



Simultaneous Force Regression and Movement Classification of Fingers via Surface EMG within a Unified Bayesian Framework

Tara Baldacchino^{1*}, William R. Jacobs¹, Sean R. Anderson¹, Keith Worden² and Jennifer Rowson^{2,3}

¹Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, United Kingdom, ²Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Sheffield, United Kingdom, ³Insigneo Institute for in silico Medicine, University of Sheffield, Sheffield, United Kingdom

This contribution presents a novel methodology for myoelectric-based control using surface electromyographic (sEMG) signals recorded during finger movements. A multivariate Bayesian mixture of experts (MoE) model is introduced which provides a powerful method for modeling force regression at the fingertips, while also performing finger movement classification as a by-product of the modeling algorithm. Bayesian inference of the model allows uncertainties to be naturally incorporated into the model structure. This method is tested using data from the publicly released NinaPro database which consists of sEMG recordings for 6 degree-of-freedom force activations for 40 intact subjects. The results demonstrate that the MoE model achieves similar performance compared to the benchmark set by the authors of NinaPro for finger force regression. Additionally, inherent to the Bayesian framework is the inclusion of uncertainty in the model parameters, naturally providing confidence bounds on the force regression predictions. Furthermore, the integrated clustering step allows a detailed investigation into classification of the finger movements, without incurring any extra computational effort. Subsequently, a systematic approach to assessing the importance of the number of electrodes needed for accurate control is performed via sensitivity analysis techniques. A slight degradation in regression performance is observed for a reduced number of electrodes, while classification performance is unaffected.

Keywords: sEMG signals, finger force regression, finger movement classification, variational Bayes, multivariate mixture of experts, prosthetic hand

1. INTRODUCTION

Estimation of user intention from surface electromyographic (sEMG) signals is a key challenge in the design of active prosthetics, and this method of control has been a thriving area of research for a few decades (Saridis and Gootee, 1982). The harnessing of sEMG signals allows the user to actively control the device, while providing a relatively simple, non-invasive and low cost human-machine-interface. Highly dexterous active arm prosthetics will enable amputees to perform tasks which are not possible with the current state-of-the-art. However, achieving refined control of upper limb prosthetics is particularly challenging due to the very high number of degrees-of-freedom (DoF) present in the extremities of the arm.

OPEN ACCESS

Edited by:

Marie-Christine Ho Ba Tho,
University of Technology of
Compiègne, France

Reviewed by:

Jason Luck,
Duke University, United States
Ahmed Samet,
Institut National des Sciences
Appliquées de Strasbourg, France

*Correspondence:

Tara Baldacchino
t.baldacchino@sheffield.ac.uk

Specialty section:

This article was submitted to
Biomechanics,
a section of the journal *Frontiers in
Bioengineering and Biotechnology*

Received: 07 August 2017

Accepted: 23 January 2018

Published: 26 February 2018

Citation:

Baldacchino T, Jacobs WR,
Anderson SR, Worden K and
Rowson J (2018) Simultaneous Force
Regression and Movement
Classification of Fingers via Surface
EMG within a Unified Bayesian
Framework.
Front. Bioeng. Biotechnol. 6:13.
doi: 10.3389/fbioe.2018.00013

TABLE 1 | A table of related work found in the literature for both finger movement classification and force regression.

# electrodes	Feature Extraction	Classifier (regressor)	Finger moves	# of subjects	Accuracy	Reference
4	WT	ANN	6 s	10H	>80%	Jiang et al. (2005)
4	TD	ANN	4 m	7H	96%	Naik et al. (2009)
32, 19	TD	MLP	10 s	5H, 1A	90%	Tenore et al. (2009)
10	–	SVM	4 m	12H	89.7%	Castellini and van der Smagt (2009)
10	–	(SVM)	4 m	12H	7.98% (NRMSE)	Castellini and van der Smagt (2009)
12	2–3 TDs	SVM	12 s	5H, 1A	>80%	Kanitz et al. (2011)
9	Filtering	(SVM)	6 s, 5 m	12H	<8% (NRMSE)	Castellini and Kõiva (2012)
2	WT	SVM	4 s	11H, 1A	93%	Kumar et al. (2013)
12, 11	TDAR	LDA	11 s, 4 m	10H, 6A	>95%	Al-Timemy et al. (2013)
12	mDWT	(NK)	6 s, 3 m	40H	91.74% (R2)	Gijsberts et al. (2014)
6	TDAR	ANN	5 s	5H	89.4%	Li et al. (2015)
6	TDAR	(Quadratic)	5 s	5H	0.15 (MAE)	Li et al. (2015)
2	TDAR	ANN	5 s, 5 m	8H	92%	Naik and Nguyen (2015)
12 pairs	Various	LDA	5 s	3A	<83%	Jarrassé et al. (2017)

"# electrodes" gives the number of electrodes for healthy, amputee. The type of feature extraction used is given in the second column: WT (wavelet transform), TD (time domain), TDAR (time domain autoregressive), and mDWT (marginal discrete wavelet transform). The classifiers/regressors are given by: ANN (artificial neural network), MLP (multilayer perceptron), SVM (support vector machine), LDA (linear discriminant analysis), and NK (nonlinear kernel). In "Finger Moves" column, "s" = single and "m" = multiple. H and A in the "# of subjects" column correspond to healthy and amputee subjects, respectively. Classifier accuracy is the percentage of correct predictions, and regressor performance uses NRMSE (normalized root mean square error), R^2 , or MAE (mean absolute error).

In order to achieve highly dexterous control of prosthetic arms, control schemes are used that combine classification of movement-type with regression for proportional force control (Castellini and van der Smagt, 2009). Classification of hand/wrist movements (including grasping) has been tackled successfully (Ferguson and Dunlop, 2002; Farrell and Weir, 2008; Atzori et al., 2012). Performing finger movement classification is a more difficult task: first, sEMG signals are smaller in amplitude for finger movements, and second, muscles responsible for finger activations lie deep beneath the skin surface and so the signals recorded at the skin are subject to nonlinear transformations by forearm tissues (Al-Timemy et al., 2013). Several authors have tackled single and multiple finger movement classification using a variety of combinations of sEMG feature extractions (FE) and classifiers, with varying degrees of success (Jiang et al., 2005; Castellini and van der Smagt, 2009; Naik et al., 2009; Tenore et al., 2009; Kanitz et al., 2011; Al-Timemy et al., 2013; Kumar et al., 2013; Li et al., 2015; Naik and Nguyen, 2015; Jarrassé et al., 2017) (refer to **Table 1** for a comprehensive list).

