Deep Metric Learning With Locality Sensitive Mining for Self-Correcting Source Separation of Neural Spiking Signals

Alexander Kenneth Clarke, *Student Member, IEEE*, and Dario Farina, *Fellow, IEEE*

***Abstract*—Automated source separation algorithms have become a central tool in neuroengineering and neuroscience, where they are used to decompose neurophysiological signal into its constituent spiking sources. However, in noisy or highly multivariate recordings these decomposition techniques often make a large number of errors. Such mistakes degrade online human-machine interfacing methods and require costly post-hoc manual cleaning in the offline setting. In this article we propose an automated error correction methodology using a deep metric learning (DML) framework, generating embedding spaces in which spiking events can be both identified and assigned to their respective sources. Furthermore, we investigate the relative ability of different DML techniques to preserve the intraclass semantic structure needed to identify incorrect class labels in neurophysiological time series. Motivated by this analysis, we propose locality sensitive mining, an easily implemented sampling-based augmentation to typical DML losses which substantially improves the local semantic structure of the embedding space. We demonstrate the utility of this method to generate embedding spaces which can be used to automatically identify incorrectly labeled spiking events with high accuracy.**

***Index Terms*—Blind source separation (BSS), deep learning, deep metric learning (DML), surface electromyography (EMG).**

I. Introduction

NEUROPHYSIOLOGICAL signals can generally be characterized as an additive mixture of repeating events from different sources, such as the motor unit activation potential (MUAP) in electromyography (EMG) or the spike potential in microelectrode cortical recordings [1], [2]. The ensemble of these activation events constitute neural codes that can provide direct insights into a neurological system of interest [3], [4], [5]. Consequently, the extraction of individual sources from noisy signal by blind source separation (BSS) algorithms has long been a major focus in computational neuroscience [4], [6], [7], [8]. Modern BSS algorithms leverage highly multivariate data, such as the output from high-density electrode arrays [9], [10], to identify the contributions of many individual spiking sources [4], [11], [12]. These sources also provide an extremely clean control signal for human-machine interfacing applications when compared to the original bulk signal [5], [13], [14].

Despite these successes, BSS algorithms frequently make mistakes when labeling source activations, which can be difficult to identify and correct automatically [15], [16]. As a result, applications that use neurophysiological data still broadly rely on bulk signal [17], [18], [19]. In offline decomposition, a degree of manual or semi-automated post-hoc cleaning is commonly employed, often using additional knowledge about the system of interest, for example the temporal statistics of the sources [20]. The nature of this manual error correction generally relates to the mixing system of interest, for example, intracortical and intramuscular EMG decompositions generally require post-hoc examination of source classes [8], [21], whilst surface EMG (sEMG) decompositions also require further inspection of individual neural activations [22]. However, whilst accurate, manual cleaning is constrained to offline methods and is an extremely time-consuming process [23]. In response, modern source separation pipelines are increasingly using additional automated post-processing steps [16], [20], [24]. A contemporary direction is to use the noisy BSS-derived labels to train a neural network classifier, which can then outperform the noisy label it was trained on [25], [26], [27]. However, such supervised deep learning approaches generally suffer from a critical threshold where the initial label noise overwhelms the ability of the model to self-correct [28].

A different approach to leveraging deep learning for neurophysiological source separation is through deep metric learning (DML), in which a model is trained to learn the relative similarity or dissimilarity between data samples rather than to directly classify them [29], [30]. In this formulation, the neural network acts as an efficient featurizer, taking high-dimensional neurophysiological input and extracting a low-dimensional embedding in which individual spiking events are easily assigned to their respective sources. In recent years DML has seen particular success in person reidentification tasks [31], [32], medical image classification [33], [34] and digital

Manuscript received 29 June 2022; revised 9 December 2022, 8 March 2023, and 6 June 2023; accepted 25 June 2023. The work of Alexander Kenneth Clarke was supported by the EPSRC Centre for Neurotechnology under Grant EP/L016737/1. The work of Dario Farina was supported by the EPSRC Transformative Healthcare Technologies 2050 under Grant EP/T020970/1. This article was recommended by Associate Editor P. Shi. *(Corresponding author: Dario Farina.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Politecnico di Torino Research Ethics Board.

The authors are with the Department of Bioengineering, Imperial College London, SW7 2BX London, U.K. (e-mail: d.farina@imperial.ac.uk; a.clarke18@imperial.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2023.3290825.

Digital Object Identifier 10.1109/TCYB.2023.3290825

pathology [35], [36]. In the theoretical domain, problems with optimization stability and computational inefficiency have been largely solved through a combination of sampling strategies [37], [38], [39] and better-distance metrics [29], [40], [41], [42]. DML is attractive in the context of spiking events, which are stable over time given fixed conditions. This means that for a well trained model, each spiking event from the same source should cluster in the same location in the embedding space, with the DML model theoretically only needing to remove the distortions caused by noise and superposition events. DML also has lower-data demands and a better tolerance for class-imbalance when compared to typical neural network-based classifiers [43], both of which are common issues in neurophysiological time series processing.

In this article we have two main objectives for a DML-trained featurizer. First, we want neural events from different sources to map to different regions of the embedding space so they can be discriminated, that is, we want good interclass variance. Second, we want neural events from the same source to preserve some intraclass structure for the purposes of identifying label noise. This second objective is complicated by the fact that maintaining a degree of intraclass variance is not usually the focus of DML methods, which are generally more interested in interclass separation [44]. Current losses tend to incentivize the model to ignore intraclass semantic differences and collapse the embedding down to a tight cluster [39]. To mitigate this, we propose a simple method of augmenting DML training to prevent this happening, allowing the embedding to be used for more than just class separation.

In summary, the main contributions of this article are as follows.

1) We robustly demonstrate that DML is an effective method of building a neural network featurizer for source separation of neurophysiological signals. Using sEMG signal, we show that the activations from different sources clearly cluster, making class discrimination by simple clustering methods trivial.
2) We implement a number of popular DML methods, such as N-pair loss and angular loss [40], [42], demonstrating that they generate embedding spaces with excellent class discriminability, but poor local semantic structure. We show that more contemporary methods which preserve such structure can be used to identify outliers within an artificially corrupted neurophysiological signal.
3) We propose locality sensitive mining (LSM), an easily implemented sampling-based method of maintaining intraclass structure that can be implemented in a variety of different DML paradigms. We go on to show that for neurophysiological signal LSM outperforms other contemporary methods of preserving local semantic structure, such as multilevel distance regularization (MDR) [44], whilst having a very simple implementation that can be added to many different DML losses.
4) For the first time we are aware of in the literature, we show that DML methods which maintain intraclass semantic structure have clear practical utility for identifying incorrectly labeled outliers in the class clusters. As an example, we leverage this to build a self-correcting source separation pipeline for neural spiking signals, which we call DeepDecomp.

II. Related Work

DML has classically been employed in discrete high-dimensional data types, such as images, rather than in the time series domain. This has not been due to a lack of theoretical grounding [45], [46], [47], but relates more to the difficulty of building discrete class pairings for signals when components have variable temporal dynamics, which may explain the lack of DML methods for neurophysiological signal processing in the literature. However, when the independent factors in the generative process have relatively short and stereotyped responses, such as spiking neurons, it becomes straightforward to break the signal into windows centered on each activation. In this case, the problem becomes similar to face reidentification, where the neural network needs to learn an invariance to confounding factors when bringing images of the same face closer together in the embedding space [31], [32], [48]. There has also been some work in face reidentification to make training robust to label noise, using external models to build metrics of label quality [49], [50], [51], [52]. Unfortunately, these methods are generally reliant on an extremely large amount of data to be effective, which precludes their use in most neurophysiological datasets.

