Employing Neighborhood Component Analysis as a Feature Selection Method in the Pattern Recognition Approach for Surface Electromyography Signal Classification

Daniyar Zhakyp and Muhammad Tahir Akhtar
Department of Electrical and Computer Engineering, School of Engineering & Digital Sciences,
Nazarbayev University, Kabanbay Batyr Ave. 53, Nur-Sultan, Kazakhstan.
Emails: {daniyar.zhakyp, muhammad.akhtar}@nu.edu.kz, akhtar@ieee.org

***Abstract*—Many research works regarding the design of a pattern recognition (PR) approach for surface electromyography (sEMG) signal analysis have attempted to find the most relevant features leading to the highest classification performance. While some classical works have outlined the general time-domain (TD) features related to sEMG signals, the research area still lacks data-driven feature selection (FS) methods that would generalize the effect of features on the classification performance. This paper proposes a novel approach of using a Neighborhood Component Analysis (NCA) algorithm as the FS step in the PR scheme. The NCA-based feature selection (NCA-FS) has resulted in an increase in the classification accuracy and F1 score of four machine learning (ML) models while reducing the total complexity of the PR scheme for two subjects in the NinaPro Database 2 (DB2). Moreover, the proposed method has elicited the most powerful TD features, positively affecting the overall performance, compared to some classical and state-of-the-art works regarding sEMG signal recognition.**

***Index Terms*—Surface Electromyography, Feature Selection, Pattern Recognition, Neighborhood Component Analysis, Signal Classification.**

I. Introduction

Surface electromyography (sEMG) is a data acquisition method for electric muscle activity, with applications in developing myoelectric prostheses for academic research and finally commercial use. Whilst most of the commercially available prosthetic devices adopted the simple control scheme based on sEMG amplitude, research has widely focused on incorporating machine learning (ML) and deep learning methods to classify the sEMG signals into separate motions of a prosthesis. This novel control scheme is called the pattern recognition (PR) approach. As the application of the PR approach in the myoelectric prosthesis design would substantially increase the dexterity of the device and eliminate the need for strict electrode arrangement, there is a need to explore the algorithms within the PR method, yielding the highest reliability and efficiency.

A flowchart of the PR approach is shown in Fig. 1. The first stage is signal preprocessing, where sEMG signals are filtered to eliminate various sources of noise, including 50- or 60-Hz power line noise [1]. In the second stage, due to the non-stationarity of the sEMG time series, the useful information is extracted in windows in the form of mathematical features [1]. The third stage is optional and mainly comprises the feature selection (FS) or reduction processes, where the features with the highest "descriptive" power are selected for further classification [1]. Lastly, various statistical ML models are trained on the resultant feature matrix to predict the motion class of the test data [1].

Many researchers have focused on finding universal features that best describe the stochastic nature of sEMG. The pioneering work [2] proposes the classical feature set, consisting of five time-domain (TD) features, namely the Mean Absolute Value (MAV), Mean Absolute Value Slope (MAVS), Zero Crossing (ZC), Slope Sign Change (SSC), and Waveform Length (WL), upon which other similar research quests have been aroused. A predefined set of features has also

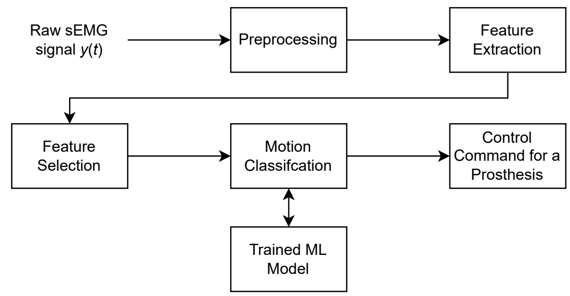

Fig. 1. Flowchart of the PR approach in sEMG signal analysis.

been deployed in the sEMG classification task in [3] and [4]. The further studies analyzed the usefulness of a feature or a predefined combination of features based on their effect on the classification error [5] or by visual inspection of their scatter plots [6]; however, very few studies deployed any data-driven FS algorithms that would ease the given problem. The proper FS algorithm would highlight the most prominent features of the given data while eliminating the unnecessary ones. This scenario overall may help in the classification task and overall reduce the complexity of the entire PR model, which could be beneficial for hardware implementation later.