Recently, regression between the user's intent, via sEMG signals, and the force applied at the fingertips has been performed (Castellini and van der Smagt, 2009; Castellini and Kõiva, 2012; Gijsberts et al., 2014; Li et al., 2015) (refer to **Table 1** for references). Regression for proportional force control is a natural and accurate form of control since sEMG signals are related to the force a muscle is applying (Luca, 1997). The whole control scheme, consisting of a classifier coupled with a regressor, is hence capable of predicting the movement being performed along with how much force the subject is exerting.

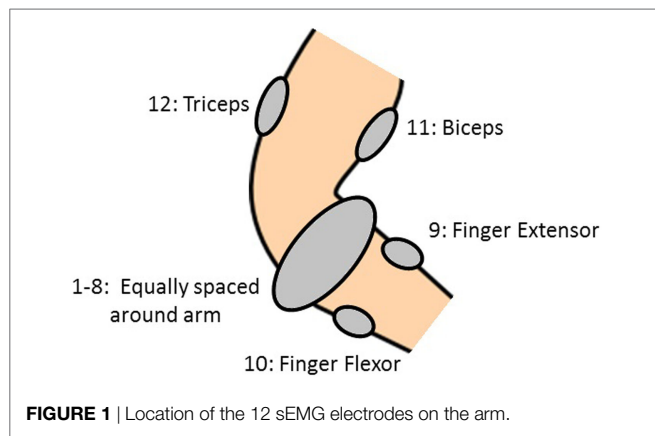
Movement dynamics of a patient observed via sEMG can vary substantially over time due to, e.g., the fatiguing of muscle, motor unit noise, and electrode movement (Luca, 1997). Due to this variability, dynamic models of movement would be improved by intrinsic characterization of uncertainty within the modeling framework to aid robust control design. To date, this aspect of user intention estimation has generally been neglected. In order to address this gap, the authors propose a new method for classifying

movement-type and regressing force based on a mixture of experts (MoE) model, identified within a Bayesian framework which intrinsically characterizes model uncertainty. Further advantages of this approach are: (i) the method simultaneously performs classification and regression, which simplifies the model representation and estimation process, (ii) the model can represent both linear and nonlinear dynamics, and (iii) the estimation algorithm consists of an iterative sequence of closed form expressions, which are computationally simple to execute.

The MoE model used in this work probabilistically divides the input space of a system by means of *gates*, while individual regression *experts* specialize on certain regions of the input space. Thus the adoption of a MoE model exploits the linear relationship between the sEMG signals and the force for each finger movement, giving rise to a more natural modeling description. Additionally, classification is also performed since it is a by-product of the MoE structure which is integral to the model. Thus force regression and movement classification of fingers are implemented within a single unified framework.

Bayesian inference naturally incorporates uncertainty into the training of the model by specifying distributions over the parameters. A useful by-product is the evaluation of confidence bounds of model predictions, hence providing a natural check on the model's predictive capabilities. Due to the Bayesian setting, overfitting is avoided and a metric naturally arises for choosing the appropriate number of experts. Bayesian inference also opens up the possibility of performing patient-specific modeling by incorporating patient-specific heuristics to guide the model building process in order to provide much more accurate predictions for clinical applications. One such approach, for example, is to instruct the Bayesian model averaging process in clinical applications, which has been successfully applied to sepsis and heart failure clinical datasets (Visweswaran et al., 2010).

Cost and power consumption are predominant factors for the success of a method for clinical application, and they can be reduced by minimizing the number of electrodes used for classification and regression. In this work, this problem is addressed



by applying sensitivity analysis (SA) techniques in order to provide a methodical framework for assessing the importance of the individual and grouped sEMG signals on the output force DoF. SA allows an increased understanding of the relationships between the input and output variables of a system/model and has been used extensively within the biomechanics literature, see for example, Batterbee et al. (2011) and Becker et al. (2011).

In this study, the authors build on previous analytical and computational work (Baldacchino et al., 2016) and demonstrate that the proposed multivariate Bayesian MoE model is a suitable candidate for finger force regression and finger classification using sEMG signals since accurate performance, comparable to that found in the literature, was achieved for both regression and classification. The novel work presented in this article provides a powerful method for actively controlling highly dexterous prosthetic and robotic hands for rehabilitation purposes.

The rest of the article is structured as follows. Section 2 describes the sEMG/force dataset used in this article, along with preprocessing methods for the sEMG signals. This section also summarizes the main analysis approaches. The multivariate MoE model is introduced in Section 2.2, and the Bayesian parameter update equations are given in Section 2.3. Section 2.6 gives details of the SA framework used in this work. The results are presented collectively in Section 3. An in-depth discussion regarding the results obtained is given in Section 4. Finally, conclusions on the work presented are drawn in Section 5.

2. THE DATA AND THE METHODOLOGY

2.1. NinaPro Data

The dataset analyzed in this article is comprised of the second version of the publicly released NinaPro (Non-invasive Adaptive Prosthetics) database (Atzori et al., 2012; Gijsberts et al., 2014). The NinaPro database contains sEMG recordings collected from 40 intact subjects while performing a large number of common hand/finger movements and grasping actions. Of particular interest is the dataset containing measurements of the sEMG and the corresponding applied forces at the fingertips for different finger and thumb movements.

Each subject had 12 electrodes attached to their forearm and upper arm in order to measure sEMG signals. The 12 electrodes were placed in such a way that some electrodes provided dense

TABLE 2 | Description of the 9 force patterns.

#	Movement description
F1–F4	Flexion of little through to index fingers.
F5/F6	Abduction/Flexion of the thumb.
F7	Flexion of the index and little finger.
F8	Flexion of the ring and middle finger.
F9	Flexion of the index finger and the thumb.

sampling while others targeted specific muscles. Eight electrodes (1–8) were equally spaced around the forearm at the height of the radio-humeral joint, one electrode each was placed on the finger extensor (9) and flexor (10) muscles, and on the biceps (11) and triceps (12), see Figure 1. Six DoF force measurements were considered: flexion of the five digits (little, ring, middle, index and thumb) along with abduction of the thumb. These force signals were collected by way of a Finger-Force Linear Sensor (FFLS) (Kõiva et al., 2012). Each subject was asked to produce a set of nine force patterns, given in Table 2, by pressing down with one or more digits of the dominant hand in response to an external stimulus. Each force pattern was repeated six times, and a rest period was enforced in between each movement and each repetition in order to prevent muscle fatigue. More information and detail regarding the setup and data collection can be found in Gijsberts et al. (2014), Nin (2014).