An alternative solution to the problem of label noise is to preserve a richer intraclass embedding, such that semantic similarity or dissimilarity between samples from the same class is better preserved, allowing identification of outliers. Early work on DML losses generally either ignored intraclass variance or actively sought to reduce it [41], [53]. Some proxy-based losses, which pull samples towards shared class proxies rather than other in-class samples [54], [55], have been designed in part with the intention of reducing overfitting by preventing intraclass collapse [56], [57]. There has also been some work in preserving intraclass characteristics using generative models, such as variational autoencoders [58], [59], with the main objective of improving generalization performance by excluding features that are not shared across a class when making embedding decisions.

More recently there have been some explicit attempts to preserve local semantic structure in the embedding space, as illustrated in Fig. 1. DML with self-supervised ranking uses an auxiliary term based on self-supervised learning [44], using a set of transform functions to augment samples to different degrees, such that more perturbed samples are embedded further away from the original image. However, this requires domain knowledge when sculpting the transform functions, for example the authors perturb image scale, viewpoint and color in their task of bird identification. Selecting similar transformations for neural signals would not be a simple task. An alternative approach is MDR, which also adds an auxiliary loss term, albeit one which aims to prevent class collapse by using proxies set at multiple distances from each sample [60].

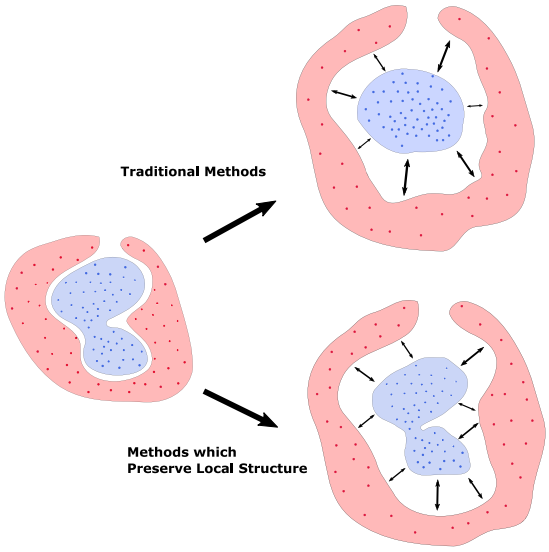

Fig. 1. Effect of traditional DML losses versus those designed to preserve local semantic structure. Traditional methods tend to obliterate intraclass variance in the embedding of samples from the same class (in blue), whilst recent directions instead preserve this, while maintaining good separation from samples belonging to different classes (in red).

III. Preliminaries

A. Deep Metric Learning

The basic aim of DML is to train a neural network to map a sample taken from one of C classes to an embedding vector $\mathbf{x}$, such that for an arbitrarily selected anchor embedding $\mathbf{x_a}$, samples from the same class $\mathbf{x_p}$ are embedded closer than samples from a different class $\mathbf{x_n}$, as measured by some distance metric $\mathbf{D}$. $\mathbf{D}$ can be a number of different metrics, such as the Euclidean distance, cosine similarity or Kullback–Leibler divergence [61]. Commonly the loss function is formulated in terms of a relative distance between positive pairs (samples from the same class) and negative pairs (samples from different classes), such that

$$\mathbf{D}\big(\mathbf{x_a}, \mathbf{x_p}\big) + m \leq \mathbf{D}(\mathbf{x_a}, \mathbf{x_n}) \tag{1}$$

where m is a margin term that specifies the objective interclass separation.

After a small amount of optimization, the bulk of negative pairs will be further away than positive pairs, making most training examples in a batch uninformative [37]. This can be partially mitigated by mining strategies; calculating the DML loss using only pairings from each batch selected using some heuristic based on the embedding space [62], for example selecting only pairings where $\mathbf{x_a}$ is further away from $\mathbf{x_p}$ than $\mathbf{x_n}$. A related approach is to use multiple negative pairings N for each anchor term, as proposed originally in the N-pair loss [40]

$$L_{\text{NPAIR}} = \frac{1}{M}\sum_{\mathbf{x}_a \in B}^{M} \log\left[1 + \sum_{\mathbf{x}_n \in B\setminus(\mathbf{x}_a,\mathbf{x}_p)}^{N} \exp\big(f_{a,p,n}\big)\right] \tag{2}$$

where $f_{a,p,n}$ is the DML loss and B is a batch of size M.

B. Locality Sensitive Mining

Whilst current methods of selecting pairings from B are suitable if the objective is to maximize interclass distance within the embedding space, they have the side effect of penalizing local semantic structure within a class [54]. When positive pairings are selected randomly, the network will generally learn to embed all samples from a particular class into a dense point [57].

To preserve local semantic structure, we instead propose a new mining procedure, LSM, using a top-k algorithm to select the k closest $\mathbf{x_n}$ to $\mathbf{x_a}$ that belong to a different class and using only the closest $\mathbf{x_p}$. GPU implementations of top-k algorithms have become extremely efficient in recent years, due to their increasing use within machine learning applications [63]. In the N-pair formulation this mining approach can be written as

$$\begin{aligned} l_{LSM} &= \frac{1}{M}\sum_{\mathbf{x}_a \in B}^{M} \sum_{\mathbf{x}_p \in B\setminus(\mathbf{x}_a,\mathbf{x}_n)} \mathbf{1}_{\mathbf{x}_p \in \Upsilon} \log\big[1+g\big] \\ g &= \sum_{\mathbf{x}_n \in B\setminus(\mathbf{x}_a,\mathbf{x}_p)} \mathbf{1}_{\mathbf{x}_p \in \Theta} \exp\big(f_{a,p,n}\big) \end{aligned} \tag{3}$$

where $\Upsilon = \text{argmax}(\angle(\mathbf{x}_a, \mathbf{x}_p))$ is the argmax set of the pairwise cosine similarity $\angle$ between $\mathbf{x_a}$ and its associated set $\mathbf{x_p}$ in B and $\Theta = \text{top}k(\angle(\mathbf{x}_a, \mathbf{x}_n))$ is the top-k values of the ordered pairwise cosine similarity $\angle$ between $\mathbf{x_a}$ and its associated set $\mathbf{x_n}$ in B. $\mathbf{1}$ is the indicator function.

Like other mining procedures, LSM can be combined with many DML losses. As examples, we selected N-pair loss combined with the popular angular loss as a distance metric [42].

C. Angular Loss

Angular loss is a stable geometric reformulation of the angular distance metric, which constructs a right-angled triangle with $\mathbf{x_n}$ and the midpoint between $\mathbf{x_a}$ and $\mathbf{x_p}$, with the final vertex being the point on the semicircle joining $\mathbf{x_a}$ and $\mathbf{x_p}$ [42]. By dropping constant terms, this geometric relationship can be used as the DML loss $f_{a,p,n}$ in (2), expressed as

$$f_{a,p,n} = 4tan^2\alpha\big(\mathbf{x}_a + \mathbf{x}_p\big)^T\mathbf{x}_n - 2\Big(1+\tan^2\alpha\Big)\mathbf{x}_a^T\mathbf{x}_p \tag{4}$$

where α is an angle in radians which sets the upper accepted bounds of the loss, analogous to m in (1).