The existing state-of-the-art works regarding FS approaches select features based on the ratio of the Euclidean distance (ED) between the feature means of two respective motions and the standard deviation (STD) value of all windows within a feature [7]. The features' discriminative power based on the Bhattacharyya distance as the filter FS method, and the Sequential Forward Search (SFS) combined with the Linear Discriminant Analysis (LDA) classifier as the wrapper FS method have been evaluated in [8]. The Particle Swarm Optimization (PSO) has been introduced to select appropriate features for sEMG classification [9]. Pearson's Correlation Coefficient (PCC) has also been used for removing irrelevant features based on statistical linear correlation between a pair of features [10]. Additionally, the Principal Component Analysis (PCA) and LDA have been employed as the dimensionality reduction methods, neglecting the entirety of a feature extracted from multiple channels but rather seeing the feature set as a matrix of samples [11].

This paper proposes to employ the Neighborhood Component Analysis (NCA) algorithm as the sEMG FS method [12]. Using NCA, the weights of each individual feature in the extracted set are calculated. Out of 44 extracted TD features appearing in the literature regarding the sEMG analysis, different feature subsets are obtained by setting distinct thresholds for feature weights in NCA. The obtained feature vector with no FS and with NCA-based feature selection (NCA-FS) using several thresholds has been trained on four ML classifiers to observe the effect on classification performance. Furthermore, the results have been compared with the performance of

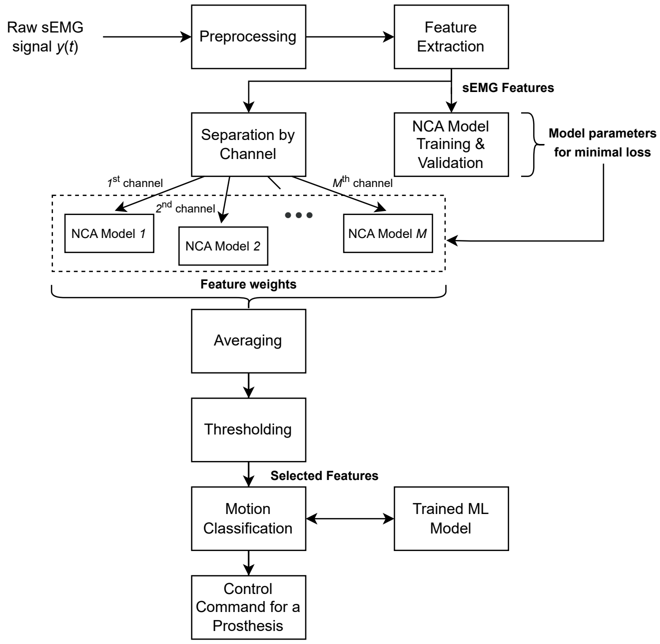

Fig. 2. Flowchart of the proposed method.

the benchmark and classical feature sets existing in the literature. The aim of this work is to evaluate the contribution of the incorporated NCA algorithm as the FS method to the change in prediction performance and complexity of the PR approach. The rest of the paper is organized as follows. Section 2 gives details of materials and methods. The results and discussion are provided in Section 3. Finally, the concluding remarks are given in Section 4.

II. Materials and Methods

A flowchart of the proposed method is shown in Fig. 2 and the detailed description of the various stages within the proposed method is as follows.

A. Dataset and Preprocessing

The Non-Invasive Adaptive Prosthetics (NinaPro) Database 2 (DB2) has been considered for experiments. The NinaPro DB2 contains 40 healthy subjects of both genders and of different ages and constitutions [13]. Each subject performed three different exercises that involve basic finger and wrist movements, grasping and functional actions, and force patterns, respectively [13]. All three exercises constitute 50 distinct motion classes, including rest, and are performed in six repetitions (sessions) [13]. Subjects repeat hand motions that appear on the laptop screen for 5 seconds for each gesture, followed by 3 seconds of rest [13]. The sEMG signals have been recorded using 12-channel Delsys Trigno wireless electrodes at a sampling rate of 2 kHz [13]. For experiments carried out in this study, only Subject 1 (S1) and Subject 5 (S5) have been considered. Both subjects are right-handed men in their twenties.