The data preprocessing steps employed in this article follow the steps proposed in Gijsberts et al. (2014), since a benchmark for force regression on the NinaPro data has been set. The data was first split into a training and testing set based on the repetition of movements; the second and fifth repetition for each movement were used for testing and the four remaining repetitions were used for training. All the data were then standardized to be zero mean and unit variance, using statistics calculated solely from the training set.

A common practice used in smoothing sEMG signals is to first segment the signal into overlapping windows, and features are then extracted from each window. The sEMG signals were sampled at 2,000 Hz, and following on from Gijsberts et al. (2014) windows of 400 ms (800 samples) with a sliding window increment of 10 ms (20 samples) were used. For computational feasibility the training set was subsampled by a factor of 10 (at regular intervals) resulting in approximately 3,000 data points for training. The subsample rate used here is slightly higher than that used by Gijsberts et al. (2014) for regression analysis but preliminary analysis indicated that a higher subsample rate achieved a better performance. The MoE model is trained using the extracted sEMG features as inputs to the modeling algorithm.

Preliminary results using multivariate MoE for force regression alluded to low-dimensional FE representations outperforming high dimensional FEs (Baldacchino et al., 2015). Thus commonly used low dimensional time-domain FEs for sEMG signals are considered in this work, see Table 3. All the FEs considered have dimensionality equal to the number of sEMG channels (in this case 12). All the FEs, except for FILT, were performed on the windowed data as outlined in the previous paragraph. For FILT FE, a zero-phase second-order low pass Butterworth filter with a cutoff frequency of 2 Hz was applied to the full wave rectified sEMG data, separately for the training and test datasets. In order

TABLE 3 | A table showing the different FEs, used on each channel of the sEMG data, analyzed in this article.

Feature extraction (FE)	Definition (per channel)	References
Mean absolute value (MAV)	$x = \frac{1}{N_w} \sum_{n=1}^{N_w} \tilde{x}_n $	Hudgins et al. (1993), Hargrove et al. (2008), Artemiadis and Kyriakopoulos (2011), and Kuzborskiy et al. (2012)
Waveform length (WL)	$x = \sum_{n=2}^{N_w} \tilde{x}_n - \tilde{x}_{n-1} $	Hudgins et al. (1993), Hargrove et al. (2008), Artemiadis and Kyriakopoulos (2011), and Kuzborskiy et al. (2012)
Root mean square (RMS)	$x = \left(\frac{1}{N_w} \sum_{n=1}^{N_w} \tilde{x}_n^2 \right)^{\frac{1}{2}}$	Gijsberts et al. (2014) and Hahne et al. (2014)
Butterworth filter (FILT)	See text	Atzori et al. (2012) and Lenzi et al. (2012)

The extracted features are denoted by x which are computed from the windowed sEMG signal \tilde{x} of length N_w . The "References" column provides references from the biomedical, robotics and exoskeletons literature where the different types of FEs have been implemented.

to obtain a similar number of data points to the other FE datasets, this dataset was subsampled at regular intervals of 200 samples.

2.2. Multivariate Bayesian Mixture of Experts

The only instance of an MoE model applied to myoprosthetic control appears in Hahne et al. (2014) for regression analysis of wrist movements. The model was limited to two experts and trained using iterative reweighted least squares separately for each DoF. In Section 2.2.1, a sophisticated MoE is defined which is capable of accommodating unlimited DoFs and achieving analytical update solutions. The MoE model structure used in this article follows the work of Ueda and Ghahramani (2002), and further developed in Baldacchino et al. (2016). This particular choice of MoE model was adopted for two reasons; tractable closed form update equations for the parameters are obtained, and it is easily extended to accommodate multivariate outputs. These two characteristics reduce the computational effort during the training phase resulting in fast training times. The model used in this article is thus referred to as the multivariate Bayesian MoE model. Since a Bayesian approach to training the model is employed, prior distributions for the random variables need to be specified and these are given in Section 2.2.2.

2.2.1. Multivariate MoE Regression Model

The d^y DoF force signal, $\mathbf{y}_n = [y_n^1, \dots, y_n^{d^y}]$, at time instant n , can be represented by a MoE model with M regression experts, given by

$$\mathbf{y}_n = \sum_{i=1}^M g_i(\mathbf{x}_n, \theta_i^g) f_i(\mathbf{x}_n, W_i), \quad (1)$$

where $\mathbf{x}_n = [x_n^1, \dots, x_n^{d^x}]$ is the d^x dimensional input vector consisting of the extracted sEMG features. The i th gating function, $g_i(\cdot)$, is a *normalized Gaussian* function, given by

$$g_i(\mathbf{x}_n, \pi_i, \theta_i^g) = \frac{\pi_i \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_i, \Lambda_i^{-1})}{\sum_{l=1}^M \pi_l \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_l, \Lambda_l^{-1})}, \quad (2)$$

where $\theta_i^g = [\boldsymbol{\mu}_i, \Lambda_i]$; $\boldsymbol{\mu} = \{\boldsymbol{\mu}_i\}_{i=1}^M$ is the mean and $\Lambda^{-1} = \{\Lambda_i^{-1}\}_{i=1}^M$ is the covariance matrix. $\boldsymbol{\pi} = \{\pi_i\}_{i=1}^M$ are the mixing coefficients satisfying $\pi_i \geq 0$ and $\sum_{i=1}^M \pi_i = 1$. $g_i(\mathbf{x}_n, \pi_i, \theta_i^g)$ is the posterior conditional probability that \mathbf{x}_n is assigned to the partition corresponding to the i th expert.

The i th parametric expert function, $f_i(\mathbf{x}_n, W_i)$ describes the relationship between the sEMG signals and the force applied at

the fingertips. This function is restricted to be a linear vector such that $f_i(\mathbf{x}_n, W_i) = W_i^T [\mathbf{x}_n \ 1]$ (1 represents a bias term) and W_i is a $(d^x + 1) \times d^y$ matrix corresponding to the weights associated with each sEMG signals and each DoF force signal. The probability distribution, $p(\mathbf{y}_n | \mathbf{x}_n, \theta_i^e)$, of the i th expert is taken to be a multivariate Gaussian distribution having mean $f_i(\mathbf{x}_n, W_i)$ and covariance χ_i^{-1} , given by

$$p(\mathbf{y}_n | \mathbf{x}_n, \theta_i^e) = \mathcal{N}(\mathbf{y}_n | W_i^T [\mathbf{x}_n \ 1], \chi_i^{-1}), \quad (3)$$

where $\theta_i^e = [W_i, \chi_i]$; $W = \{W_i\}_{i=1}^M$ is the multidimensional parameter weight matrix, and $\boldsymbol{\chi}^{-1} = \{\chi_i^{-1}\}_{i=1}^M$ is the covariance matrix.

