Clinical Neurophysiology 159 (2024) 41–55

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Artificial intelligence for automatic classification of needle EMG signals: A scoping review

S. de Jonge [a], W.V. Potters [a,b], C. Verhamme [a,*]

[a] Department of Neurology and Clinical Neurophysiology, Amsterdam Neuroscience, Amsterdam UMC Location University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
[b] TrianecT, Padualaan 8, Utrecht, The Netherlands

HIGHLIGHTS

- In clinical needle EMG interpretation, conventional machine learning methods were more often studied than deep learning methods.
- Most studies developed models for classification of contraction signals and a few addressed the classification of rest signals.
- Current research has methodological shortcomings, which prevent clinical implementation, but are surmountable.

ARTICLE INFO

Article history:
Accepted 16 December 2023
Available online 3 January 2024

Keywords:
Needle electromyography
Neuromuscular disorders
Artificial intelligence
Machine learning
Classification

ABSTRACT

Objective: This scoping review provides an overview of artificial intelligence (AI), including machine and deep learning techniques, in the interpretation of clinical needle electromyography (nEMG) signals.
Methods: A comprehensive search of Medline, Embase and Web of Science was conducted to find peer-reviewed journal articles. All papers published after 2010 were included. The methodological quality of the included studies was assessed with CLAIM (checklist for artificial intelligence in medical imaging).
Results: 51 studies were identified that fulfilled the inclusion criteria. 61% used open-source EMGlab data set to develop models to classify nEMG signal in healthy, amyotrophic lateral sclerosis (ALS) and myopathy (25 subjects). Only two articles developed models to classify signals recorded at rest. Most articles reported high performance accuracies, but many were subject to bias and overtraining.
Conclusions: Current AI-models of nEMG signals are not sufficient for clinical implementation. Suggestions for future research include emphasizing the need for an optimal training and validation approach using large datasets of clinical nEMG data from a diverse patient population.
Significance: The outcomes of this study and the suggestions made aim to contribute to developing AI-models that can effectively handle signal quality variability and are suitable for daily clinical practice in interpreting nEMG signals.

1. Introduction

Needle electromyography (nEMG) is crucial for the identification and localisation of neuromuscular disorders (Kim et al., 2019, Preston and Shapiro, 2021, Rubin, 2019). During nEMG a needle is placed in the muscle to record the activity with high spatial resolution, in resting and contracting states. Pathological waveforms present at rest and changes in motor unit action potentials (MUAP) during various levels of contraction provide diagnostic information upon a defect in the neuromuscular control pathway and thus, neuromuscular disorders (Preston and Shapiro, 2021).

The interpretation of nEMG signals is a complex task. In clinical practice, evaluation of nEMG recordings is based on audio-visual interpretation during investigation. The various sounds of the phenomena that can be heard at rest, during voluntary contraction and needle movement may resemble everyday sounds, e.g., a fibrillation potential sounds like a ticking clock, end-plate noise sounds like a seashell and a complex repetitive discharge sounds like a jackhammer (Preston and Shapiro, 2021, Rubin, 2019). Interpretation of the nEMG signals requires a lot of training and expertise to accurately recognise pathologies. However, even with adequate training and expertise, the interpretation of nEMG signals remains subjective, leading to limited agreement among different raters.

* Corresponding author at: Department of Neurology and Clinical Neurophysiology Amsterdam Neuroscience, Amsterdam UMC Location University of Amsterdam Meibergdreef 9, Amsterdam, The Netherlands.
E-mail address: c.verhamme@amsterdamumc.nl (C. Verhamme).

https://doi.org/10.1016/j.clinph.2023.12.134
1388-2457/

Studies have shown that depending on type of (pathological) waveform or level of training the interrater agreements are found to be between 47 and 91% (Kendall and Werner, 2006, Narayanaswami et al., 2016). These studies focused on the interrater agreement for the final diagnosis of radiculopathy, in which they also examined the agreement on the presence or absence of abnormal spontaneous activity and of reinnervation without standardized criteria.

The development of standardized (semi-) automated methods holds the potential to enhance the interpretation of nEMG signals and to reduce interrater variability. It is challenging to define adequate reference standards, which may be at recording, muscle, and patient-level. This complexity needs careful consideration as the aim is to improve the current diagnostic process of neuromuscular disorders.

Quantitative nEMG has been described previously in a review focusing on quantitative MUAP characterization and classification (Farkas et al., 2010). Various techniques have been employed to classify MUAPs using features in time domain, frequency domain and with time–frequency analysis as well as unsupervised-learning using self-organising feature maps (Christodoulou and Pattichis, 1999, Pattichis and Elia, 1999, Pattichis and Pattichis, 1999). Probabilistic models using an aggregation of MUAP characteristics were most successful in discriminating between healthy, neuropathic and myopathic muscle states (Farkas et al., 2010).

Recent studies have taken advantage of the huge improvements in the field of artificial intelligence (AI), including the introduction of deep learning methods, to improve waveform classification. In deep learning, hand-crafted feature selection by an expert, which is part of conventional machine learning, is not needed. Instead, features are directly obtained from the input data, removing effects of bias and level of experience of the human interpreter. Although there have been numerous publications on AI-based nEMG classification, to our knowledge, it has not yet been implemented in clinical practice. The aim of this scoping review is to present an overview of AI techniques to classify clinical nEMG signals, both at rest and during contraction, to discuss why AI has not arrived in clinical practice and to look at the future of AI application in clinical nEMG classification.

2. Materials and methods

This scoping review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018). No review protocol was registered for this study.

2.1. Search strategy

The literature search was conducted in February 2023 to identify relevant literature in English in three major health and sciences databases, i.e., Medline, Embase and Web of Science. The search strategy was defined in conjunction with an experienced librarian. The final search term included items describing the constructs "electromyography" and "artificial intelligence" (Table 1). A broad search was conducted because most articles do not specify in key words or title/abstract that nEMG data was used. Items describing "surface EMG" were excluded to narrow the search results. The complete overview of the search terms as used for the different databases is available in Supplementary Material 1, Appendix A. In addition to searching the databases, the reference lists of included articles were screened to identify additional relevant studies.

2.2. Selection criteria

Papers were included that implemented artificial intelligence to classify (human-derived) nEMG signals to support the diagnosis of neuromuscular disorders. Studies that focused primarily on model development and did not provide a clear description on the dataset used to evaluate the model, were excluded. Studies were only included when they were published from 2010 onwards. Around this time the field of AI changed substantially due to the combination of an increase in large datasets and computer power becoming more affordable (Theodoridis, 2015). The study selection criteria are summarised in Box 1.

Box 1 Inclusion and exclusion criteria.

Inclusion criteria

- Studies that use artificial intelligence in the classification of nEMG signals to support the diagnosis of neuromuscular disorders.
- nEMG signals derived from humans in a clinical setting.
- Published from 2010 onwards.
- Peer-reviewed original journal articles.

Exclusion criteria

- Studies that do not introduce a new model or do not repurpose an existing model.
- Studies that use different data sources (e.g., surface electrodes, simulation data, non-clinical data)
- Studies that have a different purpose (e.g., development of decomposition methods)
- Studies not in English language.
- Reviews, conference abstracts and proceedings, news, books, and protocols.

2.3. Selection process

Web-based Rayyan software was used to manage papers identified in the searches (Johnson and Phillips, 2018). Duplicates were removed and all papers were screened for eligibility by reading titles and abstracts. The selection criteria as previously stated were applied. All remaining papers were read in full, and inclusion and exclusion criteria were applied again. The selection process was conducted by one reviewer (SJ) and in case of uncertainty, two independent reviewers (WP, CV) were consulted.

Table 1
Search strategy in key words.

Search category	Search terms
(1) Electromyography	"Electromyography" OR "needle electromyography" OR "nEMG" OR "electromyogram" OR "muscle classification" OR "MUAP" or "motor unit"
(2) Artificial intelligence	"Artificial intelligence" OR "algorithms" OR "classification" OR "classifier" OR "deep learning" OR "machine learning" OR "automated reasoning" OR "neural network" OR "multi-classifier"
(3) Surface Electromyography	"Surface electromyography" OR "sEMG"
Combined	1 AND 2 NOT 3

2.4. Data extraction

A data extraction form was developed, shown in Supplementary Material 1, Appendix B. One reviewer (SJ) charted the data and two independent reviewers (WP, CV) were consulted in case of uncertainty. Data was extracted on study characteristics (e.g., authors, year of publication, country of origin), the purpose or aim(s) of the study, dataset characteristics (e.g., source of data, data acquisition and preprocessing methods, data population and included subjects, reference standard), and implementation of AI (e.g., type of model, feature extraction methods, feature selection methods, training and validation, model performance).

2.5. Quality assessment

Included studies were reviewed upon the quality of the documentation of methodologies in the paper. Clinical utility of AI (support) systems requires that the method is reproducible (e.g., definition of reference standard) and that the model was evaluated (e.g., with external validation). Studies were assessed through a widely accepted checklist in the field of Radiology: The Checklist for Artificial Intelligence in Medical Imaging (CLAIM) (Mongan et al., 2020). Although this checklist was developed as a guideline to aid authors and reviewers of AI in medical imaging, it also fits the objective of this review.

Papers were evaluated item-wise, so that implementation of items could be compared between papers.

3. Results

3.1. Search findings

A total of 8886 articles were identified. After duplicate removal, 5997 articles were screened and 5842 articles were excluded based on title and abstract. The 155 remaining full-text articles were read and another 106 articles were excluded as outlined in Fig. 1, leading to 49 included articles by search. An additional two articles were identified from reference lists of these articles. This resulted in a total of 51 included articles for this review.

3.2. Study characteristics and AI implementation

Table 2 provides a summary of the approach of the studies included in this review, covering the data source, whether preprocessing steps were applied, feature extraction methods, inputs for the model, type of model, classification task, training method and achieved performance. These steps, which are necessary in the development of AI models, are also visually shown in Fig. 2.

3.2.1. Data source

The functionality of the model is determined by the data source used for its training and development. The data source is briefly described in Table 2 and more details for each dataset are given in Table 3. Most (49/51) studies evaluated contraction nEMG signals. Neuropathy group was most frequently represented by patients with amyotrophic lateral sclerosis (ALS). As for the myopathy group, no particular population was represented except for one study that used patients with inclusion-body myositis (IBM) in the myopathy group (Tannemaat et al., 2023). Only two studies used data acquired during rest (Nodera et al., 2019a, 2019b).

The data sources were either open or closed source. Out of 51 studies, 30 (59%) used an open-source dataset, 18 (35%) acquired their own data (closed-source), and 3 (6%) combined their own data with an open-source dataset (Fig. 2a). Among the open-source datasets, the EMGLab dataset was used in 31 (60%) studies, while the PhysioNet dataset was used in 5 (10%) studies.