The preprocessing step first comprises the band-pass filtering with the 8^{th}-order Butterworth digital filter. The cut-off frequencies of the Butterworth band-pass filter are set to 20 and 500 Hz. The 2^{nd}-order Notch filter has been implemented to eliminate the 50 Hz power line noise. After the filtering stage, the signals have been downsampled by a factor of 2, from 2 to 1 kHz.

B. Feature Extraction

The feature extraction step implies the representation of the sEMG time series in the form of valuable and descriptive domain features.

TABLE I
Forty Four TD Features Within the Original sEMG Feature Set

Index	Feature
1	Root Mean Square (RMS) [14]
2	Mean Absolute Value (MAV) [2]
3	Average Amplitude Change (AAC) [6]
4	Average Energy (AE)
5	Absolute Value of the Summation of the $\exp^{\text{th}}$ Root (ASM) [15]
6	Absolute Value of the Summation of Square Root (ASS) [15]
7	Cardinality (CARD) [16]
8	Coefficient of Variation (COV) [17]
9	Difference Absolute Mean Value (DAMV) [14]
10	Difference Absolute Standard Deviation Value (DASDV) [14]
11	Difference Variance Value (DVARV) [18]
12	Enhanced Mean Absolute Value (EMAV) [19]
13	Enhanced Waveform Length (EWL) [19]
14	Integrated EMG (IEMG) [6]
15	Interquartile Range (IQR) [20]
16	Kurtosis (KURT) [10]
17	Log of Coefficient of Variation (LCOV) [21]
18	Log Detector (LD) [22]
19	Log of Difference Absolute Mean Value (LDAMV) [18]
20	Log of Difference Absolute Standard Deviation Value (LDASDV) [18]
21	Maximum Fractal Length (MFL) [23]
22	Mean Absolute Deviation (MAD) [20]
23	Mean Absolute Value Slope (MAVS) [2]
24	Mean Value of the Square Root (MSR) [15]
25	Modified Mean Absolute Value 1 (MMAV1) [6]
26	Modified Mean Absolute Value 2 (MMAV2) [6]
27	Myopulse Percentage Rate (MYOP) [6]
28	New Zero Crossing (ZC) [24]
29	Simple Square Integral (SSI) [6]
30	Skewness (SKEW) [10]
31	Slop Sign Change (SSC) [2]
32	Standard Deviation (STD)
33	Teager-Kaiser Energy Operator (TKEO) [21]
34	Temporal Moment (TM) [6]
35	Variance (VAR)
36	Variance of EMG (VAREMG) [22]
37	v-Order (VO) [6]
38	Waveform Length (WL) [2]
39	Willison Amplitude (WAMP) [22]
40	Zero Crossing (ZC) [2]
41	Histogram (HIST) [25]
42	Coefficients of the 4^{th} order Autoregressive Model (AR4) [26]
43	Coefficients of the 5^{th} order Autoregressive Model (AR5) [26]
44	Coefficients of the 6^{th} order Autoregressive Model (AR6) [26]

Those features are extracted in overlapping or non-overlapping windows, during the duration of which the sEMG signal is assumed to be stationary and ergodic. A rectangular window of length of 256 milliseconds (ms) is used. The window increment size of 128 ms for the training set and 64 ms for the test set is used. Table I shows TD features that constitute the original sEMG feature set with no FS. It has been decided to extract 44 different TD features, which have been used in various sEMG analysis research works. The bin size for the Histogram (HIST) feature is chosen to be 5. The orders of the Temporal Moment (TM) and v-Order (VO) features are set to 3 and 4, respectively. For the Cardinality (CARD), Myopulse Percentage Rate (MYOP), Slope Sign Change (SSC), Zero Crossing (ZC), and Willison Amplitude (WAMP) features, the threshold value ϵ is computed as [28]:

$$\epsilon = R \times \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{NM_j})^2}, \tag{1}$$

where y_{NM_j} are signal samples at rest (no motion) and R is the coefficient ranging from 0 to 4, which has been empirically set to 0.5 for results reported in this study. It is worth to mention that transitions between motion classes may contain imprecise signal information that can be removed to improve the classification accuracy [3]. Therefore, eight samples before and after the transition have been removed from each feature in the resulting feature vector. Lastly, all the features have been normalized across their mean and STD values.

