Neurological Sciences (2020) 41:3719–3727
https://doi.org/10.1007/s10072-020-04499-y

ORIGINAL ARTICLE

Electrodiagnostic accuracy in polyneuropathies: supervised learning algorithms as a tool for practitioners

Antonino Uncini[1] · Graziano Aretusi[1,2] · Fiore Manganelli[3] · Yukari Sekiguchi[4] · Laurent Magy[5] · Stefano Tozza[3] · Atsuko Tsuneyama[4] · Sophie Lefour[5] · Satoshi Kuwabara[4] · Lucio Santoro[3] · Luigi Ippoliti[2]

Received: 16 March 2020 / Accepted: 30 May 2020 / Published online: 10 June 2020

Abstract

Objective The interpretation of electrophysiological findings may lead to misdiagnosis in polyneuropathies. We investigated the electrodiagnostic accuracy of three supervised learning algorithms (SLAs): shrinkage discriminant analysis, multinomial logistic regression, and support vector machine (SVM), and three expert and three trainee neurophysiologists.

Methods We enrolled 434 subjects with the following diagnoses: chronic inflammatory demyelinating polyneuropathy (99), Charcot-Marie-Tooth disease type 1A (124), hereditary neuropathy with liability to pressure palsy (46), diabetic polyneuropathy (67), and controls (98). In each diagnostic class, 90% of subjects were used as training set for SLAs to establish the best performing SLA by tenfold cross validation procedure and 10% of subjects were employed as test set. Performance indicators were accuracy, precision, sensitivity, and specificity.

Results SVM showed the highest overall diagnostic accuracy both in training and test sets (90.5 and 93.2%) and ranked first in a multidimensional comparison analysis. Overall accuracy of neurophysiologists ranged from 54.5 to 81.8%.

Conclusions This proof of principle study shows that SVM provides a high electrodiagnostic accuracy in polyneuropathies. We suggest that the use of SLAs in electrodiagnosis should be exploited to possibly provide a diagnostic support system especially helpful for the less experienced practitioners.

Keywords Polyneuropathies · Electrodiagnosis · Diagnostic accuracy · Supervised learning algorithms

Introduction

Polyneuropathy is a common neurological disorder. The prevalence is 2.4% in the overall population, rising to 8% in subjects older than 55 [1]. There are over 200 identified causes of neuropathy, and structured approaches, including electrodiagnostic studies, have been proposed to assist the physician in the differential diagnosis [2–4]. Electrodiagnosis is considered a sensitive, specific, and validated tool to reveal a polyneuropathy also contributing to define the distribution and symmetry of the disorder; to determine whether the neuropathy is motor or sensory, primarily demyelinating or axonal; to evaluate the severity; and eventually to establish prognosis [3–5]. However, the interpretation of electrodiagnostic findings may lead to misdiagnosis. In chronic inflammatory demyelinating polyneuropathy (CIDP), an early diagnosis and treatment are critical to prevent potentially irreversible axonal injury and disability, but a misdiagnosis based on the interpretation of nerve conduction studies, which may imply costly and sometimes with

Antonino Uncini and Graziano Aretusi contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10072-020-04499-y) contains supplementary material, which is available to authorized users.

✉ Antonino Uncini
uncini@unich.it

[1] Department of Neuroscience, Imaging and Clinical Sciences, University "G. d'Annunzio", Via Luigi Polacchi 11, 66100 Chieti-Pescara, Italy

[2] Statistics Unit, Department of Economics, University "G. d'Annunzio", Chieti-Pescara, Italy

[3] Department of Neurosciences, Reproductive Sciences and Odontostomatology, University of Naples Federico II, Naples, Italy

[4] Department of Neurology, Graduate School of Medicine, Chiba University, Chiba, Japan

[5] National Reference Centre for Rare Peripheral Neuropathies and Department of Neurology, University of Limoges, Limoges, France

dangerous side effects therapies, has been reported in up to 45% of CIDP patients [6, 7]. Electrodiagnostic criteria help to reduce CIDP misdiagnosis but necessitate the normalization of electrophysiological data and are infrequently employed in the everyday practice [8]. Moreover, the diagnostic accuracy of criteria sets was assessed versus patients with diabetic neuropathy, amyotrophic lateral sclerosis, not better specified distal sensory and sensory-motor axonal neuropathies, and only rarely versus other demyelinating neuropathies [9–12]. At last, there is an overlap of electrophysiological findings among patients with different neuropathies and, given the segmental nature of the demyelinating process, some results in CIDP can be within normal ranges. All the above considerations illustrate the inherent difficulties and limitations of using a set of electrophysiologic parameters to establish reliable electrodiagnostic criteria in polyneuropathies.

Classification in medicine faces several difficulties: (i) the physician needs a long time to accumulate enough knowledge and experience to distinguish between different and similar patterns, (ii) the classification process is often labor intensive and time consuming, and (iii) the most difficult challenge occurs when differential characteristics are not prominent and hence not easily discernible. The use of expert systems can represent an important advance in medical practice, and supervised learning algorithms (SLAs), although rarely employed in electrodiagnosis, have proven to be effective in diagnosis of diabetic polyneuropathies and recently in distinguishing between demyelinating and axonal Guillain-Barré syndrome subtypes [13–15]. In this study, we aimed to investigate the electrodiagnostic accuracy of three SLAs and six neurophysiologists in cohorts of patients with demyelinating and axonal polyneuropathies. To the best of our knowledge, this is the first time that this comprehensive approach has been attempted.