The set of unknown model parameters for the model expressed in (1) is given by $[\boldsymbol{\pi}, \boldsymbol{\Theta}] = [\boldsymbol{\pi}, \boldsymbol{\theta}^g, \boldsymbol{\theta}^e]$. Given that N training data points are available, then the joint likelihood of the finger force and sEMG signals is expressed as

$$p(Y, X | \boldsymbol{\pi}, \boldsymbol{\Theta}) = \prod_{n=1}^N \sum_{i=1}^M \underbrace{\pi_i \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_i, \Lambda_i^{-1})}_{\tilde{g}_i} \times \mathcal{N}(\mathbf{y}_n | W_i^T [\mathbf{x}_n \ 1], \chi_i^{-1}), \quad (4)$$

where $X = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathcal{R}^{N \times d^x}$, and $Y = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathcal{R}^{N \times d^y}$. The gating network \tilde{g}_i is a Gaussian mixture model (GMM) and it divides the sEMG input space into separate Gaussian clusters. Hence, clustering (unsupervised classification) is performed using GMMs, and GMMs have been previously applied to hand movements as a means of performing classification with minimal computational complexity when compared to supervised learning classifiers (Chan and Englehart, 2003). The form of this gate results in singularities when maximum likelihood is used due to an ill-defined likelihood function (this causes a mixture component to collapse onto a single data point (Bishop, 2006)), hence further justifying the use of a Bayesian approach in this article. Expressing the likelihood as in equation (4) enforces soft competition between the experts such that only one expert will be dominant in a certain region of the input space (Jacobs et al., 1991), see **Figure 2**.

Discrete latent indicator variables, $Z = \{z_{ni}\}_{i=1, n=1}^{M, N} \in \mathcal{R}^{N \times M}$, are introduced such that if $(\mathbf{x}_n, \mathbf{y}_n)$ was generated from the i th expert then $z_{ni} = 1$, else it is 0. Latent variables are introduced in order to simplify expression (4) since the sum over M experts is

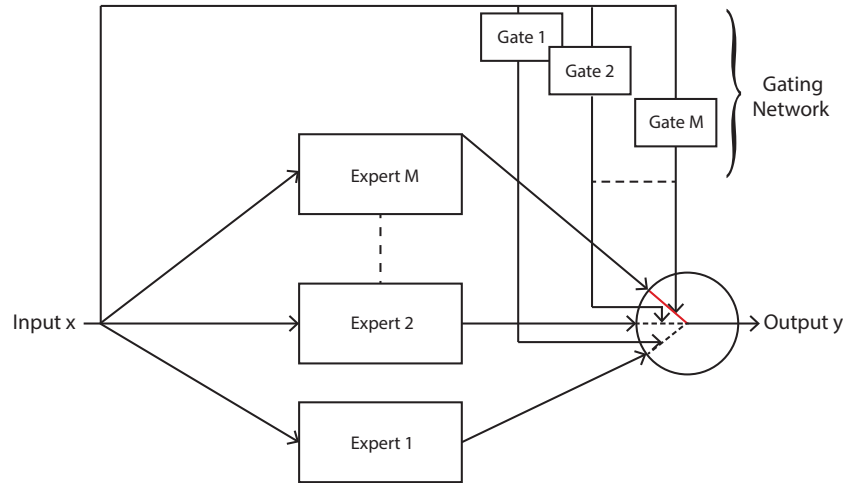


FIGURE 2 | Mixture of experts model with soft competition, whereby only one expert is active at any given time. This expert is chosen probabilistically via the gating function.

changed to a product over M . Thus the complete-data likelihood can be written as

$$p(X, Y, Z | \boldsymbol{\pi}, \boldsymbol{\Theta}) = \prod_{n=1}^N \prod_{i=1}^M \left(\pi_i \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_i, \Lambda_i^{-1}) \times \mathcal{N}(y_n | \mathbf{W}_i' [\mathbf{x}_n \ 1], \chi_i^{-1}) \right)^{z_{ni}}, \quad (5)$$

which is an equivalent representation of (4) incorporating an explicit latent variable.

2.2.2. Priors

Conjugate priors are assigned to all the parameters except for the mixing coefficients $\boldsymbol{\pi}$ (which are treated as non-random variables). The priors used here follow on from the Bayesian literature.

The Gaussian-Wishart prior is used for the mean, $\boldsymbol{\mu}$, and precision, $\boldsymbol{\Lambda}$, of the gating GMM, given by

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\boldsymbol{\mu} | \boldsymbol{\Lambda}) p(\boldsymbol{\Lambda}) = \prod_{i=1}^M \mathcal{N}(\boldsymbol{\mu}_i | \mathbf{m}_0, (\beta_0 \Lambda_i)^{-1}) \mathcal{W}(\Lambda_i | B_0, \nu_0), \quad (6)$$

where B_0 is a $d^x \times d^x$ symmetric, positive definite matrix, and $\nu_0 > d^x - 1$ is the number of degrees of freedom of the Wishart distribution. Similarly, the prior distribution of the joint sEMG weights, \mathbf{W} , and precision matrix, $\boldsymbol{\chi}$, for the linear regressor experts is given by a matrix-Normal-Wishart distribution expressed as

$$p(\mathbf{W}, \boldsymbol{\chi} | \mathbf{a}) = \prod_{i=1}^M \mathcal{MN}(\mathbf{W}_i | 0, A_i^{-1}, \chi_i^{-1}) \times \mathcal{W}(\chi_i | Q_0, \lambda_0), \quad (7)$$

where \mathcal{MN} is the matrix-Normal distribution. The matrix $A_i = \text{diag}(a_{i,1}, \dots, a_{i,d^x+1})$ and it is assigned the following hyperprior distribution

$$p(a_{i,j}) = \mathcal{G}(a_{i,j} | c_0, d_0). \quad (8)$$

The parameter A_i incorporates automatic relevance determination (ARD) since each input sEMG signal has an associated $a_{i,j}$ that forms part of the variance of the i th expert weights' distribution (MacKay, 1995). ARD automatically imposes conditions such that if $a_{i,j}^{-1} = 0$ then the corresponding sEMG signal \mathbf{x}^j will have little effect on the force.