C. Proposed NCA-Based Feature Selection (NCA-FS)

The NCA algorithm is an efficient statistical analysis tool that addresses two major limitations of the k-Nearest Neighbor (kNN)

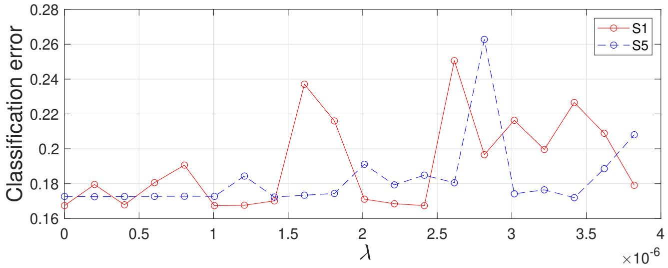

Fig. 3. Classification error of NCA for 20 different values of the regularization parameter λ for S1 and S5.

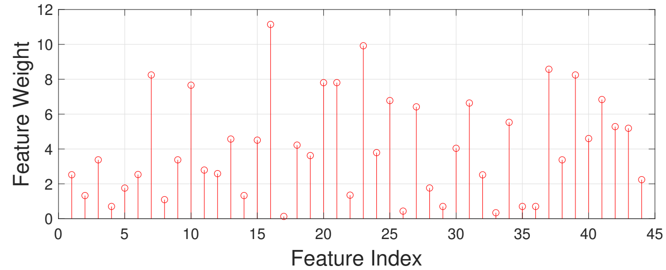

Fig. 4. Feature weights of 44 TD features computed by NCA for S1.

algorithm: high computational complexity and uncertainty in choosing the "proper" distance metric for "nearest neighbors" [12]. The kNN's high complexity is explained by its necessity to store and search through all the sample space to identify the neighbor(s) of a test sample and label it accordingly. By contrast, the NCA algorithm uses the stochastic neighbor selection (SNS) method to more efficiently identify the neighbor(s) of the input sample. However, since information about the distribution of the test data is unavailable, the NCA algorithm attempts to minimize the leave-one-out (LOO) error on the training data [12]. Consider a classification problem with a training set:

$$D = \{(\boldsymbol{x}_i, c_i), i = 1, 2, \cdots, N\}, \tag{2}$$

where N is the number of observations, $\boldsymbol{x}_i \in \mathbb{R}^p$ are the feature vectors, $c_i \in \{0, 1, \cdots, J-1\}$ are the class labels with J denoting the number of classes [12]. With some probability, NCA picks a reference sample $\boldsymbol{x}_j$ from D to label another training sample $\boldsymbol{x}_i$ based on the class label of the reference. A choice for a reference point depends on the distance between two samples based on the weight vector $\boldsymbol{w}$ and computed as [27]:

$$d_{\boldsymbol{w}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{r=1}^{p} w_r^2 |x_{ir} - x_{jr}|, \tag{3}$$

where w_r is the weight of the r^{th} feature. The probability of $\boldsymbol{x}_j$ being chosen as the reference point for $\boldsymbol{x}_i$ is computed as:

$$p_{ij} = \frac{K(d_{\boldsymbol{w}}(\boldsymbol{x}_i, \boldsymbol{x}_j))}{\sum_{k \neq i}^{N} K(d_{\boldsymbol{w}}(\boldsymbol{x}_i, \boldsymbol{x}_k))}, \; p_{ii} = 0, \tag{4}$$

where $K(z) = \exp(-\frac{z}{\sigma})$ is the kernel function and σ denotes the kernel width parameter, which controls the likelihood of any given point being chosen as the reference point [27]. Thus, the objective of NCA is to maximize the average number of observations classified correctly under the SNS rule, and it is computed as [27]:

$$F(\boldsymbol{w}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} c_{ij} p_{ij} = \frac{1}{N} \sum_{i=1}^{N} p_i, \tag{5}$$

where $c_{ij} = 1$ if $c_j = c_i$ and is 0 otherwise.

