapeutic avenues, Neurotherapeutics (2021) 1–13, doi:10.1007/s13311-021-01142-2.
[3] A.G. Smith, J.R. Singleton, Idiopathic neuropathy, prediabetes and the metabolic syndrome, J. Neurol. Sci. 242 (2006) 9–14, doi:10.1016/j.jns.2005.11.020.
[4] A.G. Smith, K. Rose, J.R. Singleton, Idiopathic neuropathy patients are at high risk for metabolic syndrome, J. Neurol. Sci. 273 (2008) 25–28, doi:10.1016/j.jns.2008.06.005.
[5] J.C. Deenen, P.A. van Doorn, C.G. Faber, A.J. van der Kooi, J.B. Kuks, N.C. Notermans, L.H. Visser, C.G. Horlings, J.J. Verschuuren, A.L. Verbeek, The epidemiology of neuromuscular disorders: age at onset and gender in the Netherlands, Neuromuscul. Disord. 26 (2016) 447–452, doi:10.1016/j.nmd.2016.04.011.
[6] P. Hafner, R. Phadke, A. Manzur, R. Smith, S. Jaiser, P. Schutz, C. Sewry, F. Muntoni, M. Pitt, Electromyography and muscle biopsy in paediatric neuromuscular disorders–Evaluation of current practice and literature review, Neuromuscul. Disord. 29 (2019) 14–20, doi:10.1016/j.nmd.2018.10.003.
[7] J.R. Daube, D.I. Rubin, Needle electromyography, Muscle Nerve 39 (2009) 244–270, doi:10.1002/mus.21180.
[8] K.R. Mills, The basics of electromyography, J. Neurol. Neurosurg. Psychiatry 76 (2005) ii32–ii35, doi:10.1136/jnnp.2005.069211.
[9] D.I. Rubin, Needle electromyography: basic concepts, Handb. Clin. Neurol. 160 (2019) 243–256, doi:10.1016/B978-0-444-64032-1.00016-3.
[10] R.G. Whittaker, The fundamentals of electromyography, Pract. Neurol. 12 (2012) 187–194, doi:10.1136/practneurol-2011-000198.
[11] J. Kimura, Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice, Oxford university press, 2013, doi:10.1093/med/9780199738687.001.0001.
[12] S.J. Oh, Clinical Electromyography: Nerve Conduction Studies, Lippincott Williams & Wilkins, 2003, doi:10.1046/j.1468-1331.2003.00654.x.
[13] J.M. Torpy, J.L. Kincaid, R.M. Glass, Peripheral Neuropathy, JAMA 299 (2008) 1096 -1096, doi:10.1001/jama.299.9.1096.
[14] F. Leblhuber, F. Reisecker, H. Boehm-Jurkovic, A. Witzmann, E. Deisenhammer, Diagnostic value of different electrophysiologic tests in cervical disk prolapse, Neurology 38 (1988) 1879 -1879, doi:10.1212/WNL.38.12.1879.
[15] R. Kendall, R.A. Werner, Interrater reliability of the needle examination in lumbosacral radiculopathy, Muscle Nerve 34 (2006) 238–241, doi:10.1002/mus.20554.
[16] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, doi:10.1038/nature14539.
[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, Jakob Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv preprint, 2020 arXiv:2010.11929.
[18] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, Deep speech 2: End-to-end speech recognition in english and mandarin, in: International conference on machine learning, 2016, pp. 173–182.
[19] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, Language models are few-shot learners, arXiv preprint, 2020 arXiv:2005.14165.
[20] O.P. Idowu, A.E. Ilesanmi, X. Li, O.W. Samuel, P. Fang, G. Li, An integrated deep learning model for motor intention recognition of multi-class EEG Signals in upper limb amputees, Comput. Methods Programs Biomed. 206 (2021) 106121, doi:10.1016/j.cmpb.2021.106121.
[21] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, Comput. Methods Programs Biomed. 161 (2018) 1–13, doi:10.1016/j.cmpb.2018.04.005.
[22] U. Erdenebayar, Y.J. Kim, J.-U. Park, E.Y. Joo, K.-J. Lee, Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram, Comput. Methods Programs Biomed. 180 (2019) 105001, doi:10.1016/j.cmpb.2019.105001.
[23] A. Bahador, M. Yousefi, M. Marashi, O. Bahador, High accurate lightweight deep learning method for gesture recognition based on surface electromyography, Comput. Methods Programs Biomed. 195 (2020) 105643, doi:10.1016/j.cmpb.2020.105643.