EMGLab dataset consists of recordings from 25 subjects: 10 healthy controls (aged 21–37 years), 8 ALS patients (aged 35–67 years) and 7 myopathy patients (aged 19–63 years) (Nikolic et al., 2001). No patients in the control group had a history nor showed signs of neuromuscular disorders. Only patients with clear diagnosis were included in the two patient groups. Six muscles (Abductor Pollicis Brevis, Biceps Brachii, Triceps Brachii, Deltoideus, Tensor Fasciae Latae, and Tibialis Anterior) were recorded at various insertion depths and locations during slight and constant contractions, resulting in a total of 938 recordings. One muscle could have up to fifteen recordings, each recording lasting 11.2 seconds, acquired at five different locations and three insertion levels. Recordings at biceps brachii were most frequent, with 442 recordings. Eighteen studies only used recordings from biceps brachii (Table 2) where only one study used all recordings in the dataset (Torres-Castillo et al., 2022).

The PhysioNet dataset consists of recordings from 3 subjects: a control patient (aged 44 years), a neuropathy patient (aged 62 diagnosed with L5 radiculopathy) and a myopathy patient (aged 57 years diagnosed with polymyositis). Recordings were made in tibialis anterior muscle for several seconds. Out of 5 articles that used this dataset, two used it as their primary data source (Bajaj and Kumar, 2015, Roy et al., 2022) and three used it alongside a different data source (Artameeyanant et al., 2016, Dubey et al., 2022, Kehri and R. N, 2018).

3.2.2. Preprocessing

The nEMG examination of a muscle consists of assessment of insertional and spontaneous activity at rest and of assessment of MUAPs during contraction. Needle movements may give artefacts. In the development of a classification model, it is necessary to implement a preprocessing step to extract the relevant data segments (Fig. 2b). In most datasets, this preprocessing step was already performed during data acquisition (Table 3). For example, recordings from EMGLab dataset were initiated when the signal quality was good and the patient were applying a constant level of contraction, resulting in pure contraction nEMG signals (Nikolic et al., 2001) (Fig. 2a). Other approaches are that annotation was performed, e.g., for the selection of spontaneous activity in resting state signals (Nodera et al., 2019a, 2019b), or that manual data selection was performed, e.g. to select contraction signals without needle movements (Tannemaat et al., 2023) (Fig. 2a).

Additional preprocessing steps were performed to prepare the data for analysis and training (Fig. 2b). These steps included reducing the known noise components in the data (e.g., with a 50 Hz notch filter), down sampling to reduce data size, automatic segmentation to increase training samples and/or reduce data size, decomposing the nEMG signal into, for example, MUAPs, and/or the creation of spectrograms.

3.2.3. Features

Features are manually extracted from the pre-processed nEMG data in conventional machine learning approaches. This step is not applicable in deep learning approaches as the features are automatically learned from the input signals. Among the studies included, 47/51 studies (92%) implemented machine learning techniques and 4/51 studies (8%) applied deep learning techniques. The feature extraction method for deep learning approaches is denoted as 'N/A' (not applicable) in features column in Table 2 as well as in the pie chart in Fig. 2c.

Feature extraction methods for machine learning approaches can generally be categorised into time domain, frequency domain and time–frequency analysis, these methods were used or a combination thereof (Fig. 2c). Not all methods fit into these categories,

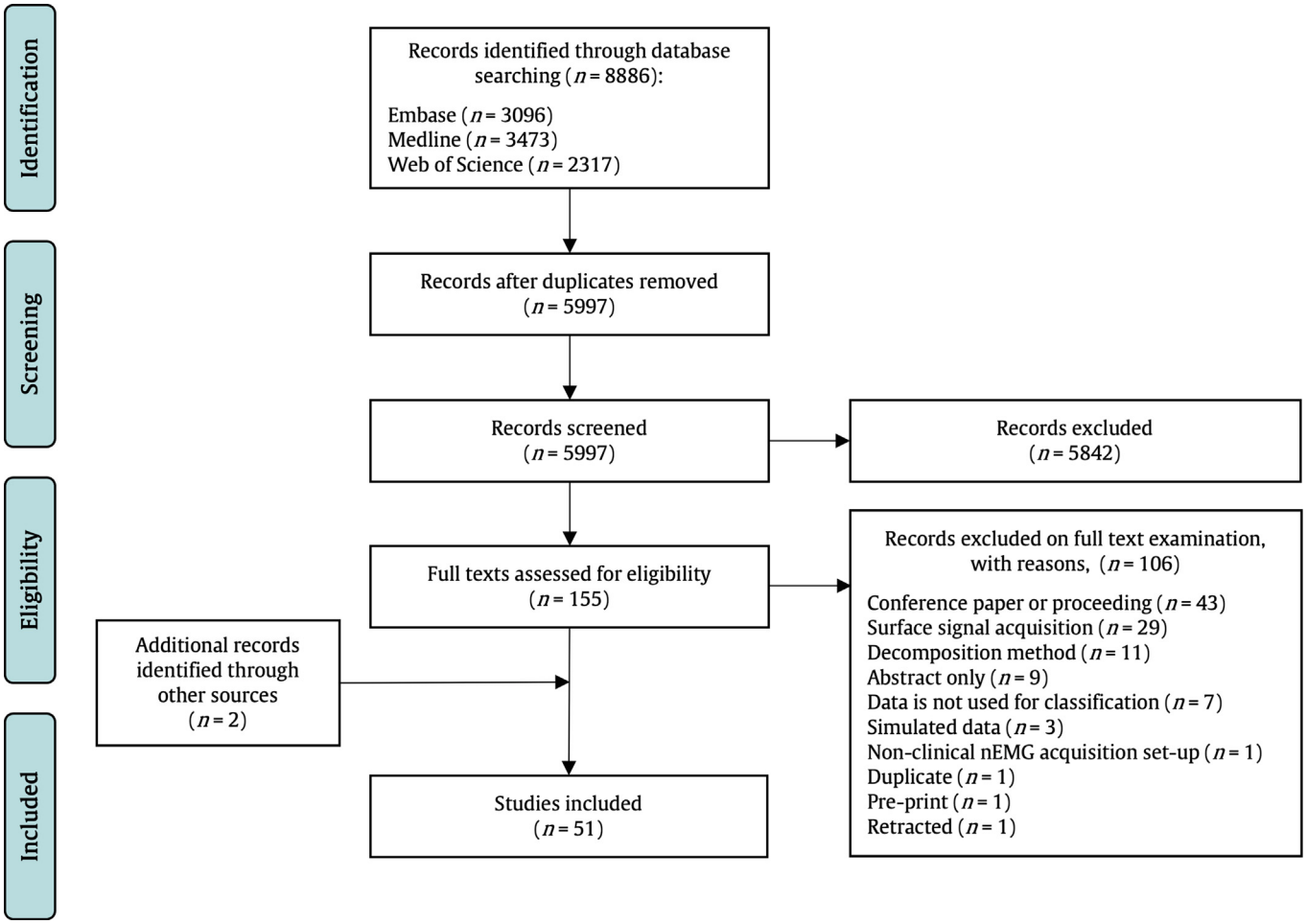

Fig. 1. PRISMA flow chart.

such as weighted visibility graph and 1-dimensional local binary pattern, recognized by 'undefined domain' in Fig. 2c. Features were most frequently acquired using time–frequency analysis methods, in 34/51 studies (67%). These features describe both time and frequency characteristics of a signal, relevant for non-stationary nEMG signal. Wavelet-based techniques, particularly the discrete wavelet transform (DWT), were employed in 18 of these studies, with Daubechies wavelet most frequently selected in 15 studies as mother wavelet. The DWT divides the signal into different frequency sub-bands using a set of wavelets derived from the mother wavelet. Time domain features were extracted in 17/51 studies (33%). These features, such as mean absolute value, zero crossing (number of times a waveform crosses from a positive amplitude to a negative amplitude), and root mean square (square root of the arithmetic mean of the squared samples amplitudes of a waveform), were derived from the nEMG signal during contraction or from MUAPs. Frequency domain features were extracted in 7/51 studies (14%), where methods such as power spectral density were used to derive relevant features. These features provide solely information about the frequency components in the nEMG signal.

Feature extraction is followed by feature selection and/ or reduction, to reduce dimensionality of the feature vector calculated over each nEMG segment. The number of features used is mainly determined by the amount of training data available to avoid overfitting. Many studies employed statistical methods, such as ANOVA and *t*-tests, to select the most discriminative features (Artameeyanant et al., 2016, Bose et al., 2021, Chatterjee et al., 2020, Chatterjee et al., 2019, Dubey et al., 2022, Roy et al., 2022). There are also studies that applied feature reduction methods, such as linear discriminant analysis (LDA) (Hazarika et al., 2019, Hazarika et al., 2018, Torres-Castillo et al., 2022). LDA aims to transform the original feature space into a new space that maximizes the separability of the different classes or groups.

3.2.4. Input

The input for machine learning based models are features, often calculated in fixed-length segments (Fig. 2d). The number of features may vary, as it may depend on the feature selection method. The duration of the segment over which the features were calculated is diverse and ranged from 175 ms to 11.2 s. A short duration means that more samples are available for training, however, when the nEMG signal is cut into very short segments, important time-related information may be lost.

In some studies, the nEMG signal was decomposed in MUAPs (Bose et al., 2021, Dobrowolski et al., 2012, Doulah et al., 2014, Kamali et al., 2014, Krishna and Thomas, 2015, Tomczykiewicz et al., 2012). For instance in Bose et al. (2021) MUAPs were extracted from nEMG signals with 175 ms duration using a template MUAP. The most dominant MUAP was selected from one segment containing multiple constituent MUAPs, resulting in a total of 7,200 dominant MUAPs extracted from 7,200 nEMG segments. The input of the model then consisted of features calculated over the dominant MUAPs and not over the unprocessed nEMG data.

The input of the deep learning models was either 1-dimensional nEMG time series (Yoo et al., 2022; Zhang et) or 2-dimensional

Table 2
Study characteristics and model implementation. Data source column shows open (O.X) and closed (C.X) data sources that are in more detail described in Table 3.