Materials and methods

A description of the background theory for SLAs and the software used for their implementation are reported in the supplementary material.

Study design and participants

This is a retrospective study in which the electrophysiological data and the reference clinical diagnoses were collected before the application of the SLAs.

Three academic, tertiary, neuromuscular centers (University of Naples, Italy; University of Chiba, Japan; University of Limoges, France) participated in the study. An independent group, working at the University of Chieti-Pescara (Italy), organized the study and performed the statistical analysis. Participants were randomly chosen in each center from the list of patients evaluated for eligibility in the years 2012–2018 and satisfying the following inclusion criteria. CIDP was diagnosed on the basis of clinical features, electrodiagnostic studies, CSF profile, supportive information, longitudinal course, and treatment response according to the European Federation of Neurological Societies (EFNS) and the Peripheral Nerve Society (PNS) guideline [16]. Atypical CIDP presentations consistent with multifocal acquired demyelinating sensory and motor neuropathy or distal acquired demyelinating symmetrical neuropathy were included[17, 18]; patients with antimyelin-associated glycoprotein antibodies were excluded. Patients with Charcot-Marie-Tooth type 1A (CMT1A) had symmetrical motor and sensory neuropathy with distal weakness and foot deformity and harbored a duplication of PMP-22 gene or were first-degree relatives with clinical signs of patients with positive molecular diagnosis. Patients with hereditary neuropathy with liability to pressure palsy (HNPP) had a history of positional paresthesias or episodic nerve palsy and carried a PMP-22 gene deletion. Patients with diabetic sensorimotor polyneuropathy (DPN) had type I or II diabetes mellitus, chronic distal symmetric symptoms or signs, and at least one conduction abnormality in each of two separate nerves (the peroneal and the sural) [4, 19]. Other causes of distal symmetric polyneuropathies were excluded by routine screening laboratory tests including B12 and methylmalonic acid dosage and serum protein immunofixation electrophoresis [4].

Controls were subjects referred for an electroneuromiographic examination that after extensive laboratory tests and follow-up did not show any evidence of neuromuscular disorders. The data of 522 anonymised subjects were sent to the Chieti-Pescara centre. The diagnosis of each subject was considered the reference standard, and the diagnoses of all subjects were available only for the Chieti-Pescara group.

Electrophysiology

Distal motor latency (DML), amplitude and duration of negative peak of compound muscle action potential from different stimulation sites, motor conduction velocity (MCV), and minimal F-wave latency of at least 16 trials were measured in median, ulnar, peroneal, and tibial motor nerves [15]. Sensory studies were performed antidromically in median, ulnar, radial, and sural nerves; amplitude of sensory action potential was measured baseline to negative peak [15]. Electrophysiological findings were normalized according to the control values of each center. Given the retrospective nature of the study and the different polyneuropathies considered, a variable number of motor (3 to 8) and sensory nerves (2 to 4) were tested in the participants. Although this variability may be suboptimal for research purposes, it reflects the everyday practice.

Data pre-processing

To handle missing data, we used simple approaches as removing participants, nerves, and electrophysiological parameters with an excessive amount of missing data. The reconstruction of missing values for specific electrophysiological parameters was made by using a per-class median imputation [15].

Supervised learning algorithms

No algorithm works optimally for every problem. Therefore, we investigated the performance of three SLAs: shrinkage discriminant analysis (SDA), multinomial logistic regression (MLR), and support vector machine (SVM) to predict the diagnostic class membership of a new subject while using a hold-out test set of data [20]. This approach is particularly appropriated when experiments are relatively specific, and the issue is a classification task with correlation among predictors as in the current study [21].

The evaluation of the discriminant power of the different SLAs includes the following steps (see supplementary material): (1) the model examines the data, (2) the model learns from its mistakes, and (3) a conclusion on how the model performs is reached. The first two steps require splitting the data into training and validation sets. The former represents the dataset that the model will use to be trained on by adjusting the parameters. On the other hand, the validation set represents the data used after training to compute the accuracy or error of the SLA. In our case study, we iteratively split the data by using a tenfold cross-validation procedure [21], where we allow 90% of the dataset to be the actual train set and the remaining 10% to be the validation set where to calculate the prediction error rate (see supplementary material). The final step regards the use of a test set which provides the "gold standard" used to evaluate the model. The test set is a portion of data that has been partitioned at the very start of the experiment. In our case, it contains 10% of data spanning the various diagnostic classes. The test set functions similarly to the validation set, except that we never used these data in the process of building or tuning the models. Finally, the test set can be used to evaluate competing models.

Each SLA provides, for each subject, a posterior probability to belong to the five diagnostic classes, and the highest probability of each subject was used as the decision rule for the final classification.