D. Cross Entropy Loss

A major difficulty in optimizing a model to detect spiking events is that spikes are relatively rare, meaning the data has a strong class imbalance toward windows with no activity [25]. This is particularly problematic in DML as the losses generally have local optima of mapping all samples to a single location in the embedding space. We found that the addition of a small auxiliary cross-entropy (CE) term with temperature was useful in avoiding this trivial solution

$$l_C = -\frac{1}{M}\sum_{i}^{M}\sum_{c}^{C} \mathbf{1}_{y_i \in C_c} \log \frac{\exp\Big[\frac{1}{\gamma}\mathbf{z}_i\Big]}{\sum_j^C \exp\Big[\frac{1}{\gamma}\mathbf{z}_j\Big]} \tag{5}$$

where $\mathbf{z} = (\mathbf{W}/\|\mathbf{W}\|)_2\mathbf{x}$ and $\mathbf{W}$ is a trainable matrix that compresses the embedding vector down to a dimension C vector for comparison with the one-hot encoded class labels. The impact of the CE term was weighted by a coefficient such that

$$l = l_{DML} + \tau l_C. \tag{6}$$

E. Neurophysiological Signal Decomposition

The timestamps used by DeepDecomp can be generated using a wide variety of manual and automated processes, however, for this study the gradient convolution kernel compensation (gCKC) algorithm was selected due to its strong performance in HD-sEMG signal decomposition [64], [65]. In the gCKC framework for BSS, the vector of spiking sources s at time t are first extended with L delayed versions of themselves, allowing the mixing problem, which is convolutive in most neurophysiological settings, to be written in instantaneous form

$$x(t) = \mathbf{H}\tilde{s}(t-l) + \omega(t) \tag{7}$$

where the signal observation vector x at time t is a linear mixture parameterized by the operation of the mixing matrix $\mathbf{H}$ on the extended source vector $\tilde{s}$ plus noise ω. In practice both the observation and source vectors are additionally extended with a further R values for reason of numerical stability during the source separation procedure.

Unlike independent component analysis methods which seek to directly estimate a separation vector for each source, gCKC seeks to include the additional statistical information that the spiking sources generate repetitive events within the signal. Sources are instead estimated indirectly using a linear minimum mean square error estimator, with the estimated jth source $\hat{s}_j$ at time point t given by

$$\hat{s}_j(t) = \hat{c}^T_{\mathbf{s}_j\tilde{\mathbf{x}}}\mathbf{C}^{-1}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{x}(t) \tag{8}$$

where $\hat{\mathbf{c}}^T_{\mathbf{s}_j\tilde{\mathbf{x}}}$ is the transposed cross-correlation vector between an activation of the jth source and extended HD-sEMG matrix and $\mathbf{C}^{-1}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ is the inverted autocorrelation matrix of the extended HD-sEMG matrix $\tilde{\mathbf{x}}$.

The vector $\hat{c}^T_{\mathbf{s}_j\tilde{\mathbf{x}}}$ is usually initialized with a time point likely to contain a source activation, which can be estimated by, for example, the Mahalanobis distance calculated on the signal [65]. Once selected, $\hat{\mathbf{c}}^T_{\mathbf{s}_j\tilde{\mathbf{x}}}$ is then optimized to find the rest of the source's signal contributions. This can be done with either a fixed-point algorithm as in [16] or in the gCKC formulation by gradient descent

$$\mathbf{c}^{\circ}_{\hat{\mathbf{s}}_j\tilde{\mathbf{x}}} = \mathbf{c}_{\hat{\mathbf{s}}_j\tilde{\mathbf{x}}} - \alpha\sum_t \frac{\partial f\left(\hat{\mathbf{s}}_j(t)\right)}{\partial \hat{\mathbf{s}}_j(t)}\tilde{\mathbf{x}}(t) \tag{9}$$

where $\mathbf{c}^{\circ}_{\hat{\mathbf{s}}_j\tilde{\mathbf{x}}}$ is the updated cross-correlation vector, α is the learning rate, and $f(\cdot)$ a contrast function designed to estimate the nongaussianity of the output source in a similar fashion to independent component analysis. Optimized sources can then be converted to timestamps using a linear threshold or a two-class k means clustering algorithm.

IV. Validation Methodology

A. Evaluated Methods

We compared a number of different DML methods, examining the utility of their respective embedding spaces for identifying two types of label noise and classifying unseen neural activations. The methods studied were as follows.

N-Pair Loss (NL) [40]: The original N-pair loss using Euclidean distance, where each sample in a batch is paired with one randomly selected in-batch positive and all in-batch negatives. A small auxiliary CE term, l_C in (5), was added to stabilize training, with both γ and τ set to 0.1.

Angular Loss With NL (AL) [42]: The N-pair loss (with the same auxiliary CE term), but using angular loss instead of Euclidean distance as the distance metric. α to 0.25 radians in both the cleaning and refitting stages. It should be noted that this method also acts as an ablation of the LSM component of training.

AL With LSM: The same loss as AL (with the same hyperparameters), but using the proposed mining approach instead. k was set to 5 throughout, which was selected after a hyperparameter study also detailed in this article.

MDR [44]: We also compare our approach with the recent MDR, which also aims to preserve intraclass variance, albeit mainly from the perspective of improving generalization performance.

CE: As a baseline we also included a classifier trained using CE, that is, purely l_C in (5) with both γ and τ set to 1. The embedding dimension is simply the output of the layer before $\mathbf{W}$.

B. Experiments

In experiment 1 we evaluated the ability of the models to clean a label set corrupted by feature-dependent noise, where label flipping probability is related to its associated features [66]. In the context of source-separated HD-sEMG, this most commonly occurs as a false-positive, where a separation vector incorrectly assigns a high probability of an in-class MUAP being present when it is not, that is, a noise class or other MU class label is flipped to the MU class of interest. To simulate this effect, we corrupted the label set by generating an artificially noisy separation vector for each MU class by randomly selecting 15 MUAP labels from that class and using the average of the associated extended HD-sEMG vectors to generate a linear minimum mean square error prediction on the extended HD-sEMG matrix. A two-class k-means clustering algorithm was then used to parameterize a linear threshold to find activations, creating a label set with a high degree of feature-dependent noise. Five levels of increasing difficulty were generated by taking an amount of false positives corresponding to 10%/20%/30%/40%/50% of the number of true labels, selected at random from the set of false positives.

In experiment 2 the models were instead evaluated on class-dependent label noise, when the probability of a label flipping to another class is stable across all labels in the class [66]. As for most neural spiking signals, in HD-sEMG source separation this error generally occurs when the separation vectors are very similar, usually due to similar MUAP waveform shapes