Author(s) (year)	Data source	Pre-pro	Features	Input	Model	Training partitions	Classification task	Performance (accuracy)	CLAIM
Zhang et al. (2023)	O.1 (a): ALS (n = 8), controls (n = 10) from BB C.1 (b): ALS (n = 189), controls (n = 133) from FDI, TA	Yes	N/A	33,113 segments with 0.25 s or 1 s duration (6,000 data samples)	Pre-training model for domain adaptation and pre-classification, followed by wavelet-based CNN	One dataset as training, one dataset as testing. Repeated experiments with Monte-Carlo simulations (five times)	Healthy vs. ALS: (T1) cross-muscle and cross-race (dataset a and b), (T2) cross-device (dataset b)	(T1) 78.9% (BB from dataset a as training set, AT from dataset b as test set), (T2) 75%	72%
Tannemaat et al. (2023)	C.2: ALS (n = 20), IBM (n = 20), controls (n = 25) from D, TA, BB, VM + 32 other muscles	Yes	TD, FD, TFA (tsfresh Python package, 794 features) + Boruta algorithm	n features × 380 segments with 5 s duration	RF	Tenfold CV, repeated five times with nested tenfold CV for hyperparameter optimization	Muscle-level and patient-level: (T1) healthy vs. ALS, (T2) IBM vs. ALS, (T3) IBM vs. healthy	(T1) 77.9% (muscle-level), (T2) 57.0% (patient-level), (T3) 68.4% (patient-level)	80%
Yoo et al. (2022)	C.3: neuropathy (n = 19), myopathy (n = 19), controls (n = 19) from distal and proximal muscles	Yes	N/A	8,070 segments with 0.4 s duration	1D-CNN followed by divide-and-vote algorithm to combine predictions, followed by LR	Fivefold CV, repeated three times	Healthy vs. neuropathy vs. myopathy using (T1) simple averaging, (T2) features-all + LR, (T3) features proximal–distal + LR	(T1) 75.4%, (T2) 81.9%, (T3) 83.7%	88%
Roy et al. (2022)	O.2: neuropathy (n = 1), myopathy (n = 1), controls (n = 1) from TA	Yes	TFA (DWT-Daubechies4) + ANOVA	n features × 310 segments with 0.25 s duration	SVM	Fivefold CV	(T1) Healthy vs. myopathy, (T2) Healthy vs. neuropathy, (T3) Healthy vs. disease	(T1) 99.2%, (T2) 100%, (T3) 99.3%	52%
Jose et al. (2022)	O.1: ALS (n = 8), myopathy (n = 7) from BB, D, VM	Yes	1D centre symmetric local binary pattern	15 features × NS segments with NS duration	kNN, SVM, DT, DA, NB, NN with by Boyer-Moore majority vote	Tenfold CV	ALS vs. myopathy in (T1) BB muscle, (T2) D, (T3) VM	(T1) 92.8%, (T2) 94.3%, (T3) 93.7%	66%
Samanta et al. (2022)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TFA (hyperbolic ST) + genetic algorithm	14 features × 12,275 segments with 175 ms duration	SVM / kNN / NB	Tenfold CV	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) ALS vs. myopathy, (T4) Healthy vs. disease	(T1) 96.8% (SVM), (T2) 99.5% (SVM), (T3) 98.8% (SVM), (T4) 98.5% (SVM)	55%
Torres-Castillo et al. (2022)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from APB, BB, D, TFL, TA, TB, VL, VM	Yes	TD and TFA (HHT) + uncorrelated LDA	2 features × 9,380 segments with 1.12 s duration	LDA / DT / kNN	Threefold CV, repeated 10 times	Healthy vs. ALS vs. myopathy using (T1) individual segments, (T2) total signal	(T1) 94.4% (EEMD and kNN) (T2) 99.4% (EEMD and kNN)	76%
Dubey et al. (2022)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, VM O.2: neuropathy (n = 1), myopathy (n = 1), controls (n = 1) from TA	Yes	TFA (Hilbert Transform) + t-test	2 features × 947 segments with 11.2 s duration and 310 segments with 0.25 s duration	Feed forward NN / SVM / DT	Train: 70%, validation: 15%, test: 15%	Healthy vs. ALS vs. myopathy	99.5% (feed forward NN)	69%
Bose et al. (2021)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	Weighted visibility graph + ANOVA	4 features × 7200 dominant mus retrieved from segments with 175 ms duration	NB / SVM / kNN / LDA / DT	Tenfold CV, repeated 10 times	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) ALS vs. myopathy, (T4) Healthy vs. ALS vs. myopathy	(T1) 98.1% (NB), (T2) 98.6% (NB), (T3) 99.5% (NB), (T4) 99.1% (NB)	58%
Bakiya et al. (2021b)	O.1: ALS (n = 8), controls (n = 10) from BB	No	TFA (ST, WVT, SET, STFT) + fractional firefly algorithm	15 features × 100 segments with 11.2 s duration	Fractional firefly NN with 3 to 20 hidden neurons	Train: 70%, test: 30%	Healthy vs. ALS	93.3% (15 hidden neurons)	59%

(*continued on next page*)

Table 2 (*continued*)

Author(s) (year)	Data source	Pre-pro	Features	Input	Model	Training partitions	Classification task	Performance (accuracy)	CLAIM
Bakiya et al. (2021a)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	No	TFA (ST, WVT, SET, STFT) features using gray level co-occurrence matrix + genetic algorithm	15 features × 150 segments with 11.2 s duration	Artificial NN	Train: 70%, test: 30%	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) ALS vs. myopathy, (T4) Healthy vs. ALS vs. myopathy	(T1) 100%, (T2) 86.8%, (T3) 96.2%, (T4) 82.2%	48%
Chatterjee et al. (2020)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	No	Multifractal detrended fluctuation analysis + ANOVA	5 features × 450 segments with 11.2 s duration	SVM / kNN / probabilistic NN / NB / LDA / RF	Tenfold CV	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) ALS vs. myopathy, (T4) Healthy vs. disease	(T1) 95.7% (PNN), (T2) 98.3% (SVM), (T3) 96.6% (kNN), (T4) 97.2% (RF)	59%
Mokdad et al. (2020)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	No	FD (bispectrum) and TFA (CWT)	4 features × *NS* segments with 11.2 s duration	SVM / kNN / LDA	Sixfold CV	Healthy vs. ALS	95.8% (SVM)	54%
(Jose et al., 2020a)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TFA (DWT tested with 71 different wavelets) + maximum likelihood estimation	n features × 180 segments with 11.2 s duration	ANN (13 different training algorithms)	Train: 60%, validation: 20%, test: 20%	Healthy vs. ALS vs. myopathy	86.5% (D11, Fletcher-Reeves update conjugate gradient)	62%
Jose et al. (2020b)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TD (fractal dimensions via Higuchi's Method, 1D LBP)	262 features × 2,250 segments with 1.24 s duration	LDA / FT / ESD / fine KNN / quadratic SVM / MLPNN with Boyer-Moore majority vote	Tenfold CV, repeated 25 times	Healthy vs. ALS vs. myopathy	99.9% (MLPNN)	69%
Kamali and Stashuk (2020)	PR: patient count unknown from D, VM, FDI, TA	Yes	TD, FD and TFA (DWT-Daubechies) + NDEC clustering	4 clusters (4 TD, 6 FD features) × 6576 Thundes from 322 segments with 15 s duration	MIL with adaptive neurofuzzy classifier (TSK) / RF / SVM	NS	Healthy vs. neuropathy vs. myopathy in (T1) TA muscle, (T2) FDI, (T3) DLT, (T4) VM	(T1) 100% (RF), (T2) 98.8% (RF, SVM), (T3) 97.9% (RF, SVM, TSK), (T4) 97.6% (SVM)	56%
Bakiya et al. (2020)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	No	TFA (SWT, WVT, ST, STFT) + particle swarm optimization with fractional velocity update	15 features × 150 segments with 11.2 s duration	NN (layers = 1 to 5 and neurons = 5 to 20)	Train: 70%, test: 30%	Healthy vs. ALS vs. myopathy	97.7% (4 layers and 5 neurons)	52%
Yaman and Subasi (2019)	C.5: neuropathy (n = 13), myopathy (n = 7), controls (n = 7) from BB	Yes	TFA (wavelet package decomposition)	6 features × n sub-bands × 2400 segments with 0.25 s duration	ANN / kNN / NB / REPTree / LADTree / C4.5 DT / RF + bagging and boosting	Tenfold CV, repeated 25 times	Healthy vs. ALS vs. myopathy	99.1% (RF with AdaBoost)	55%
Chatterjee et al. (2019)	O.1: ALS (n = 8), myopathy (n = 7) from BB	No	TFA (modified window ST optimized with PSO) + t-test	4 features × 350 segments with 11.2 s duration	SVM / kNN / NB / DT	Tenfold CV	Healthy vs. myopathy	98.6% (SVM)	57%
Nodera et al. (2019b)	C.6: 83 subjects from BB, FDI, VM, TA	Yes	N/A	330 / 20,000 / 200,000 Mel spectrograms retrieved from segments with 2 s duration	VGG16 / VGG19 / ResNet50 / ResNet152	Train: 80%, Test 20%	CRDs vs. endplate potentials vs. fasciculation potentials vs. fibrillation potentials/ PSW vs. myotonic discharges vs. noise artifact	94% (ResNet50 with data-augmentation to20,000 spectrograms)	74%
Nodera et al. (2019a)	C.6: 103 subjects from BB, FDI, VM, TA	Yes	TD, FD, TFA (INTERSPEECH 2009, 384 features) and 201, 4367 features) + NS	10 (IS-09) / 10 (IS-11) features × 389 signals with 2 s duration	GBM	Fivefold CV	CRDs vs. endplate potentials vs. fasciculation potentials vs. fibrillation potentials/ PSW vs. myotonic discharges vs. noise artifact	90.4% (INTERSPEECH 2009 feature set)	67%
Hazarika et al. (2019)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10)	Yes	TFA (DWT-Daubechies 2) on features decomposed with F-SVD + canocical	4 features × *NS*	kNN	Train: 50%, test: 50%	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS,	(T1) 96.5%, (T2) 98.1%, (T3) 97.6%	55%

Table 2 (*continued*)