Role of neurophysiologists

After having randomly ordered the subjects using a uniform random generator, the final dataset containing only the age and the electrophysiological data was sent back to the centers and examined by three expert (more than 10 years of practice) and three trainee (1–2 years of practice) neurophysiologists. Experts and trainees were asked to make independently the diagnosis according to their routine practice. Five over six neurophysiologists applied the EFNS/PNS electrodiagnostic criteria for CIDP [16], and three neurophysiologists considered an abnormal median normal sural SAP pattern helpful in CIDP diagnosis. All neurophysiologists considered diffuse slowing (< 38 m/s) of MCV as characteristic of CMT1A and five also looked for homogeneity of slowing (conduction velocity difference between nerves less than 10 m/s) for differential diagnosis with CIDP. Regarding HNPP, all neurophysiologists looked for NCV slowing limited or more evident at the entrapment sites with particular attention by three neurophysiologists to conduction slowing of the ulnar nerve across the elbow. DPN was diagnosed by at least one conduction abnormality in the peroneal and sural nerve examination in patients with NCV slowing that for five out of six neurophysiologists should not fit the EFNS/PNS criteria for CIDP. One expert neurophysiologist applied his own, stepwise, electrodiagnostic procedure consisting in first focusing on median motor and sensory conduction values, then checking the results of sural conduction and finally looking for conduction slowing and conduction blocks in other motor nerves. In the presence of electrophysiological findings indicative of carpal tunnel syndrome, he directly looked for an ulnar slowing at the elbow.

Statistical methods

The performance of SLAs and of neurophysiologists on the training and test set was summarized by means of confusion matrices and by the following four performance indicators (informative statistics): overall accuracy (proportion of correctly classified subjects for all the five diagnostic classes), precision (proportion of per-class true positive among all subjects classified as positive), sensitivity (proportion of true positives that are correctly identified as positive), and specificity (proportion of true negatives that are correctly identified as negative) [22].

Prediction results provided by SLAs and neurophysiologists were then ranked by a multidimensional comparison analysis computing the Euclidean distance of the performance indicators results for SLAs and neurophysiologists with the benchmark consisting in the ideal situation of observing a perfect (100%) correct classification in each one of the diagnostic classes.

The balance between sensitivity and specificity was studied using the Receiver Operating Characteristic plot. Finally, the McNemar with Edwards's correction test was used in CIDP to assess if SLAs and neurophysiologists made random or systematic misclassification [23].

Results

Figure 1 shows the study design. The initial raw data matrix contained 522 subjects and 96 electrophysiological parameters with 63.9% of missing values. Eighty-eight participants were excluded because of excessive amount of missing electrophysiological data. The tibial nerves and some electrophysiological parameters, as minimal F-wave latency, recorded in an insufficient number of subjects were also excluded from the analysis. Finally, by satisfying the double criteria of restricting the maximum number of missing data per subject at 30% and setting the minimum number of information available for each electrophysiological parameter within the diagnostic classes at 25% (for larger classes) and 35% (for smaller classes), it was possible to obtain a final dataset of 434 subjects with 16.6% of missing values that were reconstructed by imputation. A missing data rate of 15–20% is considered common in this type of studies [24]. The final dataset contained 434 subjects: 98 controls, 99 CIDP (14 multifocal acquired demyelinating sensory and motor neuropathy) patients, 124 CMT1A, 46 HNPP, and 67 DPN (17 with type I diabetes). The demographic characteristics of participants are reported in Table 1 of supplementary material.

Each participant was characterized by 27 electrophysiological parameters belonging to three motor (median, ulnar and peroneal) and three sensory (median, ulnar and sural) nerves (Table 2 supplementary material).

The size of the final dataset was assumed to be sufficient to ensure an acceptable amount of precision of the SLAs. Moreover, classifiers such as the SVM, which can be efficiently designed in spaces of very high dimensionality, and SDA, which is particularly suited for high-dimensional