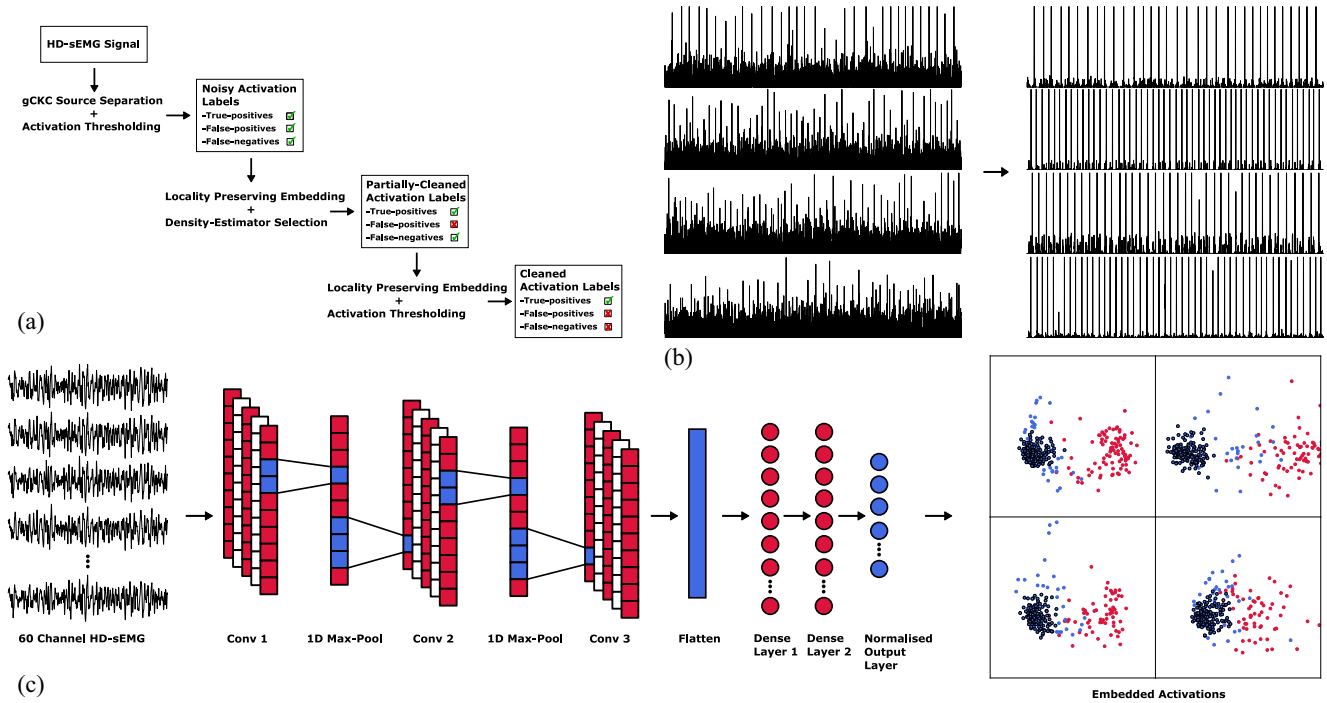

Fig. 2. (a) DeepDecomp, an example pipeline by which the noisy activation labels found from source—separating the high-density sEMG signal are cleaned. The model is trained twice—a cleaning phase to find the false positive labels and a refitting phase to find the false negatives. After the refitting phase the predicted class activity is much cleaner than that of the original source separation algorithm, as seen in (b). (c) Convolutional neural network trained by a DML loss to find neural activations. Windows of neurophysiological time series are embedded into a low-dimensional space which can be used to source separate and, if a method that preserves local semantic structure is used, to clean a noisy label set.

between two MU classes. This can be simulated by transferring a percentage of labels to a similar MU class. This was done by first averaging the MUAPs of each MU class and then cross-correlating these averages with the average MUAP of every other class in the recording, with 10%/20%/30%/40% of the class labels transferred to the class with the highest value. If labels had already been transferred to the closest class then the next closest class was selected until all classes had label transfers. A maximum label corruption of 40% was used to preserve the concept of a majority true and minority false class.

After training with the artificially corrupted label sets, methods which successfully preserved intraclass variance should have a dense central cluster due to the methods of label corruption. To avoid bias from manual selection and to demonstrate a fully automated pipeline, we elected to automatically assign samples to this cluster using a simple density-estimator. As the cleaning process could potentially bias the label set by incorrectly removing in-class outliers, we also retrained the models with the cleaned sets and used the models to find unlabeled spiking events. This also allowed demonstration of the generalization performance of the models trained with the different loss functions. Finally we explored the impact of the k hyperparameter within LSM by rerunning experiment 1 with multiple values of k.

As a demonstration of the practical utility of the proposed method for building a self-correcting source separation pipeline for neural spiking signals, we give an example pipeline, DeepDecomp, which automate the cleaning of the output of a noisy gCKC decomposition of sEMG signal [Fig. 2(a)]. By preserving local semantic structure, noisy activation label sets can be accurately cleaned [Fig. 2(b)].

C. Model and Training

To convert the source-separated HD-sEMG signal into labeled windows, first each channel of the HD-sEMG signal was standardized by z-scoring and then cut into overlapping 80-sample wide windows at a stride of 1. Each window was then labeled by reference to the predicted source activity at the final sample of the window. This meant the bulk of windows were labeled as part of the inactive class due to the sparse nature of motor neuron spiking. Due to this serious class imbalance, each minibatch was created from the entirety of the windows labeled as containing a motor neuron spike, with an additional 256 samples randomly selected from the inactive class. Each class assignment was then converted to a one-hot representation, the bulk of which had only one class active at any one time, although rarely two activations would occur simultaneously on the same time-point. As the richness of the intraclass embedding of the inactive class windows was not of any great concern, the embeddings of these windows were not used as anchor samples when calculating the DML component of the tested methods, although they were used as negative samples and in the calculation of l_C.

A convolutional neural network implemented using the tensorflow machine learning library in python was used as the embedding model [Fig. 2(c)]. Grid search optimization was used to select specific model architecture hyperparameters.

TABLE I
CLEANING RESULTS FOR DIFFERENT LEVELS OF FEATURE-DEPENDENT LABEL NOISE

Method	Level of Added Label Noise: 10%		20%		30%		40%		50%	
	True and False Labels after Cleaning (% of Original Set)									
	True	False	True	False	True	False	True	False	True	False
CE	90.2	7.9	88.8	16.2	89.3	25.1	90.9	32.6	79.7	35.6
NL[40]	92.8	7.2	90.3	14.8	87.8	20.0	86.9	24.4	77.6	25.8
AL[42]	66.7	2.2	58.4	5.7	54.6	8.0	45.1	8.0	42.0	10.8
MDR[44]	72.1	1.4	68.6	1.6	65.7	2.3	62.0	1.9	59.9	3.2
LSM	**84.1**	**1.0**	**78.3**	**1.0**	**80.3**	**1.5**	**76.6**	**1.4**	**74.1**	**2.3**

Convolution steps used a 1-D 3-sample wide kernel, with 32 filters and a drop-out of 0.2. 1-D max-pooling was completed with 2-sample wide kernels. Each densely connected layers had 64 neurons and a drop-out percentage of 0.5 during training. Both the convolution and densely connected layers used ReLU activation functions. Finally the output of the last densely connected layer was densely connected to a bias and activation-free embedding layer of 8 neurons wide, which was then divided by its L2 norm. This was an intentionally low-dimension embedding compared to standard DML due to the desire to avoid dimensionality issues during the clustering steps in the refitting phase. The additional matrix $\mathbf{W}$ used in the categorical CE was initialized with truncated normal noise, whilst the weights of the neural network layers was initialized by glorot uniform. The Adam optimization algorithm at a learning rate of 0.001 was then used to train the model over 500 epochs for both cleaning and refitting stages.

When the model was used to find neural activations, the cosine similarity of each sample to the average embedding vectors of classes was calculated, with the spiking activations assigned to MUs by way of a two-class k-means clustering algorithm. These activation labels could then be compared to the precorrupted data using the rate of agreement (RoA) metric [1], a percentage defined as the number of true positive matches divided by the total number of true positives, false positives and false negatives. When the model was instead used to clean the corrupted label set, a simple density estimator was used. First a local scale value v was estimated by finding the mean cosine similarity of the each embedding vector to its 20 nearest neighbors and taking a median of this value across all vectors. For each label the number of other labels with a cosine similarity more than v was found and the label with the highest number of neighbors was selected as the center of the cluster. All labels within a cosine similarity higher than v were then added to the refitting training set. This simple approach was generally adequate for quickly finding the densest region of the embedding space, which was usually the cluster of true labels.

D. Experimental Dataset

The HD-sEMG dataset consisted of 20-s recordings taken from the dominant tibialis anterior muscle of 10 men performing an isometric contraction at 15% of maximal force, previously used to validate source separation techniques [67]. Maximal contraction was defined as the mean force of three 5-s maximal contractions separated by 3 min of rest, with force sampled at 2048 Hz by load cells mounted on an isometric brace. Force feedback was provided to the participants by an oscilloscope. The signal from a monopolar 12×5 electrode array placed over the main muscle innervation zone was sampled at 2048 Hz having been band-pass filtered at 10–500 Hz. gCKC with an additional k-means source refinement step was implemented using the tensorflow machine learning package [16], [65]. As the label set was to be artificially corrupted it was important that the original be as noise free as possible, so additional post hoc steps were taken to maximize the likelihood that the timestamps were correct. Sources were manually cleaned by examining interspike intervals and the source-to-noise ratio of each activation. An additional step of validating decomposition accuracy was implemented by comparing the sources to those found using the DEMUSE source-separation software package [64], [65], with source cleaning completed by a different trained operator.