Author(s) (year)	Data source	Pre-pro	Features	Input	Model	Training partitions	Classification task	Performance (accuracy)	CLAIM
	from BB, APB, TA, VL C.7: ALS (n = 4), myopathy (n = 4), controls (n = 4) from BB, APB, TA, VL		correlation analysis, LDA and feature fusion				(T3) Healthy vs. ALS vs. myopathy		
Kamali and Stashuk (2018)	C.4: patient count unknown from D, VM, FDI, TA	Yes	TD and TFA (DWT-Daubechies) + Laplacian score and NDEC clustering	4 clusters (4 TD, 6 TFA features) × 6576 MUPTs from 322 segments with 15 s duration	MIL with SVM / RF	Training: leave one-out CV	Healthy vs. neuropathy vs. myopathy in (T1) TA muscle, (T2) FDI, (T3) DLT, (T4) VM	(T1) 100% (RF), (T2) 98.8% (RF, SVM), (T3) 97.9% (RF, SVM), (T4) 97.6% (SVM)	59%
Subasi et al. (2018)	C.5: neuropathy (n = 13), myopathy (n = 7), controls (n = 7) from BB	Yes	TFA (DWT-Daubechies 4)	47 features × 2400 segments with 0.25 s duration	ANN / kNN / SVM / CART, C4.5 DT / REPTree / LAD Tree / RF combined with bagging	Tenfold CV	Healthy vs. neuropathy vs. myopathy	99% (Bagging ensemble with SVM)	54%
Nagineni et al. (2018)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	No	TFA (VMD)	6 features × 329 segments with 11.2 s duration s	ELM	Tenfold CV	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) Healthy vs. disease	(T1) 99.5%, (T2) 98.5%, (T3) 97.8%	52%
Vallejo et al. (2018)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, VM, D	No	TFA (DWT-Daubechies 8) + relevance analysis with Fuzzy entropy	10 features × 120 segments with 11.2 s duration	Artificial NN	Tenfold CV	Healthy vs. myopathy vs. ALS	98.3%	75%
Kehri and R. N (2018)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB O.2: neuropathy (n = 1), myopathy (n = 1), controls (n = 1) from TA	Yes	TFA (WT) + based on thresholds	*NS*	SVM and artificial NN	Train: 70%, test: 30%	Healthy vs. disease	95% (SVM)	43%
Hazarika et al. (2018)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, APB, TA, VL C.7: ALS (n = 4), myopathy (n = 4), controls (n = 4) from BB, APB, TA, VL		TFA (DWT-Daubechies 2) on multiview features + canocical correlation analysis, LDA and feature fusion	4 features × *NS*	LDA-kNN	Twofold CV	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) Healthy vs. ALS vs. myopathy	(T1) 96.7%, (T2) 97.6% (T3) 98.8%	69%
Dostal et al. (2018)	C.8: neuropathy (n = 40), controls (n = 40) from TA	NS	TD	1 to 4 features × 80 segments with 4 s duration	SVM	Leave-one-out CV	Healthy vs. neuropathy	90% (four features)	48%
Sengur et al. (2017)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, VM, D	Yes	N/A	30,000 spectrograms / CWT / SPWVD retrieved from segments with *NS* duration	CNN	Train: 95.7%, test: 4.3%	Healthy vs. ALS	96.8% (CWT)	59%
Kamali and Stashuk (2017)	C.4: patient count unknown from TA	Yes	TD + clustering with NDEC / Chameleon / SC / DBSCAN / K-means	6–10 clusters (>5 TD features) extracted from 103 segments × 1845 MUPTs with 15 s duration	SVM	Training: leave-one-out CV	Healthy vs. neuropathy vs. myopathy	97% (NDEC)	63%
Mishra et al. (2017)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TFA (improved EMD, IMF 1–8)	4 features × *NS*	LS-SVM	NS	Healthy vs. ALS	96.3% (IMF2)	41%

(*continued on next page*)

Table 2 (*continued*)

Author(s) (year)	Data source	Pre-pro	Features	Input	Model	Training partitions	Classification task	Performance (accuracy)	CLAIM
Artameeyanant et al. (2016)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) O.2: neuropathy (n = 1), myopathy (n = 1), controls (n = 1)	Yes	Adjusted-weight visibility algorithm and statistical feature extraction + ANOVA	6 features × *NS*	kNN / multiple layer perceptron NN / SVM	Fivefold CV	Healthy vs. ALS vs. myopathy for (T1) dataset a and (T2) dataset b	(T1) 99.2%, (T2) 98.4%	58%
Mishra et al. (2016)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TFA (EMD, IMF 1–8)	6 features × 220 segments with 11.2 s duration	LS-SVM	Tenfold CV	Healthy vs. ALS	95% (IMF2)	48%
Bozkurt et al. (2016)	C.5: neuropathy (n = 13), myopathy (n = 7), controls (n = 7) from BB	No	TD (autoregressive models) / FD (multiple single classification / eigenvectors)	n features × 1200 segments with *NS* duration	Feed forward NN / combined NN using feed forward NNs	Tenfold CV	Healthy vs. neuropathy vs. myopathy	94% (combined NN with eigenvectors features)	52%
Subasi (2015)	C.5: neuropathy (n = 13), myopathy (n = 7), controls (n = 7) from BB	No	TFA (DWT-Daubechies 4)	17 features × *NS*	Evolutionary SVM trained with GA	Tenfold CV	Healthy vs. neuropathy vs. neuropathy	97%	53%
Naik et al. (2015)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TD + uncorrelated LDA	5 features × *NS* segments with 175 ms duration	LDA and majority voting (for overlapping segments)	Train: 50%, validation: 25%, test: 25%	Healthy vs. ALS vs. myopathy	98%	59%
Bajaj and Kumar (2015)	O.2: neuropathy (n = 1), myopathy (n = 1), controls (n = 1) from TA	Yes	TFA (EMD, IMF1-4)	2 features × 310 segments with 0.25 s duration	LS-SVM	Tenfold CV	Healthy vs. neuropathy vs. myopathy	99% (IMF2)	52%
Krishna and Thomas (2015)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB	Yes	TD / FD (direct spectral features) / TFA (DWT)	4 TD / 4 FD / 4 TFA features × *NS* dominant MUAPs extracted from 750 segments with 5 s duration	kNN	NS	(T1) Healthy vs. myopathy, (T2) Healthy vs. ALS, (T3) Healthy vs. ALS vs. myopathy	(T1) 83.5% (TD, TFA) (T2) 96.5% (TD, FD, and TFA) (T3) 84% (TFA)	52%
Gokgoz and Subasi (2015)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, VM	Yes	TFA (DWT-Daubechies 4)	27 features × 900 segments with 625 ms duration	CART / C4.5 DT / RF	k-fold CV k NS	Healthy vs. ALS vs. myopathy	96.7% (RF)	43%
Doulah et al. (2014)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from NS	Yes	TFA (DWT-Daubechies 2)	n features × dominant MUAPs extracted from 750 segments with 5 s duration	kNN	Leave-one-out CV	Healthy vs. ALS: (T1) small set, (T2) large set Healthy vs. ALS vs. myopathy: (T3) small set, (T4) large set	(T1) 100%, (T2) 99.5%, (T3) 100%, (T4) 98.8%	62%
Gokgoz and Subasi (2014)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from BB, VM	Yes	FD (multiple single classification) + NS	n features × 900 segments with 625 ms duration	kNN / artificial NN / SVM	Tenfold CV	Healthy vs. ALS vs. myopathy	92.6% (SVM)	50%
Kamali et al. (2014)	O.1: ALS (n = 8), myopathy (n = 7), controls (n = 10) from NS	Yes	TD / TFA (DWT with Symlet 4) + genetic algorithm	1–7 features × *NS* MUAPs extracted from *NS* segments	SVM as single classifier / multiple classifiers multiple features best voting / multiple classifiers multiple features all voting	Tenfold CV	Ensembling of binary classification (one-against one and one-against-all) for healthy, neuropathy and myopathy	97% (MCMFAV)	57%

Table 2 (*continued*)

Author(s) (year)	Data source	Pre-pro	Features	Input	Model	Training partitions	Classification task	Performance (accuracy)	CLAIM
AbdelMaseeh et al. (2014)	C.4: patient count unknown from FDI, D, TA, VM	Yes	TD + sequential floating forward search	*n* features × 1845 MUPTs (TA), 1540 MUPTs (FDI), 1272 MUPTs (VM), 1107 MUPTs (DLT) retrieved from segments with *NS* duration	Event association rules / Gaussian mixture models employing different binarization mappings	Leave-one-out CV	Healthy vs. neuropathy vs. myopathy in (T1) TA muscle, (T2) FDI, (T3) DLT, (T4) VM	(T1) 91.3% (GMM), (T2) 84% (EAR), (T3) 83.2% (GMM), (T4) 83.6% (GMM)	70%
Kamali et al. (2013)	O.1: ALS (*n* = 8), myopathy (*n* = 7), controls (*n* = 10) from unknown muscle	Yes	TD + sequential forward floating selection algorithm / TFA (DWT-Daubechies 4) + mutual information criterion and joint mutual information	6 TD / 3 TFA features × *NS* MUPTs retrieved from segments with *NS* duration	SVM in hybrid structure	Tenfold CV	Ensemble of binary classification (one-against all) for healthy, neuropathy and myopathy	91%	55%
Subasi (2013)	C.5: neuropathy (*n* = 13), myopathy (*n* = 7), controls (*n* = 7) from BB	No	TFA (DWT-Daubechies 4)	23 features × *NS*	kNN / RBFN / SVM / PSO-SVM	Tenfold CV	Healthy vs. neuropathy vs. myopathy	97.4% (PSO-SVM)	67%
Dobrowolski et al. (2012)	C.9: neuropathy (*n* = 70), myopathy (*n* = 30), controls (*n* = 200) from D, FDI, VL, TA	Yes	TFA (DWT with Symlet / Daubechies / Coiflet)	5 features × >6000 dominant MUAPs retrieved from segments with *NS* duration	SVM (linear)	Train: 30%, test: 60%	(T1) Myopathy. vs others (healthy, neuropathy), (T2) Neuropathy vs. others (healthy, myopathy)	NS	63%
Subasi (2012b)	C.5: neuropathy (*n* = 13), myopathy (*n* = 7), controls (*n* = 7) from BB	Yes	TFA (DWT-Daubechies 4)	15 features × *NS* segments with *NS* duration	LDA / ANN / C4.5 DT / SVM / fuzzy SVM	Tenfold CV	Healthy vs. neuropathy vs. myopathy	93.5% (fuzzy SVM)	65%
Subasi (2012a)	C.5: neuropathy (*n* = 13), myopathy (*n* = 7), controls (*n* = 7) from BB	Yes	TD (autoregressive) / TFA (DWT / WPE)	*n* features × 1200 segments with *NS* duration	ANN / dynamic fuzzy NN / adaptive neurofuzzy inference system	Tenfold CV	Healthy vs. neuropathy vs. myopathy	95% (ANFIS classifier and autoregressive + DWT features)	59%
Fattah et al. (2012)	O.1: ALS (*n* = 6), controls (*n* = 6) from BB	Yes	FD (FFT: spectral peak / mean frequency) / TD (autocorrelation / zero crossing rate)	1–4 features × *NS*	kNN	Leave-one-out CV	Healthy vs. ALS	100% (with spectral peak and autocorrelation features)	54%
Tomczykiewicz et al. (2012)	C.9: patient count unknown from D, FDI, VL, TA	Yes	TD / TFA (WT-Symlet4)	*n* features × 19,000 MUAPs retrieved from segments with *NS* duration	SVM	NS	Healthy vs. neuropathy vs. myopathy	NS	52%

With a "/" it is indicated that methods are compared and with a "," it is indicated that methods are combined.
General: ALS = amyotrophic lateral sclerosis; CV = cross-validation; CRD = complex repetitive discharge; IBM = inclusion body myositis; N/A = not applicable; NS = not specified; MUAP = motor unit action potential; MUPT = motor unit potential train.
Muscles: APB = abductor pollicis brevis; BB = biceps brachii; D = deltoid; FDI = first dorsal interosseous; TFL = tensor fasciae latae; TA = tibialis anterior; TB = triceps brachii; VL = vastus lateralis; VM = vastus medialis.
Feature extraction in time domain (TD), with time–frequency analysis (TFA), in frequency domain (TF): AR = autoregressive; CWT = continuous wavelet transform; DWT = discrete wavelet transform; EMD = empirical mode decomposition; F-SVD = frame singular value decomposition; HHT = Hilbert-Huang transform, IMF = intrinsic mode function; LPB = local binary pattern; PSO = particle swarm optimization; SET = synchro-extracting transform; ST = Stockwell transform; STFT = short-time Fourier transform; SPWVD = smoothed pseudo-Wigner–Ville distribution; VMD = variational mode decomposition; WPE = wavelet packet energy; WT = wavelet transform; WVT = Wigner-Ville transform.
Feature selection and reduction: DBSCAN = density-based clustering algorithm, LDA = linear discriminant analysis; NDEC = neighbourhood distances entropy consistency; SC = spectral clustering.
Classifiers: CART = classification and regression tree; CNN = convolutional neural network; DA = discriminant analysis; DT = decision tree; ELM = extreme learning machine; ESD = ensemble subspace discriminant; FT = fine tree; GA = genetic algorithm; GBM = gradient boosting machine; kNN = k-nearest neighbour; LDA = linear discriminant analysis; LS-SVM = least squares support vector machine; MIL = multiple instance learning; MLPNN = multi-layer perceptron neural network; NB = naïve-Bayes; NN = neural network; RBFN = radial basis function networks; ResNet = Residual neural network; RF = random forest; SVM = support vector machines; VGG = visual geometry group.

visualizations of nEMG data (Nodera et al., 2019b, Sengur et al., 2017) (Fig. 2d).