[24] C. Fricke, J. Alizadeh, N. Zakhary, T.B. Woost, M. Bogdan, J. Classen, Evaluation of three machine learning algorithms for the automatic classification of EMG patterns in gait disorders, Front. Neurol. 12 (2021) 666458, doi:10.3389/fneur.2021.666458.
[25] R.A. Zanini, E.L. Colombini, M.C.F.d. Castro, Parkinson's disease EMG signal prediction using neural networks, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 2446–2453, doi:10.1109/SMC.2019.8914553.
[26] K. Rezaee, S. Savarkar, X. Yu, J. Zhang, A hybrid deep transfer learning-based approach for Parkinson's disease classification in surface electromyography signals, Biomed. Signal Process. Control 71 (2022) 103161, doi:10.1016/j.bspc.2021.103161.
[27] H. Nodera, Y. Osaki, H. Yamazaki, A. Mori, Y. Izumi, R. Kaji, Classification of needle-EMG resting potentials by machine learning, Muscle Nerve 59 (2019) 224–228, doi:10.1002/mus.26363.
[28] E. Gokgoz, A. Subasi, Comparison of decision tree algorithms for EMG signal classification using DWT, Biomed. Signal Process. Control 18 (2015) 138–144, doi:10.1016/j.bspc.2014.12.005.
[29] M. Kefalas, M. Koch, V. Geraedts, H. Wang, M. Tannemaat, T. Bäck, Automated machine learning for the classification of normal and abnormal electromyography data, in: 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1176–1185, doi:10.1109/BigData50022.2020.9377780.
[30] M.U. Khan, S. Aziz, M. Bilal, M.B. Aamir, Classification of EMG signals for assessment of neuromuscular disorder using empirical mode decomposition and logistic regression, in: 2019 International Conference on Applied and En-

gineering Mathematics (ICAEM), 2019, pp. 237–243, doi:10.1109/ICAEM.2019.8853684.
[31] A. Subasi, E. Yaman, Y. Somaily, H.A. Alynabawi, F. Alobaidi, S. Altheibani, Automated EMG signal classification for diagnosis of neuromuscular disorders using DWT and bagging, Procedia Comput. Sci. 140 (2018) 230–237, doi:10.1016/j.procs.2018.10.333.
[32] H. Nodera, Y. Osaki, H. Yamazaki, A. Mori, Y. Izumi, R. Kaji, Deep learning for waveform identification of resting needle electromyography signals, Clin. Neurophysiol. 130 (2019) 617–623, doi:10.1016/j.clinph.2019.01.024.
[33] S. Nam, M.K. Sohn, H.A. Kim, H.-J. Kong, I.-Y. Jung, Development of artificial intelligence to support needle electromyography diagnostic analysis, Healthcare Inf. Res. 25 (2019) 131–138, doi:10.4258/hir.2019.25.2.131.
[34] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, M. Gabbouj, 1-d convolutional neural networks for signal processing applications, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8360–8364, doi:10.1109/ICASSP.2019.8682194.
[35] J. Amann, A. Blasimme, E. Vayena, D. Frey, V.I. Madai, Explainability for artificial intelligence in healthcare: a multidisciplinary perspective, BMC Med. Inf. Decis. Making 20 (2020) 1–9, doi:10.1186/s12911-020-01332-6.
[36] S.M. Lauritsen, M. Kristensen, M.V. Olsen, M.S. Larsen, K.M. Lauritsen, M.J. Jørgensen, J. Lange, B. Thiesson, Explainable artificial intelligence model to predict acute critical illness from electronic health records, Nat. Commun. 11 (2020) 1–11, doi:10.1038/s41467-020-17431-x.
[37] J. Yoo, T.J. Jun, Y.-H. Kim, xECGNet: Fine-tuning attention map within convolutional neural network to improve detection and explainability of concurrent cardiac arrhythmias, Comput. Methods Programs Biomed. 208 (2021) 106281, doi:10.1016/j.cmpb.2021.106281.
[38] M. Pérez-Pelegrí, J.V. Monmeneu, M.P. López-Lereu, L. Pérez-Pelegrí, A.M. Maceira, V. Bodí, D. Moratal, Automatic left ventricle volume calculation with explainability through a deep learning weak-supervision methodology, Comput. Methods Programs Biomed. 208 (2021) 106275, doi:10.1016/j.cmpb.2021.106275.
[39] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, K.-R. Müller, Unmasking clever Hans predictors and assessing what machines really learn, Nat. Commun. 10 (2019) 1–8, doi:10.1038/s41467-019-08987-4.
[40] C.J. Anders, L. Weber, D. Neumann, W. Samek, K.-R. Müller, S. Lapuschkin, Finding and removing clever hans: using explanation methods to debug and improve deep models, Inf. Fusion 77 (2022) 261–295, doi:10.1016/j.inffus.2021.07.015.