V. RESULTS

A. Feature-Dependent Label Noise

In experiment 1, which tested the effect of feature-dependent label noise by simulating noisy separation vectors, the LSM-trained network generated an embedding space with dense clusters for each class corresponding to the true labels. Surrounding each cluster is a large sparse periphery of false labels with no apparent structure, the expected result as these false samples shared fewer features. In contrast, the embedding spaces generated by training with CE, NL and AL gave a more uniform single cluster for each class (with some distant outliers for CE). The utility of these different embedding spaces for identifying false labels is particularly clear when LSM is compared to AL using a 2-D principal component space (Fig. 3).

Table I shows the cleaning results when selecting only samples from the highest-density region of a class embedding. The LSM-trained network generated an embedding space with utility for removing false labels even at the maximum tested value of 50% of total correct values, with a median post-cleaning false label retention of 2.3% of the total correct labels in the class. The number of true labels lost during the cleaning process fell as the precleaning percentage of false labels increased, but even at the highest-false label percentage tested, a median of 74.1% of the true values were still retained.

MDR also performed well at preserving some local semantic structure and so the embedding space had some utility for identifying noisy labels, although is was slightly outperformed

TABLE II
CLEANING RESULTS FOR DIFFERENT LEVELS OF CLASS-DEPENDENT LABEL NOISE

	Level of Added Label Noise							
	10%		20%		30%		40%	
Method	**True and False Labels after Cleaning (% of Original Set)**							
	True	False	True	False	True	False	True	False
CE	79.8	7.9	64.5	17.7	58.1	25.4	54.5	33.5
NL[40]	82.5	7.1	64.0	13.5	57.9	20.9	49.7	30.1
AL[42]	61.7	3.0	45.5	9.3	41.8	13.0	33.5	20.0
MDR[44]	63.1	0.8	48.1	1.6	43.9	1.8	36.4	2.2
LSM	**86.4**	**0.8**	**75.4**	**0.9**	**60.9**	**0.9**	**53.1**	**1.0**

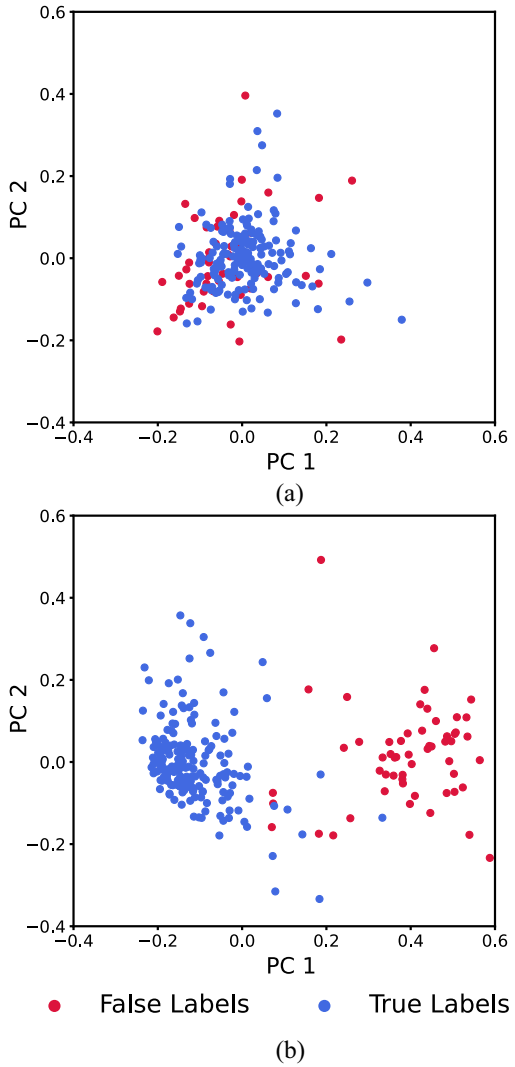

Fig. 3. Effect of two different DML losses on the embedding space for two units as shown by the first and second principle components for 40% corruption with feature-dependent label noise. (a) Effect of using the original random sampling method of angular loss, leading to all intraclass embeddings contracting down to a point. In (b) the same optimization was run again using a DML loss that preserves local structure, creating an embedding space in which the true labels cluster away from the false labels.

by LSM. In contrast, the networks trained with CE, AL and NL did not generate embedding spaces suitable for label cleaning. AL and MDR generally produced much more distributed clusters, meaning less samples in total, true or false, were selected by the density estimator.

B. Class-Dependent Label Noise

In experiment 2, when labels were randomly flipped to the MU class with the closes average MUAP shape, LSM again generated embedding spaces with clear separation between true and false separation. However, unlike in the first experiment, the false labels formed a second distinct cluster within the embedding space (Fig. 4). As the true label cluster always had more values, it was still clearly identified by the density estimator.

Table II gives the cleaning results for class-dependent label noise across the different tested methods. As in experiment 1, the LSM-trained model generated an embedding space that allowed the density estimator to identify almost all false labels. Even at a 40% transfer the median post-cleaning false label retention was only 1.0% of the total correct labels in the class. MDR also performed quite well, albeit with a slightly higher-false label burden than LSM.

As true labels were lost both to the initial transfer to other classes and to the cleaning phase, far fewer were retained in the post-cleaning dataset than in experiment 1 and would need to be recovered in the refitting stage. Once again, the CE, NL and AL trained models generated embedding spaces that were not useful for cleaning, and AL and MDR again generating a looser embedding with the consequence of less samples selected.

C. Rediscovering Unlabeled Activations

An important requirement if the cleaned label set is to be useful is that the cleaning process does not overly bias against true labels that are lost at this stage, making them difficult to recover. Lost labels tend to be more peripheral in the cluster, meaning their MUAP shapes are likely to be less similar to the MU class average, potentially due to superposition with a MUAP from a different class or due to a noise artefact. False negatives are also still used for training, but are labeled inappropriately, with a possibly detrimental effect of the model to generalize. To demonstrate that neither of these potential problems actually impacted training, after the cleaning stage of both experiments 1 and 2, the model was refitted with the LSM-cleaned label set. As an additional comparison, we also refitted using the same label set, but with the other methods compared in this study.

For both experiments the predicted activity after running the DeepDecomp pipeline was generally both sparse and clean, with MUAPs easily identifiable using all methods tested. These labels were compared to the original data using the RoA metric, with results given in Table III. LSM and MDR tended to outperform the other methods tested, most likely due to a better ability to generalize that has been noted in DML losses that preserve semantic information [44]. Most striking was the ability of the DeepDecomp pipeline to recover the original label sets in extreme degrees of label noise, as demonstrated visually in Fig. 5.