3.2.5. Model

Machine learning classifiers were commonly implemented for the classification of nEMG signals. The most frequently used algorithms included support vector machines (SVM) in 27 out of 51 articles (53%), a type of neural network (NN) in 17 articles (33%), k-nearest neighbours (kNN) in 16 articles (31%), decision tree (DT) in 8 articles (16%) and LDA in 8 articles (16%) (Fig. 2e). SVM aims to find a hyperplane that maximizes the margin between two classes to separate data points into different classes. NN is a machine learning model that is trained on manually selected features, and the goal is to minimize the difference between its predicted output and the true labels. kNN classifies a new data point based on the similarity between that data point and the data points in the training dataset. DTs recursively partition data based on feature values to separate data into homogeneous subsets by selecting the best feature and threshold. LDA can be used as both a feature reduction method and a classifier by finding a linear combination of features that maximally separates different classes in the data.

In addition, some studies used an aggregation technique to improve classification accuracy by combining the results of multiple classifiers. Random forests (RF) are ensemble machine learning classifiers to leverage the strengths of DTs which are constructed using random subsets of training data and features. RF was used in 8 articles (16%). Other aggregation methods included majority vote (Boyer-Moore) (Jose et al., 2020b, Jose et al., 2022), and bagging and boosting (Subasi et al., 2018, Yaman and Subasi, 2019). The Boyer-Moore algorithm searches for the most frequently occurring class labels across multiple classifiers. Bagging involves constructing multiple models using bootstrapped samples of the training data and averaging their outputs to reduce variance, while boosting sequentially builds models to reduce bias and increase accuracy by adjusting the weights of the training data to give more emphasis on difficult-to-class instances.

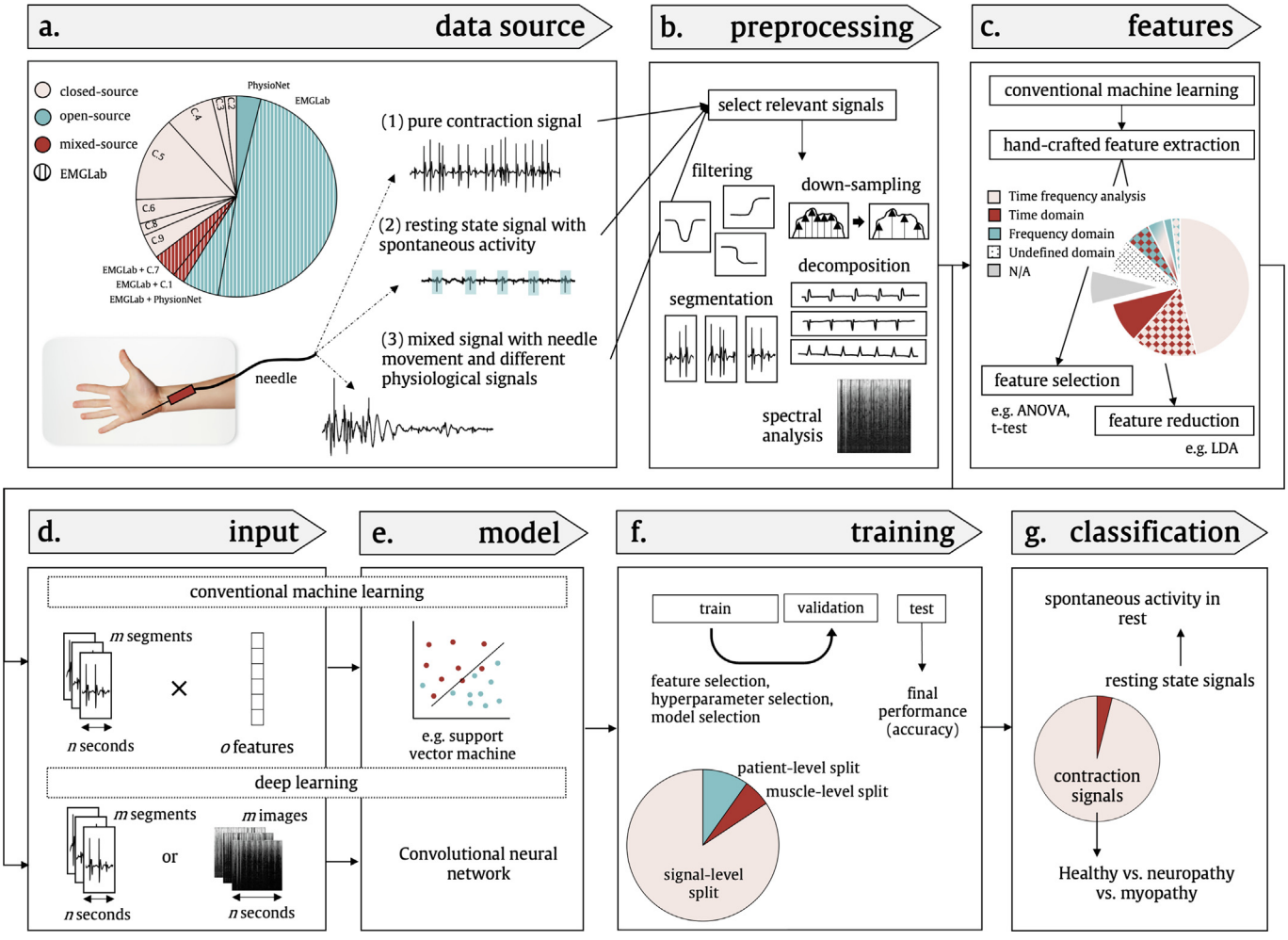

Fig. 2. Approach for the development of artificial intelligence models in the classification of needle electromyography signals of the included studies. a. The data source determines which signals are used for training. A distinction is made for open-source or closed-source (C.number) datasets. The signals in the datasets are signals recorded during contraction (1), signals recorded at rest (2) and signals with both needle movements (artefacts) and signals at rest or during contraction (3). **b.** During preprocessing the relevant signals are selected and different methods could be applied, such as filtering, down-sampling, decomposition, segmentation and/or spectral analysis. **c.** optionally, features are created. Deep learning methods do not use feature extraction/selection/reduction, as the features are automatically learned from the input signals. **d.** The input of the model is calculated over segments with specific duration. **e.** The input is fed into the model. **f.** Training is performed on a distinct training set, and the validation performance is used to make decisions on feature selection, hyperparameter selection and/or model selection. The training set should ideally be completely distinct from the validation and test set, i.e., patient-level splits. For studies where it was either not described or could not be deduced, it was assumed that a signal-level split was implemented. **g.** Models are trained for the classification of healthy, neuropathy or myopathy using contraction state signals and for the classification of spontaneous activity with resting state signals. LDA = linear discriminant analysis.

Table 3
Summary of data sources used in included studies.

Data source		Signals			Signal acquisition			No. of studies (% of total)
		Muscle activity	Subjects	Population	Muscle(s)	Band-pass filter	Sampling	
Open-source								
EMGlab	O.1	Constant and low contraction	25	ALS, myopathy, healthy controls	APB, BB, D, TFL, TA, TB, VL, VM	2 Hz – 10 kHz	23 kHz for 11.2 s with 16 bits resolution	31 (60.7%)
PhysioNet	O.2	Constant and low contraction	3	Neuropathy, myopathy, healthy controls	TA	20 Hz – 5 kHz	50 kHz for several seconds	5 (9.8%)
Closed-source								
Peking University Third Hospital of China, Peking, China	C.1	Contraction	322	ALS, healthy controls	FDI, TA	NS	24 kHz and 6 kHz	1 (2.0%)
Leiden University Medical Centre, Leiden, the Netherlands	C.2	Rest, contraction	65	ALS, IBM, healthy controls	D, TA, BB, VM + 32 other muscles	30 Hz – 3 kHz	4.8 or 5 kHz for max 40 s	1 (2.0%)
Seoul National University Hospital, Seoul, South-Korea	C.3	Contraction	57	Neuropathy, myopathy, healthy controls	Distal and proximal muscles	NS	48 kHz for 0.41–4 s	1 (2.0%)
Mayo Clinic, Arizona, USA	C.4	Constant contraction (40–60 MUAPs/s)	NS	Neuropathy, myopathy, healthy controls	D, VM, FDI, TA	10 Hz – 10 kHz	48 kHz	4 (7.8%)
University of Gaziantep, Turkey	C.5	Constant contraction at 30% of maximum voluntary contraction	27	Neuropathy, myopathy, healthy controls	BB	5 Hz – 10 kHz	20 kHz for 5 s with 12 bits resolution	7 (13.7%)
Tokushima University Hospital, Japan	C.6	At rest	83–103	Routine clinical indications	BB, FDI, VM, TA	20 Hz – 10 kHz	NS	2 (3.9%)
Guwahati Neurological Research Centre, Guwahati, Assam, India	C.7	Constant contraction	12	ALS, myopathy, healthy controls	BB, APB, TA, VL	NS	NS	2 (3.9%)
Czech Republic	C.8	Constant contraction at 30% of maximum voluntary contraction	80	Neuropathy, healthy controls	TA	5 Hz – 10 kHz	12,5 kHz for 5 s	1 (2.0%)
Military Institute of Health Service, Warsaw, Poland	C.9	Low contraction	NS	Neuropathy, myopathy, healthy controls	D, FDI, VL, TA	20 Hz – 10 kHz	20 kHz	2 (3.9%)

General: ALS = amyotrophic lateral sclerosis; IBM = inclusion body myositis; NS = not specified. **Muscles**: APB = abductor pollicis brevis; BB = biceps brachii; D = deltoid; FDI = first dorsal interosseous; TFL = tensor fasciae latae; TA = tibialis anterior; TB = triceps brachii; VL = vastus lateralis; VM = vastus medialis.