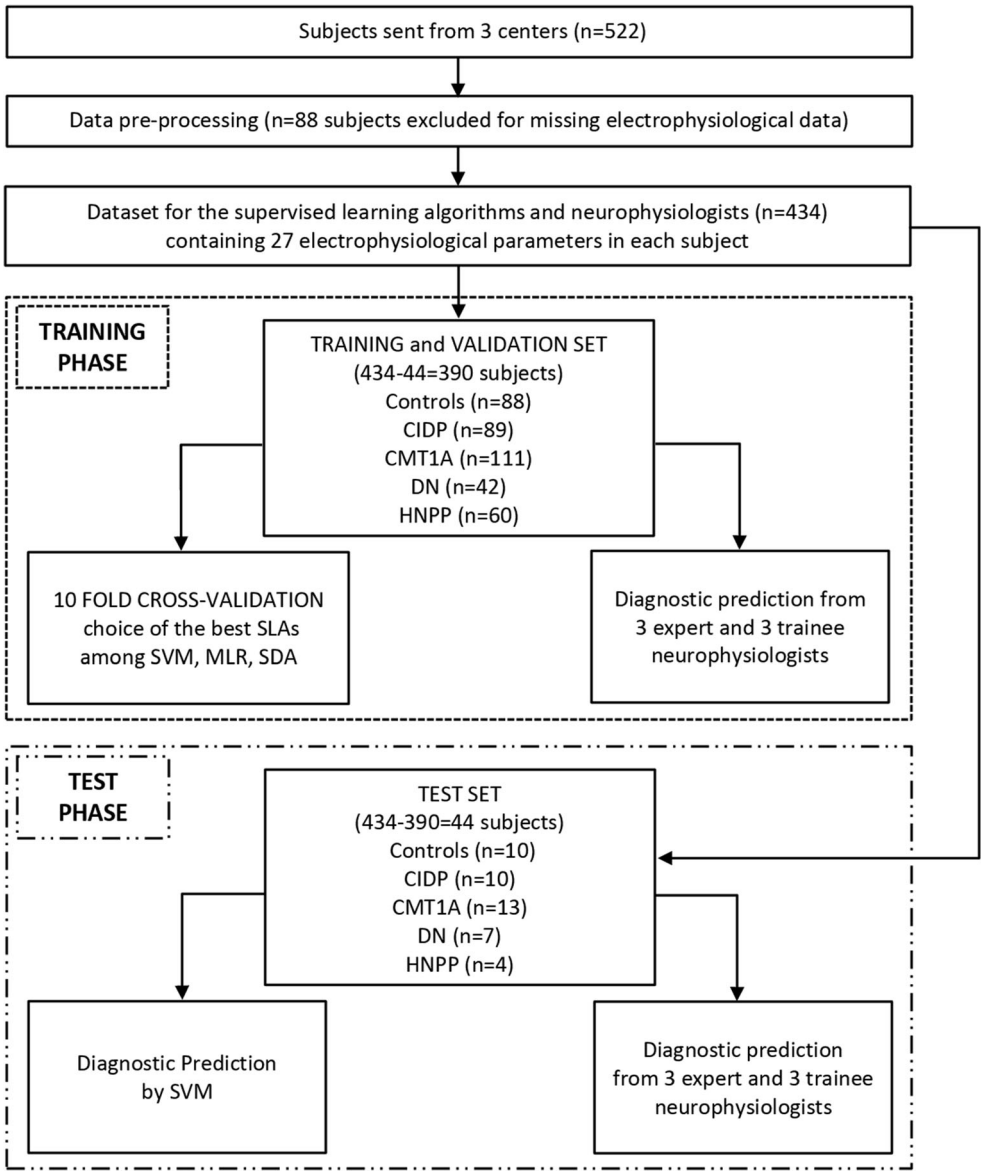

Fig. 1 Study profile. CIDP chronic inflammatory demyelinating polyneuropathy, CMT1A Charcot-Marie-Tooth disease type 1A, DPN diabetic polyneuropathy, HNPP hereditary neuropathy with liability to pressure pals, MLR multinomial logistic regression, SDA shrinkage discriminant analysis, SLA supervised learning algorithm, SVM support vector machine

classification with correlation among predictors, allow to efficiently address the problem of dimensionality [25].

Regarding the CIDP group, 98% of the 99 patients were classifiable by the clinical and electrophysiological information as belonging to the definite CIDP and 2% as possible CIDP diagnostic categories according to Table 6 of the EFSN/PNS guideline [16].

Applying, to the final dataset, containing for each subject only the 27 parameters listed in Table 2 of supplementary material, the electrodiagnostic criteria reported in Table 1 of EFNS/PNS guideline [16], 59% of patients with the CIDP reference diagnosis were classified as definite, 4% as probable, and 17% as possible, and 20% were unclassifiable. Interestingly, applying the same criteria to the other polyneuropathies, 100% of CMT1A patients were classifiable as definite CIDP because of prolongation of DML≥ 50% above the upper limit of normal and reduction of MCV≤ 30% below the lower limit of normal in at least two nerves. Six percent of HNPP were classifiable as definite, 6.5% as probable, and 17.4% as possible CIDP. Among DPN patients, 0.2% was classifiable as probable and 12% as possible CIDP.

Table 1 shows the classification results from the tenfold cross-validation procedure for SLAs and neurophysiologists on the same data (i.e., training set of 390 subjects). The overall diagnostic accuracy of SVM (90.5%) is higher than of other SLAs (83.6 and 86.7%), expert (75.1–81.5%) and trainee (54.9–76.7%) neurophysiologists. SVM demonstrates also the best precision values (83.3–98.9%) in four out of five diagnostic classes indicating that this SLA is preferable in terms of bias and variability of classifications. Moreover, considering all diagnostic classes, SVM shows the best balance in terms of sensitivity and specificity. Table 1 provides also the ranking obtained by a multidimensional comparison, and each column is color formatted to reflect the best (green) and worst (red) performance for each indicator of the SLAs and neurophysiologists in each diagnostic class. The fact that SVM achieves the best performances is evident from the highest representation of green boxes in the SVM row. SDA and MLR ranked respectively second and third whereas the best of neurophysiologist ranked fourth.

The receiver operating characteristic plots shown in Fig. 2 confirm that SLAs are highly performing in terms of sensitivity and specificity, and this is true especially for SVM. In contrast, neurophysiologists are less performing.

For the CIDP diagnosis, SVM shows a sensitivity of 79.8% and a specificity of 96% yielding an optimal balance compared with other SLAs (59.6–74.2%; 94–98.7%), expert neurophysiologists (50.6–79.8%; 81.1–96.3%), and trainees (36–51.7%; 68.4–98.3%) (Table 1). The diagnostic accuracy of SLAs and neurophysiologists in CIDP is shown in Table 2.

Table 3 shows the results of the McNemar test analyzing the difference between false negatives and false positives in CIDP.

There are no significant differences for SVM (0.36), MLR (0.53), and trainee 3 (0.31) indicating that two out of three SLAs do not make systematic error in predicting CIDP class whereas this happened for five out of six neurophysiologists. It is also remarkable that the lowest number of false negative and false positive is associated with SVM confirming its good performance.

The superiority of SVM with respect to the other SLA was also confirmed from the test set on 44 subjects whose data were not prior used (data not shown). To save space, Table 4 only shows only the classification results of SVM and neurophysiologists on the same test set of data.