D. Impact of the k Hyperparameter

Two main trends emerged when the cleaning component of experiment 1 was repeated using different values of the k

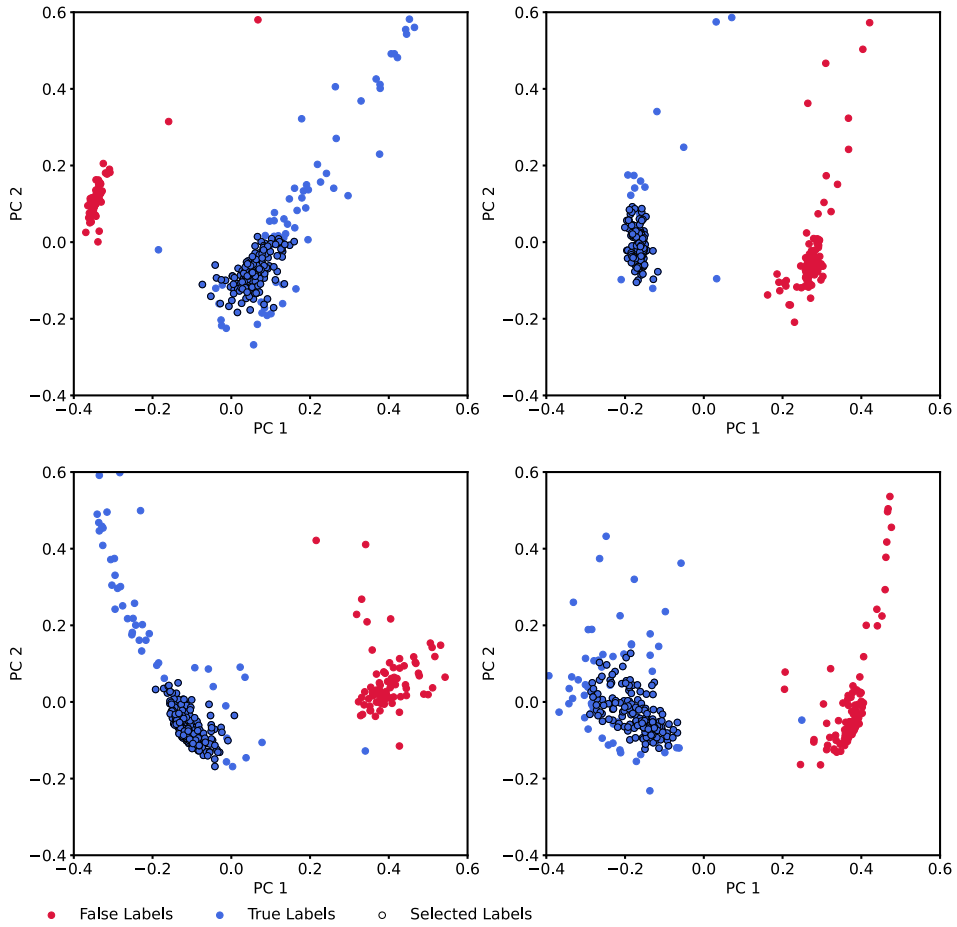

Fig. 4. Principal component plot of the embedded samples from four classes with added class-dependent label noise. The samples selected automatically for the refitting phase have been circled. The false labels have shared features as they come from the same class. This results in two tight clusters for both true and false labels, however, they are still clearly separable.

TABLE III
REFITTING RESULTS FOR DIFFERENT LEVELS OF LABEL NOISE

Label Noise Type	Method	Level of Added Label Noise 10%	20%	30%	40%	50%
		Rate of Agreement (%)				
Feature Dependent	CE	92.1	92.3	90.8	90.6	89.1
	NL[40]	94.1	94.3	91.0	92.3	92.3
	AL[42]	93.7	93.7	89.4	89.2	91.5
	MDR[44]	94.6	94.0	93.4	93.6	91.7
	LSM	**94.5**	**94.5**	**93.5**	**93.8**	**92.0**
Class Dependent	CE	95.3	92.8	91.0	89.3	
	NL[40]	97.7	95.4	91.5	89.7	
	AL[42]	95.3	93.4	91.2	90.2	
	MDR[44]	98.0	96.8	94.5	93.8	
	LSM	**98.2**	**96.8**	**95.7**	**94.0**	

hyperparameter in the LSM method, which controls the number of top-k negatives which are used to build the loss (Fig. 6). As k increases the intraclass embedding becomes increasingly dense, resulting in the density estimator selecting a greater number of samples. At the same time, the embedding begins to lose its intraclass variance, resulting in increasing numbers of false negatives being incorrectly assigned as true. This leads to an optimal value of k being 5 if the main objective is to maximize the likelihood of identifying false negatives. Its worth noting that a k of 1 is functionally similar to an easy-negative loss [39], although this is quite inefficient with respect to losing true positives to the cleaning process.

VI. CONCLUSION

In this study we demonstrate that a DML pipeline which preserves local semantic structure can embed high-dimensional neurophysiological signal into a low-dimensional space that allows accurate identification of incorrectly labeled activation events. Furthermore, we present LSM, a sampling-based augmentation to DML losses which preserves such semantic structure. By using artificially corrupted sEMG data, we show that this simple change can outperform other contemporary methods of preserving intraclass variance, making possible a practical pipeline for cleaning noisy source separated time series data. As an example, we created DeepDecomp, a pipeline which utilizes LSM-augmented DML to clean heavily corrupted sEMG decomposition data over two passes. Importantly the model was still able to perform even when the source of label corruption is class or even feature dependent, an important consideration in neurophysiological signal where mistakes often occur due correlated effects such as source superposition.

Although this study focused on source-separated HD-sEMG signal, it is important to emphasize the broader applicability

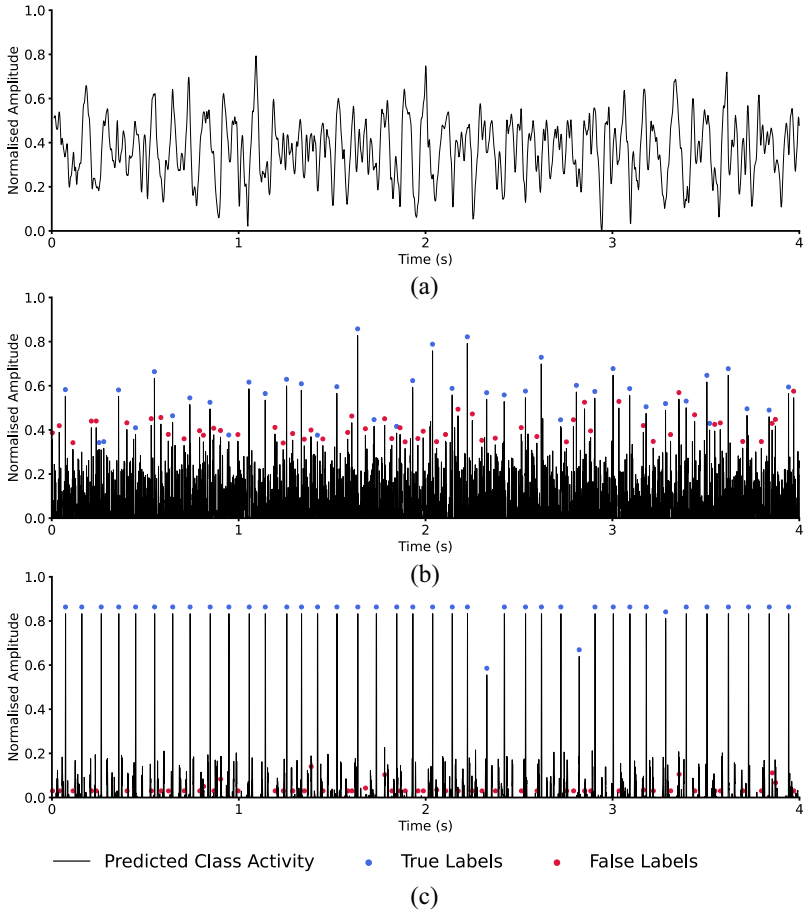

Fig. 5. Single channel of unprocessed HD-sEMG and the post-decomposition predicted activity of a single class before and after cleaning and refitting, with true and false labels. (a) demonstrates the degree of complex superposition inherent to sEMG signal as opposed to cleaner recordings such as those from intracortical sources. A linear separation filter based on an average of only 15 labels is applied to the signal to generate (b), which is consequently extremely noisy, simulating a poorly optimized filter. A number of false positives corresponding to 50% of the number of true class labels has been selected. After the cleaning and refitting phases of DeepDecomp the spiking motor neuron activity in (c) is clearly identifiable, whilst incorrect labels have been suppressed.