Convolutional neural networks (CNNs) were implemented in the four studies that applied deep learning with variations in the architecture and input data (Nodera et al., 2019b, Sengur et al., 2017, Yoo et al., 2022, Zhang et al., 2023). CNNs are commonly used for image classification tasks, because they can effectively capture spatial patterns. Two studies, however, used CNNs to extract features from the nEMG signals in the time domain (Yoo et al., 2022, Zhang et al., 2023). One study implemented a wavelet-based CNN, which uses wavelet analysis to extract features (Zhang et al., 2023). The other study implemented a CNN with residual connection (Yoo et al., 2022). Residual connection enables information to bypass certain layers, addressing the challenges of training deeper networks and preserving important input information (He et al., 2016). Two other studies transformed the nEMG signals in the time domain into images using time–frequency analysis methods, enabling the use of image-based models. One study implemented popular image-based networks, such as ResNet with 50 layers (Nodera et al., 2019b). The other study implemented a simple CNN network with two convolutional layers (Sengur et al., 2017). Since deep learning models often require a large amount of data to perform well, two studies implemented data augmentation techniques to artificially increase the number of data samples (Nodera et al., 2019b, Sengur et al., 2017).

3.2.6. Training

During model training, a model is trained on a subset of the data and evaluated on another subset of the data. This evaluation (or validation) guides in how to finetune the model's parameters, select appropriate features, and select the final model (Fig. 2f). In some cases, an additional separate set is held out, this is the test set. In 35 studies (69%) cross-validation (CV) was employed as their training method. A common approach of CV is *k*-fold CV, where the data is divided into multiple (*k*) folds: one fold is held out as the validation set, while the remaining folds are used for training. This process is repeated *k* times, ensuring that each fold serves as the validation set once. The performance obtained from each fold is then averaged to provide an overall performance estimate. In some studies *k*-fold CV was performed multiple times to ensure a more robust outcome estimate. Another approach to CV is leave-one-out CV, where training is performed using all samples except for one, which is used as the validation sample. In 13 studies (25%), a fixed split of the data was used for training, validation, and/or testing purposes. In one study (2%) external validation was used as evaluation method (Zhang et al., 2023).

The level of split of how the data samples are divided over each subset of data (train set, validation set or test set) varies and this can be at signal-level, muscle-level, or patient-level. Many of the

studies performed a signal-level split (Fig. 2f) where signals sampled from the same muscle could end up in multiple sets.

3.2.7. Classification task

We found that the most prevalent classification task was to classify healthy, neuropathy, and myopathy signals in contraction nEMG data (Fig. 2g). Twenty-five studies (49%) used multi-class classification. Eleven studies (22%) used binary classification, such as healthy vs. neuropathy, healthy vs. myopathy, neuropathy vs. myopathy, or healthy vs. disease. Eight studies (16%) addressed the multi-class problem by breaking it down into multiple binary classifications. Five studies (10%) combined multiple binary classifications and multi-class classifications.

Four studies applied the same classification task on different subsets of data. For example, on multiple muscles within the same dataset (AbdelMaseeh et al., 2014, Jose et al., 2022, Kamali and Stashuk, 2020, Kamali and Stashuk, 2018) or on multiple datasets (Artameeyanant et al., 2016). In other studies, the same classification output was compared across different subsets of data, such as signals from the same muscle or patient (Tannemaat et al., 2023), combination of features from all muscles, or split for proximal or distal muscles (Yoo et al., 2022), and individual segments or total signal (Torres-Castillo et al., 2022). Furthermore, in five studies, the classification output of individual segments were combined to reach a conclusion at recording-level, (Torres-Castillo et al., 2022), at muscle-level (Kamali and Stashuk, 2020, Kamali and Stashuk, 2018) and at patient-level (Tannemaat et al., 2023, Yoo et al., 2022). In Tannemaat et al. (2023) and in Yoo et al. (2022) the classification probabilities at segment level were used to reach a conclusion at patient level.

Albeit most studies focused on contraction nEMG, two out of 51 studies classified spontaneous activity at rest: endplate potentials, fibrillation potentials or positive sharp waves (PSW), complex repetitive discharges (CRDs), fasciculation potentials, myotonic discharges, and artefacts (Nodera et al., 2019a, 2019b).

3.2.8. Performance results

Many studies reported high performance across binary and multi-class classification tasks (Table 2). Four studies reported 100% accuracy for a binary classification task: healthy vs. myopathy data samples or healthy vs. neuropathy data samples (Bakiya et al., 2021a; Doulah et al., 2014; Fattah et al., 2012; Roy et al., 2022) and two studies reported a 100% accuracy for a multi-class classification task (Kamali and Stashuk, 2020, Kamali and Stashuk, 2018). More recent studies reported lower performance accuracies compared to earlier published studies (Fig. 3).

3.3. CLAIM checklist

The CLAIM checklist results are presented as a percentage in the final column of Table 2, with results per item available in Supplementary Material 2. Some items were not scored for the studies because they did not suit the studies in this field. The CLAIM items provide insights in potential biases that may exist in the development of the models. A higher score indicates that a study reported on more aspects that are important in model development.

The results of the CLAIM in relation to the reported accuracies of all classification tasks of individual studies is shown in Fig. 3. The dots in the figure depict the CLAIM value, a larger dot means a higher value and hence a higher reporting quality. The figure shows that more recent studies reported lower performance accuracies compared to earlier published studies; however, their CLAIM score were overall higher.

4. Discussion

4.1. Key findings

This scoping review provided an overview of AI classification of clinical nEMG signals at rest and during contraction. Most studies classified nEMG signals recorded during contraction (49 out of 51) and only two studies classified signals at rest. Despite excellent classification accuracy of healthy, neuropathic or myopathic signals, the proposed methods have methodological limitations which make them not yet suitable for clinical implementation. Studies that scored high on the CLAIM checklist, suggesting higher reliability of their results, generally report a lower accuracy.

4.2. Findings of included studies

Many studies included in this review selected data in highly specific patient groups, such as the EMGLab dataset (used in 31 out of 51 studies; 61%) or the PhysioNet dataset (used in 5 out of 51 studies; 10%). Although classification in such groups provide a good proof-of-concept, the resulting method is not generalizable to a broad population of patients nor to a population with varying stages of disease progression. Additionally, many studies used binary classification, which simplifies the classification process but does not reflect the complexity of clinical multiclass problems. A few studies collected data during routine clinical examinations (AbdelMaseeh et al., 2014, Kamali and Stashuk, 2020, Kamali and Stashuk, 2017, 2018, Nodera et al., 2019a, 2019b, Tannemaat et al., 2023, Yoo et al., 2022), which is the preferred approach to translate results to daily clinical practice.

Many studies have adopted a training process that may suffer from data leakage, leading to overfitting and overestimation of performance estimates. Data leakage frequently occurred during feature selection. This is for example seen when the whole dataset is used to select the most discriminative features and applied to both the training and validation sets during training (Dubey et al., 2022; Roy et al., 2022). An unbiased and accurate estimation on the test set can only be achieved when all aspects of model training, including feature selection, model selection, and hyperparameter tuning, are performed exclusively on the training set (Kaufman et al., 2012, Varma and Simon, 2006).

Many of the included studies applied a signal-level split (Fig. 2f). It is preferred, however, that data is disjointed at patient-level. This will ensure that training and test data are not intertwined. A recent study found that using an incorrect split at slice-level instead of subject-level on 3D magnetic resonance imaging (MRI) data resulted in 40–55% higher accuracy on smaller datasets and 25–45% higher accuracy on larger datasets when evaluating the performance of different 2D CNNs (Yagis et al., 2021).

External validation can prevent biases in performance estimates due to data leakage or other limitations of the training dataset (Altman et al., 2009, Collins et al., 2014). Unfortunately, only one article performed external validation, with a resulting performance of 78.9% accuracy in the classification of healthy vs. neuropathy (ALS) (Zhang et al., 2023). Not using external validation, likely results in overfitted models with extreme classification performances of 95–100% accuracy (Bakiya et al., 2021a, 2021b, Bose et al., 2021, Chatterjee et al., 2020, Doulah et al., 2014, Fattah et al., 2012, Hazarika et al., 2019, Hazarika et al., 2018, Krishna and Thomas, 2015, Mishra et al., 2017, Mishra et al., 2016, Mokdad et al., 2020, Nagineni et al., 2018, Samanta et al., 2022, Sengur et al., 2017), highlighting the importance of adopting external validation.