Discussion

This study should be seen as a first approach to apply some classification algorithms to the electrodiagnosis of polyneuropathies and should be considered as proof of principle study. SVM showed the highest electrodiagnostic accuracy in all diagnostic classes and ranked first considering all the performance indicators both in the training and test set. The very good performance of SVM is probably due to the fact that it is a high flexible approach especially when the

Table 1 Results on training set of performance indicators of supervised learning algorithms (based on tenfold cross-validation procedure) and of neurophysiologists for each diagnostic class. Each column is color formatted to reflect the best (green) and worst (red) performance for each indicator of the supervised learning algorithms and neurophysiologists in each diagnostic class

	Ranking	Euclidean Distance	Overall Accuracy	Precision					Sensitivity					Specificity				
				CONTROL	CIDP	CMT1A	HNPP	DN	CONTROL	CIDP	CMT1A	HNPP	DN	CONTROL	CIDP	CMT1A	HNPP	DN
SVM	1	0.407	90.5%	98.9%	85.5%	92.1%	89.5%	83.3%	100.0%	79.8%	94.6%	81.0%	91.7%	99.7%	96.0%	96.8%	98.9%	96.7%
SDA	2	0.603	86.7%	95.5%	93.0%	90.2%	78.3%	71.4%	95.5%	59.6%	99.1%	85.7%	91.7%	98.7%	98.7%	95.7%	97.1%	93.3%
MLR	3	0.680	83.6%	96.4%	78.6%	87.3%	66.0%	79.3%	90.9%	74.2%	92.8%	73.8%	76.7%	99.0%	94.0%	94.6%	95.4%	96.4%
EXP 3	4	0.788	81.5%	96.5%	76.3%	92.5%	58.5%	69.6%	93.2%	65.2%	89.2%	73.8%	80.0%	99.0%	94.0%	97.1%	93.7%	93.6%
EXP 1	5	0.815	81.8%	96.6%	80.4%	94.4%	70.5%	60.0%	96.6%	50.6%	91.0%	73.8%	95.0%	99.0%	96.3%	97.8%	96.3%	88.5%
TR 1	6	1.035	76.7%	93.6%	86.5%	87.2%	75.0%	48.2%	100.0%	36.0%	91.9%	57.1%	88.3%	98.0%	98.3%	94.6%	97.7%	82.7%
EXP 2	7	1.374	75.1%	96.5%	55.5%	93.1%	30.8%	66.1%	94.3%	79.8%	84.7%	9.5%	68.3%	99.0%	81.1%	97.5%	97.4%	93.6%
TR 3	8	1.433	66.7%	91.6%	46.0%	90.4%	31.6%	58.3%	98.9%	51.7%	59.5%	28.6%	81.7%	97.4%	82.1%	97.5%	92.5%	89.4%
TR 2	9	1.679	54.9%	73.1%	31.2%	90.6%	31.0%	71.2%	98.9%	48.3%	26.1%	31.0%	70.0%	89.4%	68.4%	98.9%	91.7%	94.8%

SDA shrinkage discriminant analysis, *MLR* multinomial logistic regression, *SVM* support vector machine, *EXP* expert neurophysiologist, *TR* trainee neurophysiologist, *CIDP* chronic inflammatory polyradiculoneuropathy, *CMT1A* Charcot-Marie-Tooth disease type 1A, *HNPP* hereditary neuropathy with liability to pressure palsy, *DPN* diabetic polyneuropathy