For the FS operation, the regularized version of the objective function is computed as [27]:

$$F(\boldsymbol{w}) = \frac{1}{N} \sum_{i=1}^{N} p_i - \lambda \sum_{r=1}^{p} w_r^2, \tag{6}$$

where λ is the regularization parameter, which can drive many weights in $\boldsymbol{w}$ to zero and be adjusted through cross-validation [27]. The NCA's task is to find the optimal weights $\boldsymbol{w}$ that maximize $F(\boldsymbol{w})$; however, after setting the kernel width $\sigma = 1$, the maximization of $F(\boldsymbol{w})$ is equivalent to the following minimization problem for given λ [27]:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left(\frac{1}{N} \sum_{i}^{N} \sum_{j=1, j \neq i}^{N} p_{ij} l(c_i, c_j) + \lambda \sum_{r=1}^{p} w_r^2\right), \tag{7}$$

where $\hat{\boldsymbol{w}}$ is the optimal feature weight vector and $l(c_i, c_j)$ is a classification error, which is 1 when $c_i \neq c_j$ and is 0 otherwise. Thus, the found feature weight vector is the one that minimizes the LOO classification error of NCA. By keeping the features with only non-zero weights or with weights greater than some certain threshold, the distance metric can be made low-rank and solve the aforementioned computational complexity problem of kNN.

The FS phase is associated with identifying and deploying the most powerful features with the goal of improving classification performance and reducing PR model complexity. The NCA model has been trained and tuned for subjects S1 and S5 separately. Fig. 3 shows the NCA classification error for different values of the regularization parameter λ for S1 and S5. The NCA model's error optimization has been performed via the leave-one-repetition-out (LORO) strategy. The LORO scheme implies training the model on five repetitions and using the sixth repetition for the test. By repeating this process five more times, the total accuracy is then averaged across all the combinations.

Fig. 4 shows the feature weights for all 44 extracted TD features for S1. A similar trend is observed for S5 and is omitted for the sake of space. After identifying the λ parameter corresponding to the lowest error, this value is used to train twelve separate NCA models based on data from 12 distinct sEMG channels. As each individual feature takes up 12 columns (except AR coefficients, whose size depends on the order of the AR model) of the feature matrix, the standard FS technique may assign little weight to some of the channels. The proposed methodology suggests developing the NCA-FS scheme on all 12 sEMG channels and then averaging their weights. This method allows the entire feature to be preserved without missing channels as columns. From Fig. 4, it can be observed that the 16^{th} feature has the highest weight across all 12 channels on average.

Finally, several thresholds have been applied to the feature weight vector to come up with different feature subsets. These feature subsets are classified using ML models and compared in terms of their performance and execution time.

III. Results and Discussion

Four different statistical ML models have been used for classification: Support Vector Machine (SVM), kNN, LDA, and Ensemble of Bagged Decision Trees (EBDT). The SVM model is used with the Radial Basis Function (RBF) kernel and the *'auto'* kernel scale parameter in MATLAB. For kNN, the number of neighbors equals 3, as it has been empirically deduced that this value results in good classification performance. The EBDT model has the following parameters: *minLeafSize* = 10, *maxNumSplits* = 300, *Split criterion* = 'deviance'. The data has been classified using the 'one-vs-all' (OvA) strategy and based on the previously mentioned LORO scheme.

The performance for different feature sets resulted from NCA-FS has been assessed using the accuracy metric, being defined as [29]:

$$\text{Accuracy} = \frac{1}{J} \sum_{i=0}^{J-1} \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}, \tag{8}$$

TABLE II
THE AVERAGE CLASSIFICATION ACCURACY FOR FOUR ML CLASSIFIERS WITH NO FS AND NCA-FS OF SIX DIFFERENT THRESHOLDS.