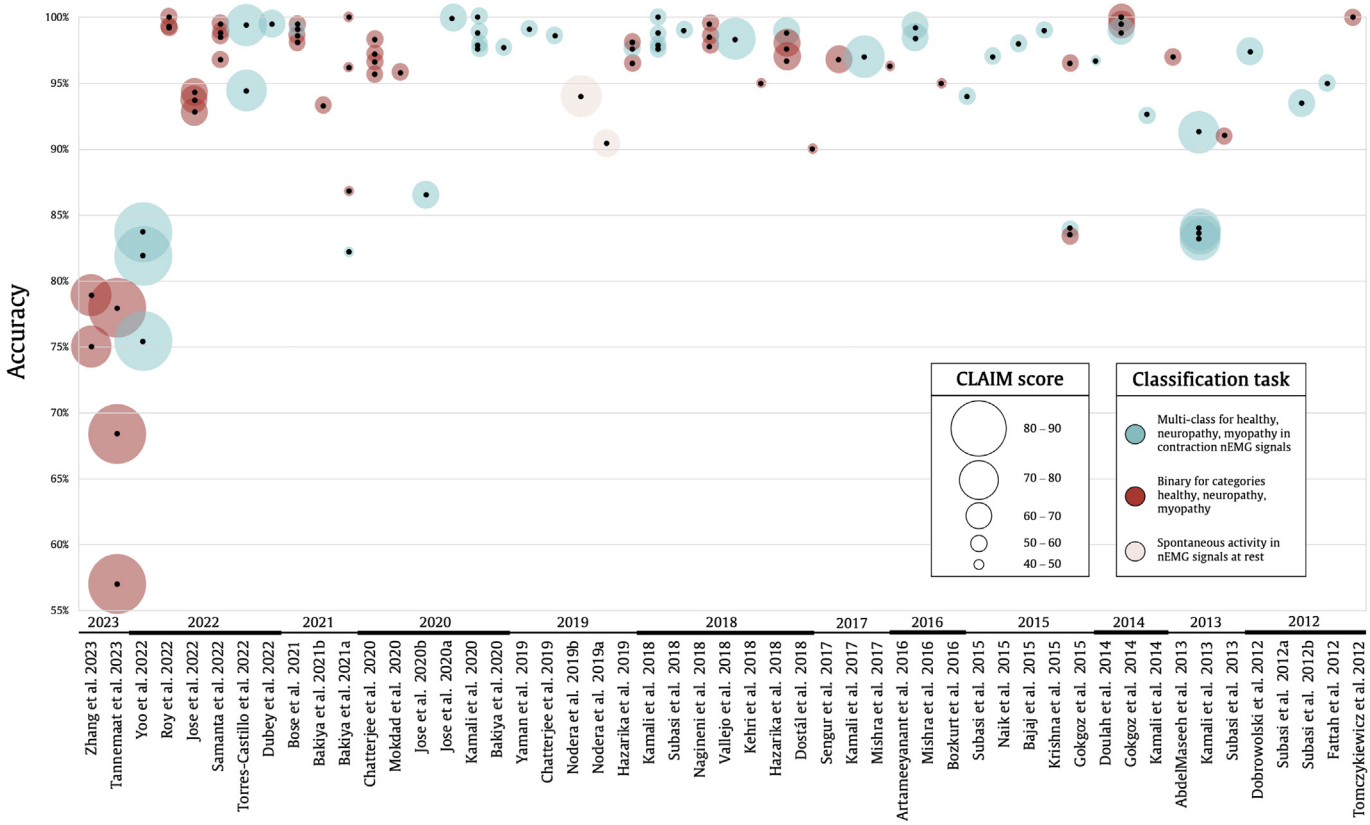

Fig. 3. Visualisation of classification accuracy vs. Checklists for Artificial Intelligence in Medical Imaging (CLAIM) score. All included studies are shown on the x-axis, ordered chronologically. Black dots indicate the accuracy (if available) of the classification tasks. The colour represents the type of classification task researched: blue for multi-class classification task to classify healthy, neuropathy and myopathy in contraction nEMG signals; red for a binary classification of two of those three classes in contraction nEMG signals; light-pink for multi-class classification task to identify spontaneous activity in nEMG signals at rest. For studies with multiple classification tasks, multiple dots were added. The circle radius corresponds to the CLAIM score, where larger is better.

Clinical nEMG signal acquisition encompasses recording muscle activity during rest, different levels of contraction, and of artefacts, such as caused by needle movement. However, all studies either acquired nEMG signals during a pure muscle state or manually segmented parts of the signal to exclude unwanted artefacts. To close the gap between clinical practice and the assessed studies, new research may focus more on direct assessment of raw clinical data as acquired in clinical practice without preselection. This improves applicability, usability and ultimately the feasibility of AI in clinical practice.

Already a few studies have incorporated majority voting algorithms on the complete recordings (albeit only from contraction signals) by combining classification results of individual segments: at recording-level, (Torres-Castillo et al., 2022), at muscle-level (Kamali and Stashuk, 2020, Kamali and Stashuk, 2018) and at patient-level (Tannemaat et al., 2023, Yoo et al., 2022). This approach avoids bias by preselection of signals and more closely aligns with clinical practice, where the clinician may combine the signals from multiple needle positions in the muscles and from different muscles. In Yoo et al. (2022), part of model development was based on findings in proximal and distal muscles to make a distinction between myopathies, polyneuropathies and normal. Depending on the clinical question, appropriate muscles should be examined, e.g., inclusion of proximal muscles for suspected myopathy, muscles innervated by different radices at and above and below the level of suspected radiculopathy, muscles in different segments for suspected motor neuron disease (Menkes and Pierce, 2019). Moreover, it is suggested to combine the classification results of rest and contraction signals as well, to reach a more complete classification at muscle and patient level that aligns with clinical practice (Blijham et al., 2006, Pugdahl et al., 2021, Shefner et al., 2020).

4.3. Implications for future research

Future research (Box 2) should focus on models suitable for daily clinical practice. Current research has mainly focused on high-quality signals, but it is important to investigate methods for effectively handling hospital-acquired nEMG signals that may contain low-quality signals and artefacts. Models should be able to handle variability in signal quality or provide real-time feedback on the signal quality during examination.

Box 2 List of priorities for future research.

- Focus on relevant clinical classification problems
- Develop reliable and reproducible reference standards for nEMG classifications
- Develop models using raw, unselected data as collected during routine clinical examination
- Make more data available for training
- Contribute to open-source datasets

Ideally, models are capable in identifying diseases in early stages with high sensitivity and specificity. This will benefit current clinical practice, where for some diseases multiple examinations are performed to reach a definitive diagnosis. For model development it is, however, essential that more (open-source) data

becomes available with reliable reference standards. Approaches such as in Tannemaat et al. (2023), incorporating nEMG recordings of the first examination where the final diagnosis is not yet definitive, should be further investigated. Furthermore, we strongly urge all nEMG manufacturers to facilitate access to raw nEMG data, as automated raw data access is still a major hurdle when creating open-source datasets. We also urge all clinicians to make their nEMG data and gold standard diagnosis available in a standardized and anonymized way, which has already shown to accelerate innovation in the field of electroencephalography (EEG) (Gorgolewski et al., 2016, Harati et al., 2014). During model development some other factors may also be considered to facilitate subsequent clinical implementation, such as usability, explainability, and expected costs. Finally, it is important to note that implementation of (semi) automated methods for interpretation EMG signals will not solve any sampling errors. However, other techniques during interpretation, such as counting the recording sites per muscle, or the use of muscle ultrasound for needle guidance, may reduce this chance.

To improve the methodological quality of future research, it is recommended that authors follow checklists such as the CLAIM. These checklists encourage comprehensive descriptions of study approaches, leading to increased transparency and reliability of results. Some specific methodological choices or learning points for authors to consider: (1) train with as much data as possible, (2) do not focus on binary classification tasks for which no clinical problem exists, (3), always use independent test data, preferably with external validation, (4) train correctly where all training steps are exclusively performed on the training set to prevent data-leakage, (5) consider patient level data splitting.

4.4. Strengths and limitations of this scoping review

In this scoping review we provided an extensive overview of the current literature on automatic classification of clinical nEMG signals using machine and deep learning. We compared (the reporting of) essential steps during different stages of model development and their final performance and identified several shortcomings, which may help to improve future studies.

The analysis of the included studies and the checklist scoring were performed by the same reviewer (SJ), who consulted reviewer WVP and CV in case of doubt. Although the same reviewer guarantees consistency in scoring, the scores may be more reliable when performed by the consensus of two or more independent reviewers.

4.5. Conclusion

In this scoping review an elaborate overview is provided of various machine and/or deep learning techniques employed for the automatic classification of clinical nEMG signals. High performance accuracies were reported which are, however, subject to bias, and published methods are not yet ready for clinical implementation. Based on the published studies, it is not clear which approach for model development will perform best in the end. We suggest that future research focuses on training with as much data as possible, using data acquired from a diverse patient population, and report performances on independent test data. With these suggestions we aim to contribute to the development of models that can effectively handle signal quality variability and that are suitable for daily clinical practice.

Conflict of interest statement

S. de Jonge and C. Verhamme declare no conflicts of interest relevant to the manuscript. W.V Potters is co-founder, employee, and shareholder of TrianecT.

Acknowledgments

The neuromuscular team of Amsterdam UMC, location AMC, is member of the Netherlands Neuromuscular Centre (NL-NMD) and the European Reference Network for rare neuromuscular diseases (ERN-EURO-NMD).

Funding

This research received no external funding.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT from OpenAI in order to improve language and readability of parts of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinph.2023.12.134.

References

AbdelMaseeh M, Chen TW, Poupart P, Smith B, Stashuk D. Transparent muscle characterization using quantitative electromyography: different binarization mappings. IEEE Trans Neural Syst Rehabil Eng 2014;22(3):511–21.
Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338(may28 1):b605-b.
Artameeyanant P, Sultornsanee S, Chamnongthai K. An EMG-based feature extraction method using a normalized weight vertical visibility algorithm for myopathy and neuropathy detection. Springerplus 2016;5(1):2101.
Bajaj V, Kumar A. Features based on intrinsic mode functions for classification of EMG signals Features based on intrinsic mode functions for classification of EMG signals 157. Int J Biomed Eng Technol 2015;18:156–67.
Bakiya A, Kamalanand K, Rajinikanth V. Assessment of electromyograms using genetic algorithm and artificial neural networks. Evol Intell 2021a;14(2):261–71.
Bakiya A, Kamalanand K, Rajinikanth V. Automated diagnosis of amyotrophic lateral sclerosis using electromyograms and firefly algorithm based neural networks with fractional position update. Phys Eng Sci Med 2021b;44(4):1095–105.
Bakiya A, Kamalanand K, Rajinikanth V, Nayak RS, Kadry S. Deep neural network assisted diagnosis of time-frequency transformed electromyograms. Multimed Tools Appl 2020;79(15–16):11051–67.
Blijham PJ, Hengstman GJ, Hama-Amin AD, van Engelen BG, Zwarts MJ. Needle electromyographic findings in 98 patients with myositis. Eur Neurol 2006;55(4):183–8.
Bose R, Samanta K, Modak S, Chatterjee S. Augmenting Neuromuscular Disease Detection Using Optimally Parameterized Weighted Visibility Graph. IEEE J Biomed Health Inform 2021;25(3):685–92.
Bozkurt MR, Subasi A, Koklukaya E, Yilmaz M. Comparison of AR parametric methods with subspace-based methods for EMG signal classification using stand-alone and merged neural network models. Turk J Electr Eng Comput Sci 2016;24(3):1547–59.
Chatterjee S, Roy SS, Bose R, Pratiher S. Feature extraction from multifractal spectrum of electromyograms for diagnosis of neuromuscular disorders. IET Sci Meas Technol 2020;14(7):817–24.
Chatterjee S, Samanta K, Choudhury NR, Bose R. Detection of Myopathy and ALS Electromyograms Employing Modified Window Stockwell Transform. IEEE Sens Lett 2019;3(7):1–4.
Christodoulou CI, Pattichis CS. Unsupervided pattern recognition for the classification of EMG signals. IEEE Trans Biomed Eng 1999;46(2):169–78.
Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014;14(1):40.
Dobrowolski AP, Wierzbowski M, Tomczykiewicz K. Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. Comput Methods Programs Biomed 2012;107(3):393–403.
Dostal O, Vysata O, Pazdera L, Prochazka A, Kopal J, Kuchynka J, et al. Permutation Entropy and Signal Energy Increase the Accuracy of Neuropathic Change Detection in Needle EMG. Comput Intell Neurosci 2018;2018:5276161.
Doulah AB, Fattah SA, Zhu WP, Ahmad MO. Wavelet domain feature extraction scheme based on dominant motor unit action potential of EMG signal for