The joint distribution of all the random variables can be expressed hierarchically as,

$$p(Y, X, Z, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{W}, \boldsymbol{\chi}, \mathbf{a} | \boldsymbol{\pi}) = p(Y, X | Z, \boldsymbol{\pi}, \boldsymbol{\Theta}) p(Z | \boldsymbol{\pi}) p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\mathbf{W}, \boldsymbol{\chi} | \mathbf{a}) p(\mathbf{a}), \quad (9)$$

where $\mathbf{a} = (a_1, \dots, a_M)$. In order to simplify the assignment of hyperparameter values for the prior distributions, the extracted feature dataset was again standardized to zero mean and unit variance. This alleviates the issue of assigning hyperparameter values, removing the need to perform hyperparameter optimization. The hyperparameter values were set accordingly so as to define broad priors: \mathbf{m}_0 is set using K-means, $\beta_0 = 0.01$, $B_0 = I$, and $\nu_0 = d^x$ for the gates; and $Q_0 = I$, $\lambda_0 = d^y$, $c_0 = 0.01$, and $d_0 = 0.0001$ for the experts.

2.3. Variational Bayesian Framework

Given the multivariate MoE and priors distributions for the random variables described in the previous sections, a learning algorithm is needed to train the model. An approximate Bayesian framework is desired in order to find the posterior distribution of the parameters $p(\boldsymbol{\Theta}, \mathbf{a} | Y)$. The marginal likelihood $p(Y)$ (which appears in the denominator of Bayes' theorem) is analytically intractable due to a multi-dimensional integral over the parameter space. The choice of conjugate-exponential distributions, along with a latent variable model is elegantly accommodated by the variational Bayes expectation-maximization (VBEM) framework. In short, the VBEM algorithm forms a lower bound of the marginal likelihood which is maximized iteratively so as to obtain a tight bound. Detailed information regarding the VBEM

algorithm can be found in Beal (2003) and Beal and Ghahramani (2003).

Since conjugate priors for the model parameters were used, the functional form of the variational distributions will thus be the same as the priors. The optimal variational distributions are noted below, and expressed as $q^*(\cdot)$. The VBM-step update equations are given in (10)–(16) while the VBE equations for the latent variables Z are shown in equations (17)–(19). The equations given here are similar to the ones detailed in Baldacchino et al. (2016), however, some equations have been modified to cater for the multivariate output force signals.

The joint variational posterior distribution of the gates' mean and covariance is a Gaussian-Wishart distribution, given by

$$q^*(\boldsymbol{\mu}_i, \Lambda_i) = \mathcal{N}(\boldsymbol{\mu}_i | \mathbf{m}_i, (\beta_i \Lambda_i)^{-1}) \mathcal{W}(\Lambda_i | B_i, \nu_i), \quad (10)$$

where

$$\begin{aligned} \mathbf{m}_i &= \frac{\beta_0 \mathbf{m}_0 + \sum_{n=1}^N r_{ni} \mathbf{x}_n}{\beta_i}, \quad \beta_i = \beta_0 + N_i, \\ B_i^{-1} &= B_0^{-1} + \sum_{n=1}^N r_{ni} \mathbf{x}_n \mathbf{x}_n' + \beta_0 \mathbf{m}_0 \mathbf{m}_0' - \beta_i \mathbf{m}_i \mathbf{m}_i' \\ \nu_i &= \nu_0 + N_i, \quad N_i = \sum_{n=1}^N r_{ni}, \quad r_{ni} = \mathbb{E}[z_{ni}]. \end{aligned} \quad (11)$$

The joint variational posterior distribution of the expert functions' mean and covariance is a matrix-Normal-Wishart distribution having the following form

$$q^*(W_i, \chi_i) = \mathcal{MN}(W_i | \hat{W}_i, \chi_i^{-1}, L_i) \mathcal{W}(\chi_i | \lambda_i, Q_i), \quad (12)$$

where

$$\begin{aligned} \hat{W}_i &= L_i [X \mathbf{1}_{N \times 1}]' V_i Y \\ L_i &= ([X \mathbf{1}_{N \times 1}]' V_i [X \mathbf{1}_{N \times 1}] + \Upsilon_i)^{-1} \\ V_i &= \text{diag}(\mathbb{E}[z_{1i}], \dots, \mathbb{E}[z_{Ni}]) \\ \lambda_i &= \lambda_0 + N_i \\ Q_i^{-1} &= Q_0^{-1} + Y' V_i Y - \hat{W}_i' (X' V_i X + \Upsilon_i) \hat{W}_i. \end{aligned} \quad (13)$$

The term $\Upsilon_i = \mathbb{E}[A_i]$ is defined in equation (16). The variational distribution for the ARD parameters is

$$q^*(a_{i,j}) = \mathcal{G}(a_{i,j} | c_i, d_{i,j}), \quad (14)$$

where

$$\begin{aligned} c_i &= c_0 + 0.5d^y, \quad d_{i,j} = d_0 + 0.5\xi_{i,j} \\ \xi_{i,j} &= d^y(L_i)_{j,j} + \lambda_i \hat{w}_{i,j} Q_i \hat{w}_{i,j}', \end{aligned} \quad (15)$$

where $(L_i)_{j,j}$ is the j th diagonal element of L_i , and $\hat{w}_{i,j} \in \mathcal{R}^{1 \times d^y}$ is the j th row of \hat{W}_i . Using the statistic of a mean from a Gamma distribution, then

$$\Upsilon_i = \mathbb{E}[A_i] = \text{diag}\left(\frac{c_i}{d_{i,1}}, \dots, \frac{c_i}{d_{i,d^x+1}}\right). \quad (16)$$

The VBE-step consists of updating the variational distribution of Z and the relevant equations are listed below. The variational distribution for the latent indicator variables follows a multinomial distribution, such that

$$\ln q^*(Z) = \sum_{n=1}^N \sum_{i=1}^M z_{ni} \ln r_{ni} \quad \text{and} \quad r_{ni} = \frac{\gamma_{ni}}{\sum_{l=1}^M \gamma_{nl}}, \quad (17)$$

$$\gamma_{ni} = \pi_i \exp\left\{\frac{1}{2} \left(\ln \hat{\Lambda}_i + \ln \hat{\chi}_i - \varpi_{ni} - \xi_{ni}\right)\right\}. \quad (18)$$