	no FS	**NCA-FS**					
	t = 0 (p = 44.0)	t = 2.0 (p = 29.5)	t = 3.0 (p = 23.0)	t = 4.0 (p = 17.5)	t = 5.0 (p = 13.5)	t = 6.0 (p = 10.0)	t = 7.0 (p = 6.0)
SVM	0.7965	0.7966	**0.8009**	0.7994	0.7970	0.7918	0.7875
kNN	0.7597	0.7595	**0.7616**	0.7556	0.7505	0.7579	0.7542
LDA	0.6488	0.6958	**0.7459**	0.7448	0.7433	0.7361	0.7183
EBDT	0.7771	0.7803	0.7801	**0.7813**	0.7803	0.7787	0.7757

TABLE III
THE AVERAGE F1 SCORE FOR FOUR ML CLASSIFIERS WITH NO FS AND NCA-FS OF SIX DIFFERENT THRESHOLDS.

	no FS	**NCA-FS**					
	t = 0 (p = 44.0)	t = 2.0 (p = 29.5)	t = 3.0 (p = 23.0)	t = 4.0 (p = 17.5)	t = 5.0 (p = 13.5)	t = 6.0 (p = 10.0)	t = 7.0 (p = 6.0)
SVM	0.6882	0.6945	**0.6971**	0.6932	0.6890	0.6831	0.6674
kNN	0.6426	**0.6473**	0.6468	0.6392	0.6283	0.6359	0.6190
LDA	0.5782	0.5826	0.6330	**0.6387**	0.6267	0.6147	0.5637
EBDT	0.6393	0.6433	0.6457	0.6499	**0.6499**	0.6439	0.6460

where TP_i, FP_i, TN_i, and FN_i are True Positives, False Negatives, True Negatives, and False Negatives, respectively, of the class i. Another classification metric is F1 score, being computed as [29]:

$$\mathrm{F1} = \frac{1}{J}\sum_{i=0}^{J-1}\frac{TP_i}{TP_i + \frac{1}{2}(FP_i + FN_i)}. \tag{9}$$

The execution time is computed for the entire PR model, beginning with the preprocessing step and finishing with the motion prediction. The experiments have been performed using a desktop PC that has an Intel Core i9 vPro 9th Gen processing unit with 8 cores. Feature extraction and classification have been carried out via the *Parallel Computing Toolbox* in MATLAB R2022a.

A. Effect of NCA-FS on Classification Performance

Table II and Table III show the average accuracy and F1 scores of four classification methods based on seven different threshold values, where a threshold t of 0 means there is no FS used and an average number of features p (complexity) is equal to 44.0. Table IV shows the average execution time of the PR model for different classifiers and thresholds. Overall, all four classifiers demonstrated the peaking phenomenon after applying NCA-FS and eliminating redundant features. While reducing the PR complexity and achieving a faster execution time by decreasing the feature set to 23.0 features, the SVM and kNN models achieved even slightly higher accuracy and F1 score than by no FS and/or NCA-FS with milder thresholds. For the LDA model, the NCA-FS approach at the same threshold value shows an increase in accuracy of almost 10% and the second best result for the F1 score. The EBDT classifier achieves its peak accuracy and F1 score using NCA-FS when the average number of features p is 17.5 and 13.5, respectively. Apparently, once applying higher threshold values and further decreasing the complexity, all four models' classification performances eventually drop. The SVM model using the NCA-FS approach has shown superior performance over other classifiers and is thus chosen for comparison with the existing benchmark feature sets.

B. Performance Comparison with Benchmark Feature Sets

In this section, the best classification performance of a feature set derived from NCA-FS is compared to the best performance of the benchmark sEMG feature sets existing in the literature. Table V shows the SVM model's average accuracy, F1 score, and execution time for different feature set complexities of corresponding works.

TABLE IV
THE AVERAGE PR EXECUTION TIME IN SECONDS (S) FOR FOUR ML CLASSIFIERS WITH NO FS AND NCA-FS OF SIX DIFFERENT THRESHOLDS.

	no FS	**NCA-FS**					
	t = 0 (p = 44.0)	t = 2.0 (p = 29.5)	t = 3.0 (p = 23.0)	t = 4.0 (p = 17.5)	t = 5.0 (p = 13.5)	t = 6.0 (p = 10.0)	t = 7.0 (p = 6.0)
SVM	830.48	685.82	**635.65**	565.11	493.39	414.28	372.67
kNN	2459.02	**2001.84**	1641.95	1342.22	1206.27	718.56	455.22
LDA	760.86	667.80	611.07	**519.86**	487.92	354.84	312.77
EBDT	592.66	550.50	513.38	454.37	**421.49**	357.75	334.87