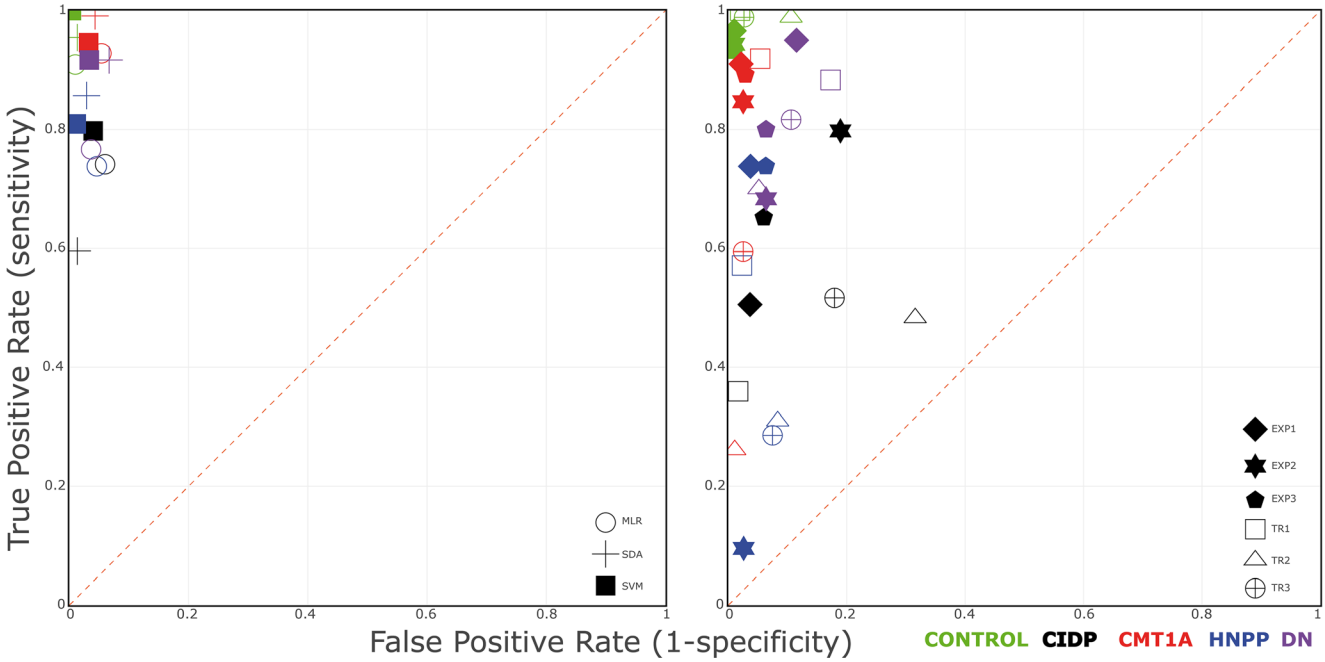

Fig. 2 Receiver operating characteristic plots of supervised learning algorithms (left) and neurophysiologists (right) for the five diagnostic classes. Note that supervised learning algorithms are highly performing in terms of sensitivity and specificity as the data plots are concentrated in the upper left quadrant; on the other hand, neurophysiologists are less performing as the data plot is more scattered. CIDP chronic inflammatory demyelinating polyneuropathy, CMT1A Charcot-Marie-Tooth disease type 1A, DPN diabetic polyneuropathy, HNPP hereditary neuropathy with liability to pressure pals, MLR multinomial logistic regression, SDA shrinkage discriminant analysis, SVM support vector machine

classes overlap. Moreover, when the prediction accuracy is the main goal, as in our case, SVM demonstrated to be one of the best SLAs [26]. The error rate in CIDP diagnosis based on the interpretation of nerve conduction studies has been reported to be 45% [7], and it was, in the current study, 36.2% for the neurophysiologist with the worst performance. Therefore, the possibility to reduce the error rate to 7.7% by employing SVM seems quite attractive. Several electrodiagnostic criteria sets have been devised to improve CIDP diagnosis [27]. The EFNS/PNS criteria, in a comparative study with two other electrodiagnostic criteria sets utilizing data from four motor nerve, showed, when considering only the definite/probable CIDP categories, an optimal balance between sensitivity (81.3%) and specificity (96.2%) versus patients with distal sensory or sensory-motor axonal polyneuropathies [12]. The introduction of the possible CIDP electrodiagnostic category improved the sensitivity by 15.4% but worsened specificity to 69% [12]. In the current study, SVM showed for CIDP a sensitivity of 79.8% and a specificity of 96.2% in the training test by using tenfold cross-validation procedure and a sensitivity of 80% and a specificity of 97.1% in the test set. These results are very similar to those found with the EFSN/PNS criteria, but it should be underlined that 37% of patients with the CIDP reference diagnosis were, after data pre-processing, unclassifiable or possible CIDP according to the EFNS/PNS criteria. Moreover, specificity and sensitivity were assessed versus other polyneuropathies considered demyelinating such as CMT1A (100% of patients fulfilling the electrodiagnostic criteria for definite CIDP) and HNPP (12.5% of patients fulfilling the definite and probable categories). Although the EFNS/PNS criteria are the most frequently used in CIDP research worldwide [28], in a recent survey of 100 community

Table 2 Accuracy on training set of supervised learning algorithms (based on tenfold cross-validation procedure) and of neurophysiologists for the CIDP diagnostic class. Each column is color formatted to reflect the best (green) and worst (red) performance for each indicator of the supervised learning algorithms and neurophysiologists in the five diagnostic classes

	Ranking	Euclidean Distance	Accuracy
SVM	1	0.263	92.3%
MLR	2	0.357	89.5%
SDA	3	0.423	89.7%
EXP 3	4	0.444	87.4%
EXP 1	5	0.552	85.9%
EXP 2	6	0.559	80.8%
TR 1	7	0.674	84.1%
TR 3	8	0.787	75.1%
TR 2	9	0.986	63.8%

SVM support vector machine, *MLR* multinomial logistic regression, *SDA* shrinkage discriminant analysis, *EXP* expert neurophysiologist, *TR* trainee neurophysiologist

Table 3 Results of the McNemar with Edwards's correction test on training set. Significance of the difference between false-negative and false-positive values for supervised learning algorithms and neurophysiologists in the chronic inflammatory demyelinating polyneuropathy diagnostic class

	CIDP false negative	CIDP false positive	McNemar Edwards statistic	*P* value
SDA	36	4	24.03	0.00
MLR	23	18	0.39	0.53*
SVM	18	12	0.83	0.36*
EXP1	44	11	18.62	0.00
EXP2	18	57	19.25	0.00
EXP3	31	18	2.94	0.09
TR1	57	11	41.95	0.00
TR2	46	95	16.34	0.00
TR3	43	54	1.03	0.31*