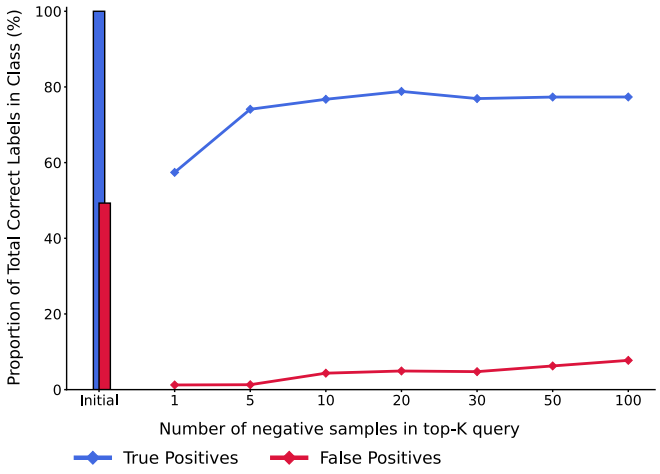

Fig. 6. Impact of the number of top-k negative samples used in LSM during the cleaning phase, after the data labels were corrupted with feature-dependent label noise. As k rises, more true positives are densely clustered in the embedding space, but the network begins to lose its intraclass variance, causing the density estimator to fail. A k of 5 was found to be optimal for maximizing the number of true positives whilst preventing class collapse.

of this approach to any imperfectly labeled neurophysiological time series data characterized by repeating events. Whilst the study focused specifically on action potentials, the proposed methodology could also be used for pattern recognition in bulk neurophysiological signal, for example by supplementing contemporary prosthetics and exoskeletons [68], [69]. Additionally, the labeling process need not be by a BSS algorithm. For example, a DeepDecomp-like pipeline could be applied to a dataset for which only a small component of the data has been manually labeled by an expert operator, recovering the rest of the labels accurately. In this way, the proposed approach can be viewed as a minimally supervised method for decomposing neurophysiological time series into individual cell activities.

REFERENCES

[1] R. Merletti and D. Farina, *Surface Electromyography: Physiology, Engineering, and Applications*. Hoboken, NJ, USA: Wiley, 2016.
[2] E. Stark and M. Abeles, "Predicting movement from multiunit activity," *J. Neurosci.*, vol. 27, no. 31, pp. 8387–8394, Aug. 2007.
[3] E. Drebitz, B. Schledde, A. K. Kreiter, and D. Wegener, "Optimizing the yield of multi-unit activity by including the entire spiking activity," *Front. Neurosci.*, vol. 13, p. 83, Feb. 2019.
[4] D. Farina and A. Holobar, "Characterization of human motor units from surface EMG decomposition," *Proc. IEEE*, vol. 104, no. 2, pp. 353–373, Feb. 2016.
[5] T. Kapelner et al., "Predicting wrist kinematics from motor unit discharge timings for the control of active prostheses," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, p. 47, Apr. 2019.

[6] W. Simon, "The real-time sorting of neuro-electric action potentials in multiple unit studies," *Electroencephalogr. Clin. Neurophysiol.*, vol. 18, no. 2, pp. 192–195, Feb. 1965.

[7] W. H. Calvin, "Some simple spike separation techniques for simultaneously recorded neurons," *Electroencephalogr. Clin. Neurophysiol.*, vol. 34, no. 1, pp. 94–96, Jan. 1973.

[8] H. G. Rey, C. Pedreira, and R. Q. Quiroga, "Past, present and future of spike sorting techniques," *Brain Res. Bull.*, vol. 119, pp. 106–117, Oct. 2015.

[9] J. J. Jun et al., "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, Nov. 2017.

[10] S. Muceli et al., "A thin-film multichannel electrode for muscle recording and stimulation in neuroprosthetics applications," *J. Neural Eng.*, vol. 16, no. 2, Feb. 2019, Art. no. 26035.

[11] M. Pachitariu, N. Steinmetz, S. Kadir, M. Carandini, and K. D. Harris, "Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels," bioRxiv Preprint, Jun. 2016.

[12] A. D. Vecchio and D. Farina, "Interfacing the neural output of the spinal cord: Robust and reliable longitudinal identification of motor neurons in humans," *J. Neural Eng.*, vol. 17, no. 1, Oct. 2019, Art. no. 16003.

[13] J. E. Downey, N. Schwed, S. M. Chase, A. B. Schwartz, and J. L. Collinger, "Intracortical recording stability in human brain–computer interface users," *J. Neural Eng.*, vol. 15, no. 4, 2018, Art. no. 46016.

[14] C. Chen, W. Guo, C. Ma, Y. Yang, Z. Wang, and C. Lin, "sEMG-based continuous estimation of finger kinematics via large-scale temporal convolutional network," *Appl. Sci.*, vol. 11, no. 10, p. 4678, 2021.

[15] C. Rossant et al., "Spike sorting for large, dense electrode arrays," *Nat. Neurosci.*, vol. 19, no. 4, pp. 634–641, 2016.

[16] F. Negro, S. Muceli, A. M. Castronovo, A. Holobar, and D. Farina, "Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation," *J. Neural Eng.*, vol. 13, no. 2, Feb. 2016, Art. no. 26027.

[17] S. Pancholi, A. Giri, A. Jain, L. Kumar, and S. Roy, "Source aware deep learning framework for hand kinematic reconstruction using EEG signal," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4094–4106, Jul. 2023.

[18] H. Lorach et al., "Walking naturally after spinal cord injury using a brain–spine interface," *Nature*, vol. 126, pp. 126–133, May 2023.

[19] J.-H. Jeong, J.-H. Cho, B.-H. Lee, and S.-W. Lee, "Real-time deep neurolinguistic learning enhances noninvasive neural language decoding for brain–machine interaction," *IEEE Trans. Cybern.*, early access, Oct. 17, 2022, doi: 10.1109/TCYB.2022.3211694.

[20] R. I. Kumar, M. M. Mallette, S. S. Cheung, D. W. Stashuk, and D. A. Gabriel, "A method for editing motor unit potential trains obtained by decomposition of surface electromyographic signals," *J. Electromyogr. Kinesiol.*, vol. 50, Feb. 2020, Art. no. 102383.

[21] K. C. McGill, Z. C. Lateva, and H. R. Marateb, "EMGLAB: An interactive EMG decomposition program," *J. Neurosci. Methods*, vol. 149, no. 2, pp. 121–133, Dec. 2005.

[22] F. Hug et al., "Analysis of motor unit spike trains estimated from high-density surface electromyography is highly reliable across operators," *J. Electromyogr. Kinesiol.*, vol. 58, Jun. 2021, Art. no. 102548.

[23] D. Carlson and L. Carin, "Continuing progress of spike sorting in the era of big data," *Current Opinion Neurobiol.*, vol. 55, pp. 90–96, Apr. 2019.

[24] P. Yger et al., "A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo," *eLife*, vol. 7, Mar. 2018, Art. no. e34518.

[25] A. K. Clarke et al., "Deep learning for robust decomposition of high-density surface EMG signals," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 526–534, Feb. 2021.

[26] Y. Wen, S. Avrillon, J. C. Hernandez-Pavon, S. J. Kim, F. Hug, and J. L. Pons, "A convolutional neural network to identify motor units from high-density surface electromyography signals in real time," *J. Neural Eng.*, vol. 18, no. 5, 2021, Art. no. 56003.

[27] Y. Wen, S. J. Kim, S. Avrillon, J. T. Levine, F. Hug, and J. L. Pons, "A deep CNN framework for neural drive estimation from HD-EMG across contraction intensities and joint angles," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2950–2959, 2022.

[28] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artif. Intell. Rev.*, vol. 33, no. 4, pp. 275–306, 2010.

[29] H. Wang et al., "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5265–5274.

[30] M. Kaya and H. Ş. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, p. 1066, Aug. 2019.

[31] X. Yang, P. Zhou, and M. Wang, "Person reidentification via structural deep metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 2987–2998, Oct. 2019.

[32] S. Liao and L. Shao, "Graph sampling based deep metric learning for generalizable person re-identification," 2021, *arXiv:2104.01546*.