neuromuscular disease classification. IEEE Trans Biomed Circuits Syst 2014;8 (2):155–64.
Dubey R, Kumar M, Upadhyay A, Pachori RB. Automated diagnosis of muscle diseases from EMG signals using empirical mode decomposition based method. Biomed Signal Process Control 2022;71.
Farkas C, Hamilton-Wright A, Parsaei H, Stashuk DW. A review of clinical quantitative electromyography. Crit Rev Biomed Eng 2010;38(5):467–85.
Fattah SA, Iqbal A, Jumana MA. Identifying the Motor Neuron Disease in EMG Signal Using Time and Frequency Domain Features with Comparison. Signal Img Process Int J 2012;3(2):99–114.
Gokgoz E, Subasi A. Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders. J Med Syst 2014;38(4):31.
Gokgoz E, Subasi A. Comparison of decision tree algorithms for EMG signal classification using DWT. Biomed Signal Process Control 2015;18:138–44.
Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 2016;3(1) 160044.
Harati A, Lopez S, Obeid I, Picone J, Jacobson M, Tobochnik S. The TUH EEG CORPUS: A big data resource for automated EEG interpretation. In: 2014 IEEE signal processing in medicine and biology symposium (SPMB): IEEE. p. 1–5.
Hazarika A, Barthakur M, Dutta L, Bhuyan M. F-SVD based algorithm for variability and stability measurement of bio-signals, feature extraction and fusion for pattern recognition. Biomed Signal Process Control 2019;47:26–40.
Hazarika A, Dutta L, Boro M, Barthakur M, Bhuyan M. An automatic feature extraction and fusion model: application to electromyogram (EMG) signal classification. Int J Multimed Inf Retr 2018;7(3):173–86.
He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-8. doi: 10.1109/CVPR.2016.90.
Johnson N, Phillips M. Rayyan for systematic reviews. J Electron Resour Librariansh 2018;30(1):46–8.
Jose S, George ST, Roopchand PS. DWT-based electromyogram signal classification using maximum likelihood-estimated features for neurodiagnostic applications. Signal Image Video Process 2020a;14(3):601–8.
Jose S, George ST, Subathra MSP, Handiru VS, Jeevanandam PK, Amato U, et al. Robust Classification of Intramuscular EMG Signals to Aid the Diagnosis of Neuromuscular Disorders. IEEE Open J Eng Med Biol 2020b;1:235–42.
Jose S, Selvaraj TG, Samuel K, Philip JT, Nanjappan Jothiraj S, Muthu Swamy Pandian S, et al. Intramuscular EMG classifier for detecting myopathy and neuropathy. Int J Imaging Syst Technol 2022.
Kamali T, Boostani R, Parsaei H. A hybrid classifier for characterizing motor unit action potentials in diagnosing neuromuscular disorders. J Biomed Phys Eng 2013;3(4):145–54.
Kamali T, Boostani R, Parsaei H. A Multi-Classifier Approach to MUAP Classification for Diagnosis of Neuromuscular Disorders. IEEE Trans Neural Syst Rehabil Eng 2014;22(1):191–200.
Kamali T, Stashuk D. Transparent Electrophysiological Muscle Classification From EMG Signals Using Fuzzy-Based Multiple Instance Learning. IEEE Trans Neural Syst Rehabil Eng 2020;28(4):842–9.
Kamali T, Stashuk DW. A Density-Based Clustering Approach to Motor Unit Potential Characterizations to Support Diagnosis of Neuromuscular Disorders. IEEE Trans Neural Syst Rehabil Eng 2017;25(7):956–66.
Kamali T, Stashuk DW. Electrophysiological Muscle Classification Using Multiple Instance Learning and Unsupervised Time and Spectral Domain Analysis. IEEE Trans Biomed Eng 2018;65(11):2494–502.
Kaufman S, Rosset S, Perlich C. Leakage in data mining: formulation, detection, and avoidance. ACM Trans Knowl Discov Data 2012;6(4):1–21.
Kehri V, R. N A. EMG Signal Analysis for Diagnosis of Muscular Dystrophy Using Wavelet Transform, SVM and ANN. Biomed Pharmacol J 2018;11(3):1583-91.
Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve 2006;34(2):238–41.
Kim J-E, Seok JM, Ahn S-W, Yoon B-N, Lim Y-M, Kim K-K, et al. Basic concepts of needle electromyography. Ann Clin Neurophysiol 2019;21(1):12–7.
Krishna A, Thomas P. Classification of EMG Signals Using Spectral Features Extracted from Dominant Motor Unit Action Potential. Int J Eng Adv Technol 2015:2249–8958.
Menkes DL, Pierce R. Needle EMG muscle identification: A systematic approach to needle EMG examination. Clin Neurophysiol Pract 2019;4:199–211.
Mishra VK, Bajaj V, Kumar A, Sharma D, Singh GK. An efficient method for analysis of EMG signals using improved empirical mode decomposition. Aeu-Int J Electron C 2017;72:200–9.
Mishra VK, Bajaj V, Kumar A, Singh GK. Analysis of ALS and normal EMG signals based on empirical mode decomposition. Iet Sci Meas Technol 2016;10 (8):963–71.
Mokdad A, Debbal SME, Meziani F. Diagnosis of amyotrophic lateral sclerosis (ALS) disorders based on electromyogram (EMG) signal analysis and feature selection. Polish J Medical Phys Eng 2020;26(3):155–60.
Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell 2020;2(2) e200029.
Nagineni S, Taran S, Bajaj V. Features based on variational mode decomposition for identification of neuromuscular disorder using EMG signals. Health Inf Sci Syst 2018;6(1):13.
Naik GR, Selvan SE, Nguyen HT. Single-Channel EMG Classification With Ensemble-Empirical-Mode-Decomposition-Based ICA for Diagnosing Neuromuscular Disorders. IEEE Trans Neural Syst Rehabil Eng 2015;24(7):734–43.
Narayanaswami P, Geisbush T, Jones L, Weiss M, Mozaffar T, Gronseth G, et al. Critically re-evaluating a common technique: Accuracy, reliability, and confirmation bias of EMG. Neurology 2016;86(3):218–23.
Nikolic MSM, Krarup C, Sørensen JA. Detailed Analysis of Clinical Electromyography Signals. Faculty of Health Science. Copenhagen: University of Copenhagen; 2001.
Nodera H, Osaki Y, Yamazaki H, Mori A, Izumi Y, Kaji R. Classification of needle-EMG resting potentials by machine learning. Muscle Nerve 2019a;59(2):224–8.
Nodera H, Osaki Y, Yamazaki H, Mori A, Izumi Y, Kaji R. Deep learning for waveform identification of resting needle electromyography signals. Clin Neurophysiol 2019b;130(5):617–23.
Pattichis CS, Elia AG. Autoregressive and cepstral analyses of motor unit action potentials. Med Eng Phys 1999;21(6–7):405–19.
Pattichis CS, Pattichis MS. Time-scale analysis of motor unit action potentials. IEEE Trans Biomed Eng 1999;46(11):1320–9.
Preston DC, Shapiro BE. Electromyography and Neuromuscular Disorders. Philadelphia: Elsevier; 2021. p. 134–259.
Pugdahl K, Camdessanché J-P, Cengiz B, de Carvalho M, Liguori R, Rossatto C, et al. Gold Coast diagnostic criteria increase sensitivity in amyotrophic lateral sclerosis. Clin Neurophysiol 2021;132(12):3183–9.
Roy SS, Dey D, Karmakar A, Roy AS, Ashutosh K, Choudhury NR. Detection of abnormal electromyograms employing DWT-based amplitude envelope analysis using Teager energy operator. Int J Biomed Eng Technol 2022;40(3):224–40.
Rubin DI. Needle electromyography: Basic concepts. Handb Clin Neurol. 2019;160:243–56.
Samanta K, Chatterjee S, Bose R. Neuromuscular disease detection based on feature extraction from time-frequency images of EMG signals employing robust hyperbolic Stockwell transform. Int J Imaging Syst Technol 2022;32 (4):1251–62.
Sengur A, Akbulut Y, Guo Y, Bajaj V. Classification of amyotrophic lateral sclerosis disease based on convolutional neural network and reinforcement sample learning algorithm. Health Inf Sci Syst 2017;5(1):9.
Shefner JM, Al-Chalabi A, Baker MR, Cui L-Y, de Carvalho M, Eisen A, et al. A proposal for new diagnostic criteria for ALS. Clin Neurophysiol 2020;131(8).
Subasi A. Classification of EMG signals using combined features and soft computing techniques. Appl Soft Comput 2012a;12(8):2188–98.
Subasi A. Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. Comput Biol Med 2012b;42 (8):806–15.
Subasi A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Comput Biol Med 2013;43(5):576–86.
Subasi A. A decision support system for diagnosis of neuromuscular disorders using DWT and evolutionary support vector machines. Signal Image Video Process 2015;9(2):399–408.
Subasi A, Yaman E, Somaily Y, Alynabawi HA, Alobaidi F, Altheibani S. Automated EMG Signal Classification for Diagnosis of Neuromuscular Disorders Using DWT and Bagging. Procedia Comput Sci 2018;140:230–7.
Tannemaat M, Kefalas M, Geraedts V, Remijn-Nelissen L, Verschuuren A, Koch M, et al. Distinguishing normal, neuropathic and myopathic EMG with an automated machine learning approach. Clin Neurophysiol 2023;146:49–54.
Theodoridis S. Machine Learning A Bayesian and Optimization Perspective. Academic press; 2015. p. 1–16.
Tomczykiewicz K, Dobrowolski AP, Wierzbowski M. Evaluation of motor unit potential wavelet analysis in the electrodiagnosis of neuromuscular disorders. Muscle Nerve 2012;46(1):63–9.
Torres-Castillo JR, López-López CO, Padilla-Castañeda MA. Neuromuscular disorders detection through time-frequency analysis and classification of multi-muscular EMG signals using Hilbert-Huang transform. Biomed Signal Process Control 2022;71 103037.
Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018;169(7):467–73.
Vallejo M, Gallego CJ, Duque-Munoz L, Delgado-Trejos E. Neuromuscular disease detection by neural networks and fuzzy entropy on time-frequency analysis of electromyography signals. Expert Systems 2018;35(4).
Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinform 2006;7(1):91.
Yagis E, Atnafu SW, Seco García, de Herrera A, Marzi C, Scheda R, Giannelli M, et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. Scientific Reports 2021;11(1) 22544.
Yaman E, Subasi A. Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. Biomed Res Int 2019;2019 9152506.
Yoo J, Yoo I, Youn I, Kim S-M, Yu R, Kim K, et al. Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability. Comput Methods Programs Biomed 2022;226 107079.
Zhang H, Liu Y, Qing Z, He J, Teng S, Wang X, et al. Domain Contrast Network for cross-muscle ALS disease identification with EMG signal. Biomed Signal Process Control 2023;82 104582.