[41] C. Olah, A. Mordvintsev, L. Schubert, Feature visualization, Distill 2 (2017) e7, doi:10.23915/distill.00007.
[42] D. Erhan, Y. Bengio, A. Courville, P. Vincent, Visualizing Higher-Layer Features of a Deep Network, 1341, University of Montreal, 2009, p. 1.
[43] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, 2005, doi:10.1002/047174882X.
[44] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint, 2014 arXiv:1409.1556.
[45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.
[46] A. Zaeemzadeh, N. Rahnavard, M. Shah, Norm-preservation: why residual networks can become extremely deep? in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, doi:10.1109/TPAMI.2020.2990339.
[47] J.Y. Yam, T.W. Chow, A weight initialization method for improving training speed in feedforward neural network, Neurocomputing 30 (2000) 219–232, doi:10.1016/S0925-2312(99)00127-7.
[48] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations, 2015.
[49] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[51] J. Gorodkin, Comparing two K-category assignments by a K-category correlation coefficient, Comput. Biol. Chem. 28 (2004) 367–374, doi:10.1016/j.compbiolchem.2004.09.006.
[52] C.M. Bishop, Pattern recognition, Mach. Learn. 128 (2006).
[53] L. McInnes, J. Healy, J. Melville, Umap: Uniform Manifold Approximation and Projection for Dimension Reduction, arXiv preprint, 2018 arXiv:1802.03426.
[54] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: a survey, Mech. Syst. Sig. Process. 151 (2021) 107398, doi:10.1016/j.ymssp.2020.107398.
[55] E.C. Ifeachor, B.W. Jervis, Digital Signal Processing: A Practical Approach, Pearson Education, 2002.
[56] A.V. Oppenheim, A.S. Willsky, S.H. Nawab, G.M. Hernández, Signals & Systems, Pearson Educación, 1997.
[57] U.R. Acharya, H. Fujita, O.S. Lih, Y. Hagiwara, J.H. Tan, M. Adam, Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inf. Sci. 405 (2017) 81–90, doi:10.1016/j.ins.2017.04.012.
[58] C.D. Galloway, A.V. Valys, J.B. Shreibati, D.L. Treiman, F.L. Petterson, V.P. Gundotra, D.E. Albert, Z.I. Attia, R.E. Carter, S.J. Asirvatham, Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram, JAMA Cardiol. 4 (2019) 428–436, doi:10.1001/jamacardio.2019.0640.
[59] C. Xiong, X. Zhao, D. Tang, K. Jayashree, S. Yan, T.-K. Kim, Conditional convolutional neural network for modality-aware face recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3667–3675.
[60] I.-C. Yoo, K. Lee, S. Leem, H. Oh, B. Ko, D. Yook, Speaker anonymization for personal information protection using voice conversion techniques, IEEE Access 8 (2020) 198637–198645, doi:10.1109/ACCESS.2020.3035416.
[61] G. Lee, B. Kang, K. Nho, K.-A. Sohn, D. Kim, MildInt: deep learning-based multimodal longitudinal data integration framework, Front. Genetic. 10 (2019) 617, doi:10.3389/fgene.2019.00617.
[62] L. Gao, H. Pan, F. Liu, X. Xie, Z. Zhang, J. Han, A.s.D.N. Initiative, Brain disease diagnosis using deep learning features from longitudinal MR images, in: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data, 2018, pp. 327–339, doi:10.1007/978-3-319-96890-2_27.
[63] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115, doi:10.1016/j.inffus.2019.12.012.
[64] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (xai): toward medical xai, in: IEEE Transactions on Neural Networks and Learning Systems, 2020, doi:10.1109/TNNLS.2020.3027314.
[65] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller, Causability and explainability of artificial intelligence in medicine, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 9 (2019) e1312, doi:10.1002/widm.1312.
[66] E.M. Cahan, T. Hernandez-Boussard, S. Thadaney-Israni, D.L. Rubin, Putting the data before the algorithm in big data addressing personalized healthcare, NPJ Digit. Med. 2 (2019) 1–6, doi:10.1038/s41746-019-0157-2.
[67] Y. Hu, J. Jacob, G.J. Parker, D.J. Hawkes, J.R. Hurst, D. Stoyanov, The challenges of deploying artificial intelligence models in a rapidly evolving pandemic, Nature Mach. Intelligence 2 (2020) 298–300, doi:10.1038/s42256-020-0185-2.
[68] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[69] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of Machine Learning, The MIT Press, 2018.