[33] A. Zhong et al., "Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101993.

[34] J. V. Sundgaard et al., "Deep metric learning for otitis media classification," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102034.

[35] Z. Chai, L. Luo, H. Lin, H. Chen, A. Han, and P.-A. Heng, "Deep semi-supervised metric learning with dual alignment for cervical cancer cell detection," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, 2022, pp. 1–5.

[36] P. Pati, A. Foncubierta-Rodríguez, O. Goksel, and M. Gabrani, "Reducing annotation effort in digital pathology: A co-representation learning framework for classification tasks," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101859.

[37] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*.

[38] Y. Suh, B. Han, W. Kim, and K. M. Lee, "Stochastic class-based hard example mining for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7244–7252.

[39] H. Xuan, A. Stylianou, and R. Pless, "Improved embeddings with easy positive triplet mining," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 2463–2471.

[40] K. Sohn, "Improved deep metric learning with multi-class N-pair loss objective," in *Advances in Neural Information Processing Systems*, vol. 29, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran Assoc., Inc., 2016.

[41] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1320–1329.

[42] J. Wang, F. Zhou, S. Wen, X. Liu, and Y. Lin, "Deep metric learning with angular loss," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2612–2620.

[43] X. Li, L. Yu, C.-W. Fu, M. Fang, and P.-A. Heng, "Revisiting metric learning for few-shot image classification," *Neurocomputing*, vol. 406, pp. 49–58, Sep. 2020.

[44] Z. Fu, Z. Mao, C. Yan, A.-A. Liu, H. Xie, and Y. Zhang, "Self-supervised synthesis ranking for deep metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4736–4750, Jul. 2022.

[45] H. Coskun, D. J. Tan, S. Conjeti, N. Navab, and F. Tombari, "Human motion analysis with deep metric learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 693–710.

[46] Z. Che, X. He, K. Xu, and Y. Liu, "DECADE: A deep metric learning model for multivariate time series," in *Proc. KDD Workshop Min. Learn. Time Series*, 2017, pp. 1–9.

[47] Z. Chen et al., "Deep multiple metric learning for time series classification," *IEEE Access*, vol. 9, pp. 17829–17842, 2021.

[48] D. Cheng, Z. Li, Y. Gong, and D. Zhang, "Fusion of multiple person re-id methods with model and data-aware abilities," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 561–571, Feb. 2020.

[49] M. Ye and P. C. Yuen, "PurifyNet: A robust person re-identification model with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2655–2666, 2020.

[50] Y. Yuan, W. Chen, Y. Yang, and Z. Wang, "In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1454–1463.

[51] D. Fu et al., "Large-scale pre-training for person re-identification with noisy labels," 2022, *arXiv:2203.16533*.

[52] F. Xu, B. Ma, H. Chang, and S. Shan, "PRDP: Person reidentification with dirty and poor data," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 11014–11026, Oct. 2022.

[53] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 499–515.

[54] O. Rippel, M. Paluri, P. Dollar, and L. Bourdev, "Metric learning with adaptive density discrimination," 2015, *arXiv:1511.05939*.

[55] Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, "No fuss distance metric learning using proxies," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 360–368.

[56] S. Kim, D. Kim, M. Cho, and S. Kwak, "Proxy anchor loss for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3238–3247.

[57] Q. Qian et al., "SoftTriple loss: Deep metric learning without triplet sampling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6449–6457.
[58] X. Lin, Y. Duan, Q. Dong, J. Lu, and J. Zhou, "Deep variational metric learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 714–729.
[59] Y. Liu, X. Gao, J. Han, and L. Shao, "A discriminative cross-aligned variational autoencoder for zero-shot learning," *IEEE Trans. Cybern.*, vol. 53, no. 6, pp. 3794–3805, Jun. 2023.
[60] Y. Kim and W. Park, "Multi-level distance regularization for deep metric learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 1827–1835.
[61] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao, and Y. Gao, "Kullback–Leibler divergence metric learning," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2047–2058, Apr. 2022.
[62] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[63] N. Shazeer et al., "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," 2017, *arXiv:1701.06538*.
[64] A. Holobar and D. Zazula, "Gradient convolution kernel compensation applied to surface electromyograms," in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, 2007, pp. 617–624.
[65] A. Holobar and D. Zazula, "Multichannel blind source separation using convolution kernel compensation," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4487–4496, Sep. 2007.
[66] G. Algan and İ. Ulusoy, "Label noise types and their effects on deep learning," 2020, *arXiv:2003.10471*.
[67] A. Holobar, M. A. Minetto, A. Botter, F. Negro, and D. Farina, "Experimental analysis of accuracy in the identification of motor unit spike trains from high-density surface EMG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, pp. 221–229, 2010.
[68] Z. Li, Q. Li, P. Huang, H. Xia, and G. Li, "Human-in-the-loop adaptive control of a soft exo-suit with actuator dynamics and ankle impedance adaptation," *IEEE Trans. Cybern.*, early access, Feb. 14, 2023, doi: 10.1109/TCYB.2023.3240231.
[69] M. A. Díaz et al., "Human-in-the-loop optimization of wearable robotic devices to improve human–robot interaction: A systematic review," *IEEE Trans. Cybern.*, early access, Dec. 20, 2022, doi: 10.1109/TCYB.2022.3224895.

Alexander Kenneth Clarke (Student Member, IEEE) received the joint degrees in physics in 2012 and in medicine and surgery in 2016 from the University of Warwick, Coventry, U.K. He is currently pursuing the Ph.D. degree with the Department of Bioengineering, Imperial College London, London, U.K., within the EPSRC Centre for Doctoral Training in Neurotechnology.

Prior to this he was an Academic Foundation Doctor with the University of Newcastle, Newcastle upon Tyne, U.K. His main research interest lies in the intersection of classical blind source separation, machine learning for time series, and biomedical signal processing.

Dario Farina (Fellow, IEEE) received the first Ph.D. degree in automatic control and computer science from Ecole Centrale de Nantes, Nantes, France, in 2001, and the second Ph.D. degree in electronics and communications engineering from Politecnico di Torino, Turin, Italy, in 2002, and the Honorary Doctorate degree in medicine from Aalborg University, Aalborg, Denmark, in 2018.

He is currently a Full Professor and a Chair in Neurorehabilitation Engineering with the Department of Bioengineering, Imperial College London, London, U.K. He has previously been a Full Professor with Aalborg University and with the University Medical Center, Georg-August University, Göttingen, Germany, where he founded and directed the Department of Neurorehabilitation Systems, acting as the Chair in Neuroinformatics of the Bernstein Focus Neurotechnology Göttingen. His research focuses on biomedical signal processing, neurorehabilitation technology, and neural control of movement. Within these areas, he has (co)authored more than 500 papers in peer-reviewed journals and has been the Editor of the books *Introduction to Neural Engineering for Motor Rehabilitation* (Wiley-IEEE Press, 2013), *Surface Electromyography: Physiology, Engineering and Applications* (Wiley-IEEE Press, 2016), and *Bionic Limb Reconstruction* (Springer, 2021).

Prof. Farina has been the recipient of the IEEE Engineering in Medicine and Biology Society (EMBS) Early Career Achievement Award, the IEEE EMBS Technical Achievement Award, and the Royal Society Wolfson Research Merit Award. He has been the President of the International Society of Electrophysiology and Kinesiology (ISEK) and is currently the Editor-in-Chief of the Official Journal of this Society, the *Journal of Electromyography and Kinesiology*. He is also currently an Editor for *Science Advances*, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, IEEE TRANSACTIONS ON MEDICAL ROBOTICS AND BIONICS, and IEEE REVIEWS IN BIOMEDICAL ENGINEERING. He has been elected as a Fellow of AIMBE, ISEK, and EAMBES.