neurologists in the USA, only 13% employed these guidelines and approximately half endorsed electrophysiological parameters that do not help in CIDP electrodiagnosis, underlining the difficulties in the use of diagnostic guidelines and the problems in the interpretation of neurophysiologic findings in the everyday practice [8]. Interestingly, even though all CMT1A patients could be classifiable as definite CIDP according to EFSN/PNS electrodiagnostic criteria, SVM was able to catch the characteristic differential features in a very high percentage of CMT1A subjects. The above observations indicate the possible utility of SVM as a reliable support decision tool. This study has limitations. Because of the retrospective nature with a high number of missing values in the initial dataset, we had to limit the number of nerves and electrophysiological parameters to analyze. Rather than removing variables with missing data, or filling missing values by using a per-class median imputation strategy, non-ignorable missing data models or imputation techniques in multilevel data structures could have been used. However, applying a model-based imputation approach would require the knowledge of the missing-data mechanism and we leave this modeling approach to future works. A prospective study with the same number of nerves and electrophysiological parameters studied in each subject should be the ideal method to validate our results. Although we deem that the electrodiagnosis is an extension of clinical examination, this study was based only on electrophysiological findings and clinical information were not given to neurophysiologists, which could have used them to improve their diagnostic accuracy, and SLAs. It could be theoretically possible "to feed" the SLAs with information associated with history and clinical examination and ensure that both neurophysiologists and SLAs could work with the "whole picture" of the subject. However, this would require the use of several scales that should have been validated and uniformly applied by the participating centers. Moreover, the reported high diagnostic error rate in CIDP was based on the incorrect interpretation of electrodiagnostic data [7], and our first aim was to verify how neurophysiologists and SLAs performed only on the basis of the electrophysiological data. On the other way round, the above limitations can be seen as points of strength of the study as SVM correctly classified a very high number of subjects by a limited electrophysiological dataset and without adjunctive clinical information.

Table 4 Results on test set of performance indicators of support vector machine and neurophysiologists for each diagnostic class. Each column is color formatted to reflect the best (green) and worst (red) performance for each indicator of the supervised learning algorithms and neurophysiologists in each diagnostic class

	Ranking	Euclidean Distance	Overall Accuracy	Precision					Sensitivity					Specificity				
				CONTROL	CIDP	CMT1A	HNPP	DN	CONTROL	CIDP	CMT1A	HNPP	DN	CONTROL	CIDP	CMT1A	HNPP	DN
SVM	1	0.374	93.2%	100.0%	88.9%	100.0%	80.0%	85.7%	100.0%	80.0%	100.0%	100.0%	85.7%	100.0%	97.1%	100.0%	97.5%	97.3%
EXP1	2	0.843	81.8%	100.0%	85.7%	100.0%	57.1%	55.6%	90.0%	60.0%	92.3%	100.0%	71.4%	100.0%	97.1%	100.0%	92.5%	89.2%
EXP3	3	0.866	86.4%	100.0%	88.9%	100.0%	57.1%	60.0%	100.0%	80.0%	100.0%	100.0%	42.9%	100.0%	97.1%	100.0%	92.5%	94.6%
TR1	4	0.963	79.5%	100.0%	100.0%	100.0%	75.0%	42.9%	100.0%	40.0%	92.3%	75.0%	85.7%	100.0%	100.0%	100.0%	97.5%	78.4%
TR3	5	0.981	77.3%	100.0%	66.7%	100.0%	42.9%	66.7%	100.0%	60.0%	69.2%	75.0%	85.7%	100.0%	91.2%	100.0%	90.0%	91.9%
EXP2	6	1.639	75.0%	100.0%	50.0%	92.3%	0.0%	71.4%	100.0%	60.0%	92.3%	0.0%	71.4%	100.0%	82.4%	96.8%	95.0%	94.6%
TR2	7	1.729	54.5%	90.9%	15.4%	83.3%	40.0%	55.6%	100.0%	20.0%	38.5%	50.0%	71.4%	97.1%	67.6%	96.8%	92.5%	89.2%

SDA shrinkage discriminant analysis, *MLR* multinomial logistic regression, *SVM* support vector machine, *EXP* expert neurophysiologist, *TR* trainee neurophysiologist, *CIDP* chronic inflammatory polyradiculoneuropathy, *CMT1A* Charcot-Marie-Tooth disease type 1A, *HNPP* hereditary neuropathy with liability to pressure palsy, *DPN* diabetic polyneuropathy

Conclusions

SVM seems to be a promising tool to increase the electrodiagnostic accuracy in polyneuropathies and has the advantage of assigning to each subject the probability of belonging to the different diagnostic classes.

We are not proposing an automatic approach to electrodiagnosis of polyneuropathies that replaces the final global judgment of the physician, but we think that the use of SLAs in electrodiagnosis should be exploited to possibly provide a diagnostic support system especially helpful for the less experienced practitioners.

Authors' contributions AU and GA equally contributed to the study. AU, GA, and LI designed the study and AU acted as study supervisor. FM, YS, LM, ST, AT, SK, and LS collected the data. GA made the statistical analysis. AU, GA, and LI analyzed and interpreted the results. All the authors contributed to drafting and revising the manuscript and gave their approval to the final version of the manuscript.

Data Availability Deidentified participant data, data dictionary and R script for the implementation of the statistical analysis will be available on request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethical approval The study was carried out in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University Federico II of Naples (320/17) and the institutional Medical Ethics Research Committee of each center.

Consent to participate Not applicable.

Consent for publication All subjects signed a written informed consent that allowed the utilization of electrophysiological data for research purposes.

Informed consent All subjects signed a written informed consent that allowed the utilization of electrophysiological data for research purposes.

Code availability Not applicable (software application or custom code).