The terms required in equation (18) are,

$$\begin{aligned} \ln \tilde{\Lambda}_i &= \mathbb{E}[\ln |\Lambda_i|] \\ &= \sum_{j=1}^{d^x} \psi\left(\frac{\nu_i + 1 - j}{2}\right) + d^x \ln 2 + \ln |B_i|, \\ \ln \tilde{\chi}_i &= \mathbb{E}[\ln |\chi_i|] \\ &= \sum_{j=1}^{d^y} \psi\left(\frac{\lambda_i + 1 - j}{2}\right) + d^y \ln 2 + \ln |Q_i|, \quad (19) \\ \varpi_{ni} &= \text{Tr}(\mathbb{E}[\Lambda_i] \mathbb{E}[(\mathbf{x}_n - \boldsymbol{\mu}_i)(\mathbf{x}_n - \boldsymbol{\mu}_i)']) \\ &= \nu_i (\mathbf{x}_n - \mathbf{m}_i) B_i (\mathbf{x}_n - \mathbf{m}_i)' + \beta_i^{-1} d^x, \\ \xi_{ni} &= \text{Tr}(\mathbb{E}[\chi_i] \mathbb{E}[(\mathbf{y}_n - [\mathbf{x}_n \mathbf{1}]) W_i (\mathbf{y}_n - [\mathbf{x}_n \mathbf{1}]) W_i']) \\ &= \lambda_i (\mathbf{y}_n - [\mathbf{x}_n \mathbf{1}] \hat{W}_i) Q_i (\mathbf{y}_n - [\mathbf{x}_n \mathbf{1}] \hat{W}_i)' \\ &\quad + d^y [\mathbf{x}_n \mathbf{1}]' L_i [\mathbf{x}_n \mathbf{1}]. \end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

The update equation for the mixing coefficients is obtained by maximizing the variational lower bound with respect to the mixing coefficients. The update equation is given by

$$\pi_i = \frac{1}{N} \sum_{n=1}^N r_{ni}. \quad (20)$$

The equation given above was used to determine the number of mixtures in GMMs (Corduneanu and Bishop, 2001), and it was later extended to a MoE framework in Baldacchino et al. (2016). Using (20), any surplus experts will have their π_i driven to zero, effectively removing them from the model. This equation needs to be interleaved into the iterative procedure since (18) depends on $\boldsymbol{\pi}$. Following on from previous authors the mixing coefficients are updated after a pass of the variational posterior distribution update equations. The equations presented in this section are summarized in **Algorithm 1**.

2.4. Posterior Predictive Distribution

The posterior predictive distribution allows one to perform predictions of the output force signals to new unseen sEMG signals. This distribution also gives information regarding the variance of the predictions, and thus confidence bounds can be calculated. The posterior predictive distribution is expressed as $p(y_{n'} | \mathbf{x}_{n'}, \mathcal{D})$,

ALGORITHM 1 | Multivariate VBEM MoE.**Initialize the hyperparameters:**

- GMM gates, \mathbf{m}_0 , β_0 , \mathbf{B}_0 and ν_0 ,
- Linear experts ρ_0 , λ_0 , c_0 , d_0 and $\Upsilon_j^{(0)} = \frac{c_0}{d_0} \mathbf{I}_{d_x} \forall i$.
- Initialize $\gamma_{n'}^{(0)} \sim \mathcal{U}[0, 1] \forall i, n$.

for $k = 0$: stopping criteria

VBEM-step:

- Evaluate mixing coefficients $\pi^{(k+1)}$ using (20)
- Update the gate parameters using (11).
- Update the expert parameters using (13), (15) and (16).

VBE-step:

- Update for latent variables Z using (17)–(19).

end for

where $\mathcal{D} = [Y, X]$, and $n' = N + 1$ is the new data point. For the mixture of experts described in this article, the predictive distribution is given by,

$$p(\mathbf{y}_{n'} | \mathbf{x}_{n'}, \mathcal{D}) = \sum_{i=1}^M \Omega_{i,n'} \mathcal{T}(\mathbf{y}_{n'} | [\mathbf{x}_{n'} \ 1] \hat{W}_i, \kappa_i \Sigma_i, \kappa_i) \quad (21)$$

where $\{\Omega_{i,n'}\}_{i=1}^M$ take value 1 with probabilities $\{g_{i,n'}^{MAP}\}_{i=1}^M$, respectively. $g_{i,n'}^{MAP} = g_i(\mathbf{x}_{n'}, \pi_i, \theta_{i,MAP}^g)$ using (2) at the maximum *a posteriori* (MAP) estimates $\theta_{MAP}^g = \{\boldsymbol{\mu}_{MAP}, \boldsymbol{\Lambda}_{MAP}\}$ obtained from the posterior distribution (10), and the final value for π . At any given time n' only one $\{\Omega_{i,n'}\}_{i=1}^M$ can be 1 (the rest are zero) corresponding to the gate with the largest probability. $\mathcal{T}([\mathbf{x}_{n'} \ 1] \hat{W}_i, \kappa_i \Sigma_i, \kappa_i)$ is a multivariate Student-*t* distribution with $\kappa_i = \lambda_i - d^y + 1$ degrees of freedom, mean $[\mathbf{x}_{n'} \ 1] \hat{W}_i$ and scale matrix $\kappa_i \Sigma_i$, where $\Sigma_i = Q_i(1 + [\mathbf{x}_{n'} \ 1]' L_i [\mathbf{x}_{n'} \ 1])^{-1}$. The relevant statistics for prediction are

$$\begin{aligned} \mathbb{E}[\mathbf{y}_{n'}] &= \sum_{i=1}^M \Omega_{i,n'} [\mathbf{x}_{n'} \ 1] \hat{W}_i, \\ \text{cov}[\mathbf{y}_{n'}] &= \sum_{i=1}^M \Omega_{i,n'} \frac{\Sigma_i^{-1}}{\kappa_i - 2}. \end{aligned} \quad (22)$$

2.5. Implementation

The multivariate VBEM MoE algorithm, given in **Algorithm 1**, was trained on the sEMG/force data from all 40 subjects, and all the 6 force signals were trained together. The gate divisions on the input space are common to all the outputs. Since 9 movements and a rest period between repetitions and movements are considered here, the algorithm is initialized with 10 experts. The algorithm was run 100 times with random initialization for γ and \mathbf{m}_0 so as to avoid the local maxima problem. The model with the largest lower bound was chosen as being the best model to represent the data. The algorithm was run for all the FE representations listed in **Table 3**, and a comparison between them is considered in Section 3.