TABLE V
COMPARISON OF THE RESULTS OF THE BENCHMARK FEATURE SETS WITH THOSE OBTAINED USING NCA-FS FOR THE SVM CLASSIFIER

	Accuracy	F1	Ex. time (s)	Complexity, p
Hudgins et al. set	0.7684	0.6378	**280.44**	**5.0**
Atzori et al. set▼	0.7827	0.6678	367.61	7.0
Phinyomark et al. set	0.7680	0.6427	353.33	15.0
Souza et al. set	0.7804	0.6671	368.18	7.0
Proposed method (a)	**0.8009**	**0.6971**	635.65	23.0
Proposed method (b)	0.7875	0.6674	372.67	6.0

(a) - NCA-FS of t = 3.0, (b) - NCA-FS of t = 6.0, ▼ - a modified version

There is a classical Hudgins et al. [2] feature set consisting of five TD features (p = 5.0); a modified version of the Atzori et al. [25] feature set, which consists of the Hudgins et al. set and two more TD features (without a marginal Discrete Wavelet Transform (mDWT), which is a time-frequency domain (TFD) feature); a Phinyomark et al. [6] set of fifteen TD features selected based on scatter plots of different combinations of channels of the respective features; and a Souza et al. [10] set selected based on PCC values of different combinations of features. The feature set derived from NCA-FS, with a complexity of 23.0, shows higher classification performance in terms of accuracy and F1 score than other feature sets while having slower execution. However, after decreasing the complexity to 6.0 features using NCA-FS, the classification accuracy still has a greater value than in other sets. The resultant execution time is comparable with other approaches. Thus, the NCA algorithm serves as a robust data-driven FS method that could enhance the classification performance while reducing the model complexity and execution time by finding the most powerful features within the original sEMG feature set.

IV. CONCLUSION

In this work, the NCA algorithm has been proposed as the data-driven FS method for sEMG signal classification. The proposed method suggests finding average weights of individual TD features across all the channels and then applying different threshold values to get the optimal feature subset leading to the best trade-off between performance and complexity. With this approach, the results have shown a slight improvement in accuracy for SVM, kNN, and EBDT classification rules and a substantial improvement in the given metric for the LDA model, while reducing their total complexity. Additionally, the NCA-FS method helped to identify the best feature sets acquired from the original one of 44 features that lead to a higher classification performance and comparable complexity than with benchmark feature sets.

Future work may consider generalizing results of NCA-FS to a larger number of subjects from NinaPro DB2 to reveal the most powerful features across all the subjects. Moreover, further studies may focus on using the NCA algorithm as a multivariate FS technique, which would receive the entire feature matrix and provide information about the redundancy of some sEMG channels. The performance of NCA may be compared with other FS methods like Relief or PCC.

References

[1] A. Phinyomark, R. N. Khushaba, and E. Scheme, "Feature extraction and selection for myoelectric control based on wearable EMG Sensors," *Sensors*, vol. 18, no. 5, p. 1615, 2018. pp. 959-962.

[2] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.

[3] A. D. C. Chan and G. Green, "Myoelectric control development toolbox," in *Proc. 2007 30th Conf. Canadian Medical Biological Engineering Society (CMBES)*, 2007, pp. 2-4.

[4] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. C. Chan, , "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.,*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.

[5] F. Douglas, H. Glover, C. Docherty, G. Shields, K. Leventi, and G. Di Caterina, "An exploration of the optimal feature-classifier combinations for transradial prosthesis control ," in *Proc. 2022 44th Annual Int. Conf. IEEE Eng. Medicine Biology Society (EMBC)*, 2022, pp. 2-4.

[6] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG Signal Classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012.

[7] A. Phinyomark, S. Hirunviriya, C. Limsakul, and P. Phukpattaranont, "Evaluation of EMG feature extraction for hand movement recognition based on euclidean distance and standard deviation," in *Proc. 2010 7th Int. Conf. Electrical Engineering/Electronics, Computer, Telecommunications Information Technology (ECTI-CON)*, 2010, pp. 3-5.