References

1. Martyn CN, Hughes RA (1997) Epidemiology of peripheral neuropathy. J Neurol Neurosurg Psychiatry 62:310–318
2. Dyck PJ, Dyck PJ, Grant IA, Fealey RD (1996) Ten steps in characterizing and diagnosing patients with peripheral neuropathy. Neurology 47:10–17
3. England JD, Asbury AK (2004) Peripheral neuropathy. Lancet 363: 2151–2161
4. England JD, Gronseth GS, Franklin G, Miller RG, Asbury AK, Carter GT, Cohen JA, Fisher MA, Howard JF, Kinsella LJ, Latov N, Lewis RA, Low PA, Sumner AJ (2005) Distal symmetric polyneuropathy: a definition for clinical research, report of the AAN, AAEM, and AAPM&R. Neurology 64:199–207
5. Watson JC, Dyck PJ (2015) Peripheral neuropathy: a practical approach to diagnosis and symptom management. Mayo Clin Proc 90:940–945
6. Allen JA, Lewis RA (2015) CIDP diagnostic pitfalls and perception of treatment benefit. Neurology 85:498–504
7. Allen JA, Lewis RA (2018) Electrodiagnostic errors contribute to chronic inflammatory demyelinating polyneuropathy misdiagnosis. Muscle Nerve 57:542–549
8. Gelinas D, Katz J, Nisbet B, England JD (2018) Current practice pattern in CIDP: a cross-sectional survey of neurologists in the United States. J Neurol Sci 397:84–91
9. Nicolas G, Maisonobe T, Le Forestier N, Léger JM, Bouche P (2002) Proposed revised electrophysiological criteria for chronic inflammatory demyelinating polyradiculoneuropathy. Muscle Nerve 25:26–30
10. Van den Bergh PY, Pieret F (2004) Electrodiagnostic criteria for acute and chronic inflammatory demyelinating polyradiculoneuropathy. Muscle Nerve 29:565–574
11. Wilson J, Chawla J, Fisher M (2005) Sensitivity and specificity of electrodiagnostic criteria for CIDP using ROC curves: comparison to patients with diabetic and MGUS associated neuropathies. J Neurol Sci 231:19–28
12. Rajabally YA, Nicolas G, Pieret F, Bouche P, van den Bergh PY (2009) Validity of diagnostic criteria for chronic inflammatory demyelinating polyneuropathy: a multicentre European study. J Neurol Neurosurg Psychiatry 80:1364–1368
13. Rondinelli RD, Stolov WC, Fujimoto WY, Osberg JS (1988) Electrodiagnosis of diabetic peripheral polyneuropathy. A multivariate analytic approach. Am J Phys Med Rehabil 67:12–23
14. Rondinelli RD, Robinson LR, Hassanein KM, Stolov WC, Fujimoto WY, Rubner DE (1994) Further studies on the electrodiagnosis of diabetic peripheral polyneuropathy using discriminant function analysis. Am J Phys Med Rehabil 73:116–1123
15. Uncini A, Ippoliti L, Shahrizaila N, Sekiguchi Y, Kuwabara S (2017) Optimizing the electrodiagnostic accuracy in Guillain-Barré syndrome subtypes: criteria sets and sparse linear discriminant analysis. Clin Neurophysiol 128:1176–1183
16. Joint Task Force of the EFNS and the PNS (2010) European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of chronic inflammatory demyelinating polyneuropathy: report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society-First Revision. J Peripher Nerv Syst 2010(15):1–9
17. Saperstein DS, Amato AA, Wolfe GI, Katz JS, Nations SP, Jackson CE et al (1999) Multifocal acquired demyelinative sensory and motor neuropathy: the Lewis-Sumner syndrome. Muscle Nerve 22:560–566
18. Katz JS, Saperstein DS, Gronseth G, Amato AA, Barohn RJ (2000) Distal acquired demyelinating symmetric neuropathy. Neurology 54:615–620
19. Dyck PJ, Albers JW, Andersen H, Arezzo JC, Biessels GJ, Bril V, Feldman EL, Litchy WJ, O'Brien PC, Russell JW, on behalf of The Toronto Expert Panel on Diabetic Neuropathy (2011) Toronto expert panel on diabetic neuropathy. Diabetic polyneuropathies: update on research definition, diagnostic criteria and estimation of severity. Diabetes Metab Res Rev 27:620–628
20. Vapnik VN (2000) The nature of statistical learning theory. Springer-Verlag
21. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining; inference and prediction, 2nd edn. Springer-Verlag, pp 241–251
22. Flach P (2012) Machine learning. The art and science of algorithms that make sense of data. Cambridge University Press, pp 53–58

23. Edwards A (1948) Note on the 'correction for continuity' in testing the significance of the difference between correlated proportions. Psychometrika 13:185–187
24. Enders CK (2003) Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. Psychol Methods 8:322–337
25. Theodoridis S (2015) Machine Learning: A Bayesian and Optimization Perspective. Academic Press p 314
26. James G, Vitten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning: with applications in R. Springer, pp 24–26
27. Bromberg M (2011) Review of the evolution of the electrodiagnostic criteria for chronic inflammatory demyelinating neuropathy. Muscle Nerve 43:780–794
28. Rajabally YA, Fowle AJ, Van den Bergh PY (2015) Which criteria for research in chronic inflammatory demyelinating polyneuropathy? An analysis of current practice. Muscle Nerve 51:932–933