2.6. Sensitivity Analysis

In this article, the authors apply SA techniques with the intention of assessing the importance of the number of electrodes to the performance of the modeling framework. Interested readers are

referred to Saltelli et al. (2000, 2008) for an in-depth exploration of SA. Here, the regression-based global SA technique developed in Xu and Gertner (2008) is applied since this method accommodates correlated input signals. The contribution of an individual sEMG signal to the variance of the model force output, V_j , is decomposed into two parts: the uncorrelated contribution VU_j , which describes variations unique to a sEMG input \mathbf{x}^j which are independent of the other sEMG inputs, and correlated contribution VC_j , which explains variations of an sEMG input \mathbf{x}^j which are correlated with other sEMG signals (Xu and Gertner, 2008)

$$V_j = VU_j + VC_j. \quad (23)$$

This discrimination between uncorrelated and correlated contributions highlights, respectively, if the sEMG input itself dominates or if the correlated variations among the sEMG signals dominate. The three indices found in equation (23) can be calculated using the regression-based SA equations found in Xu and Gertner (2008). The final indices of interest are the first-order sensitivity indices given by the total S_j , uncorrelated SU_j and correlated SC_j contributions of sEMG input \mathbf{x}^j such that

$$S_j = \frac{V_j}{V}, \quad SU_j = \frac{VU_j}{V}, \quad SC_j = \frac{VC_j}{V}, \quad (24)$$

where $V = \frac{1}{N_{lhs}} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})^2$ is the estimated variance of the force output, and $\bar{\mathbf{y}}$ is the mean of the force signal. A Latin hypercube sampling (LHS) procedure is used to generate N_{lhs} rank-correlated samples for X from which the output $\{\mathbf{y}\}_{n=1}^{N_{lhs}}$ can be generated using the final MoE model. The LHS strategy used here follows that given in Iman and Conover (1982) since it takes into account the correlations between the inputs when generating samples. Using the sensitivity indices given in equation (24), examination of spurious sEMG signals is possible since they will affect the force output only due to their strong correlations with other significant sEMG inputs (that is, they will have a very low SU_j compared to SC_j). The regression-based SA approach relies on an approximate linear relationship between the sEMG signals and the force. This is not a bad assumption for the data used in this article since in Gijsberts et al. (2014), an average $R^2 = 60\%$ was achieved for a linear model using RMS FE.

3. RESULTS

The multivariate Bayesian MoE algorithm for simultaneous regression/classification was applied to the problem of predicting finger force dynamics and classifying finger movements from the Ninapro dataset. The results are reported here with a comparative evaluation against previous results (Gijsberts et al., 2014): firstly the results of finger force regression are presented and secondly classification. Lastly, SA is performed in order to assess the importance of the electrodes' contribution to the output force signals.

3.1. Finger Force Regression Using Test Data

The performance of the final regression models is analyzed using the normalized root mean square error (NRMSE) measure,

given by

$$NRMSE = \frac{\sqrt{(1/N) \sum_{n=1}^N (y_n - \hat{y}_n)^2}}{y^{max} - y^{min}}, \quad (25)$$

and the coefficient of determination (R^2), given by

$$R^2 = 1 - \frac{\sum_n (y_n - \hat{y}_n)^2}{\sum_n (y_n - \bar{y})^2}. \quad (26)$$

where \hat{y}_n is the model predicted output at time instance n , and \bar{y} is the mean of y . These two statistics were chosen because of their wide use by the bio-robotics and rehabilitation/prosthetics communities in assessing the quality of their models.

The average normalized root mean square error (NRMSE) for each force DoF across all subjects, is plotted for all the different FEs in **Figure 3**. All FEs perform similarly and low NRMSE values are recorded. The multivariate VBEM MoE algorithm performs consistently well across all subjects since the SDs associated with each DoF are small compared to their mean values (i.e., the coefficients of variation are overall less than 30%). Focusing on the FILT FE as being fairly typical, the best results are obtained for the thumb abduction force signal ($3.6 \pm 1.2\%$), while the worst result is for the thumb flexion force signal ($4.96 \pm 1.5\%$). Prediction of little, ring, middle and index flexion force signals have similar performance.

Figure 4 shows the average NRMSE for all subjects (averaged over the 6 DoF force activations) for each FE, including unit SD (given by the black vertical lines). The performance of the algorithm is fairly consistent across all subjects, with most subjects achieving a level of NRMSE below 5%. Subject 37 for MAV has the worst overall NRMSE of $9.61 \pm 1.76\%$ while Subject 36 performs the worst for the other FEs. The best results were achieved by Subject 8 for WL and Subject 33 for MAV having an NRMSE of 2.64 ± 0.46 and $2.64 \pm 0.22\%$, respectively (Subject 33 has the worst NRMSE for RMS and FILT). The mean (over all DoF and all subjects) and corresponding unit SD for both NRMSE (given by red and black horizontal lines in **Figure 4**) and R^2 are reported in **Table 4**. Overall, the FILT FE has the best performance with the other FEs having very similar performance.

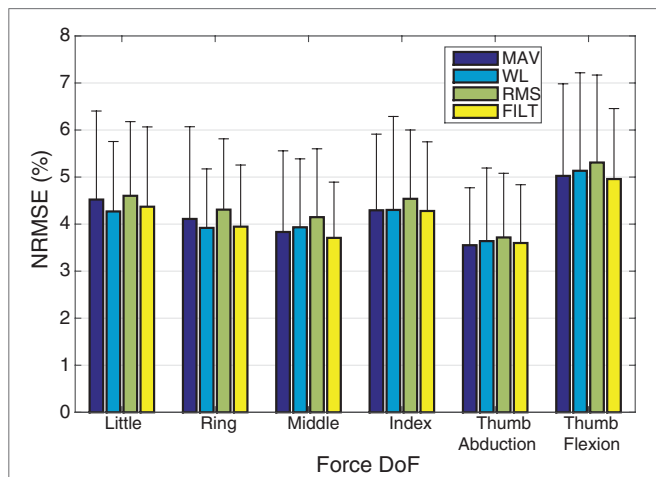


FIGURE 3 | A plot of NRMSE averaged over the 40 subjects for the 6 DoF. The error bars indicate unit SD.

TABLE 4 | Average R^2 and NRMSE including 1 SD.