[8] A. A. Adewuyi, L. J. Hargrove, and T. A. Kuiken, "Evaluating emg feature and classifier selection for application to partial-hand prosthesis control," *Frontiers in Neurorobotics*, vol. 10, pp. 4-9, 2016.

[9] Q. Li, A. Zhang, Z. Li, and Y. Wu, "Improvement of EMG Pattern Recognition model performance in repeated uses by combining feature selection and Incremental Transfer Learning," *Frontiers in Neurorobotics*, vol. 15, 2021.

[10] J. O. de Oliveira de Souza, M. D. Bloedow, F. C. Rubo, R. M. de Figueiredo, G. Pessin, and S. J. Rigo, "Investigation of different approaches to real-time control of prosthetic hands with electromyography signals," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20674–20684, 2021.

[11] D. Zhang, A. Xiong, X. Zhao, and J. Han, "PCA and LDA for EMG-based control of bionic mechanical hand," in *Proc. Int. Conf. Inf. Autom*, 2012, pp. 960–965.

[12] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. 2004 17th Int. Conf. Neural Inf. Process. Sys. (NIPS)*, 2004, pp. 1-5.

[13] M. Atzori et al., "Characterization of a Benchmark Database for Myoelectric Movement Classification," *IEEE Trans. Neural Systems Rehabilitation Eng.*, vol. 23, no. 1, pp. 73-83, 2015.

[14] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Current Applied Physics*, vol. 11, no. 3, pp. 740–745, 2011.

[15] O. W. Samuel, H. Zhou, X. Li, H. Wang, H. Zhang, A. K. Sangaiah, and G. Li, "Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion classification," *Computers and Electrical Engineering*, vol. 67, pp. 646–655, 2018.

[16] A. Waris and E. N. Kamavuako, "Effect of threshold values on the combination of EMG time domain features: Surface versus intramuscular EMG," *Biomed. Sig. Process. Control*, vol. 45, pp. 267–273, 2018.

[17] P. A. Karthick and S. Ramakrishnan, "Surface electromyography based muscle fatigue progression analysis using modified B distribution time–frequency features," *Biomed. Sig. Process. Control*, vol. 26, pp. 42–51, 2016.

[18] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "Feature extraction of the first difference of EMG time series for EMG Pattern Recognition," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 247–256, 2014.

[19] J. Too, A. Rahim, and N. Mohd, "Classification of hand movements based on discrete wavelet transform and enhanced feature extraction," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.

[20] A. R. Verma and B. Gupta, "Detecting neuromuscular disorders using EMG signals based on TQWT features," *Augmented Human Research*, vol. 5, no. 1, 2019.

[21] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "A framework of temporal-spatial descriptors-based feature extraction for improved Myoelectric Pattern Recognition," *IEEE Trans. Neural Systems Rehabilitation Eng.*, vol. 25, no. 10, pp. 1821–1831, 2017.

[22] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal NeuroEngineering Rehabilitation*, vol. 7, no. 1, 2010.

[23] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Fractal analysis features for weak and single-channel upper-limb EMG signals," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11156–11163, 2012.

[24] D. C. Toledo-Perez, J. Rodriguez-Resendiz, and R. A. Gomez-Loenzo, "A study of computing zero crossing methods and an improved proposal for EMG Signals," *IEEE Access*, vol. 8, pp. 8783–8790, 2020.

[25] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data*, vol. 1, no. 1, 2014.

[26] A. D. Chan and A. R. Goge, "Investigating classification parameters for continuous myoelectrically controlled prostheses," in *Proc. 2004 28th Conf. Canadian Medical Biological Engineering Society (EMBC)*, 2004, pp. 141–144.

[27] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *Journal of Computers*, vol. 7, no. 1, 2012.

[28] E. N. Kamavuako, E. J. Scheme, and K. B. Englehart, "Determination of optimum threshold values for EMG time domain features; a multi-dataset investigation," *Journal Neural Engineering*, vol. 13, no. 4, 2016.

[29] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing Management*, vol. 45, no. 4, pp. 427-437, 2009.