Feature type	Average R^2	Average NRMSE
MAV	$90.54 \pm 6.93\%$	$4.22 \pm 1.33\%$
WL	$90.75 \pm 5.07\%$	$4.20 \pm 1.08\%$
RMS	$89.81 \pm 5.00\%$	$4.44 \pm 1.06\%$
FILT	$90.91 \pm 4.50\%$	$4.14 \pm 0.92\%$

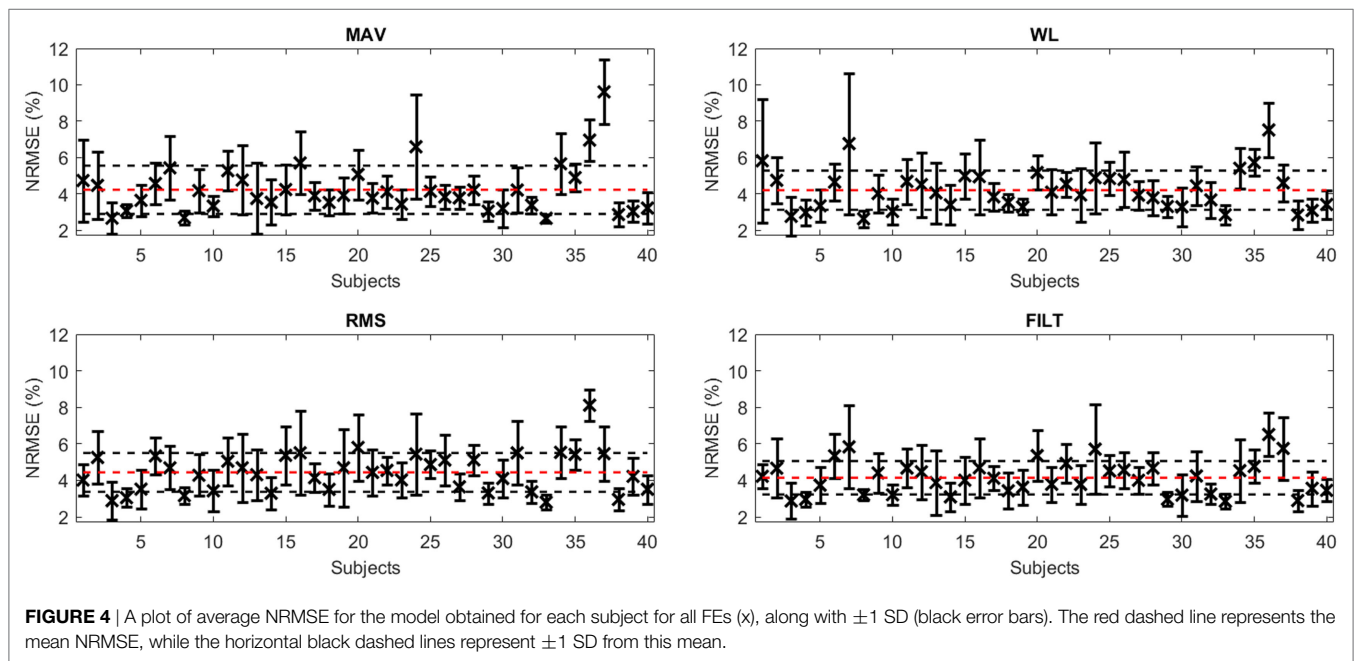
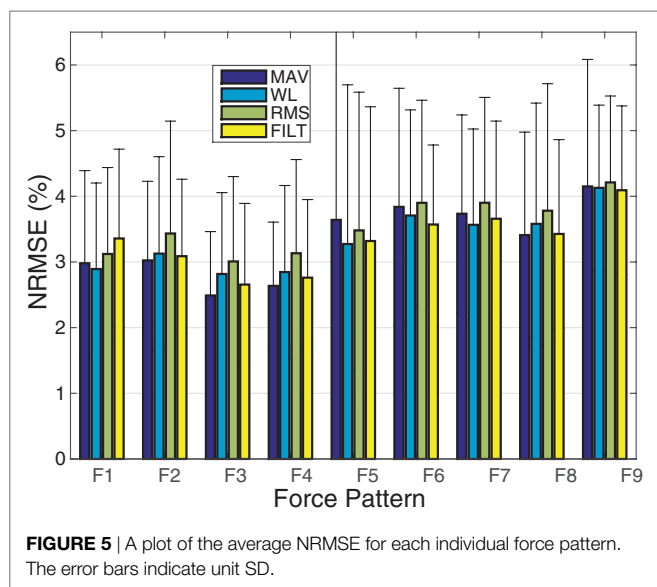


FIGURE 4 | A plot of average NRMSE for the model obtained for each subject for all FEs (x), along with ± 1 SD (black error bars). The red dashed line represents the mean NRMSE, while the horizontal black dashed lines represent ± 1 SD from this mean.



To investigate the performance of the models on each of the 9 force patterns, the average NRMSE per pattern is shown in **Figure 5**. Similar to that reported in *Gijsberts et al. (2014)*, patterns involving the individual activation of the four fingers (F1-F4) are all characterized by slightly better performance overall (i.e., by NRMSE values around 3%). The remaining movements (F5-F9) have a slightly worse performance (between 3.5 and 5%), especially those movements involving flexion of the thumb (F6 and F9); this is consistent with the results shown in **Figure 3**, where thumb flexion shows the worst NRMSE result.

3.1.1. Results for Subject 1

The results reported in this section concern Subject 1 (so as to compare to *Gijsberts et al. (2014)*) and the model obtained with the RMS FE is used since this model has the lowest NRMSE for this particular subject.

The top plot of **Figure 6** shows the predictions (red) obtained on the test data (blue) for the fourth force DoF (index flexion). The signals are plotted on a background of colors where each color corresponds to a particular expert. The black dashed lines represent the 99% confidence intervals which are calculated using (22) hence arising naturally from the Bayesian inference framework. The confidence intervals visibly enclose most of the observed data (blue). The model predictions of the force closely follow the measured data indicating that the multivariate MoE model is a suitable model for performing force regression from sEMG signals; at least so far as Subject 1 is concerned.

An equally interesting result is the assignment of experts to the data points. The different movements are immediately identified as being separate due to being allocated a different expert (denoted by a different color in **Figure 6**). The experts allotted to the different movements in the training set are also the same ones used in the test set for corresponding movements. Details of expert assignment are discussed further in Section 3.2.

The predictions for all 6 DOF force measurements on the test data is shown in the bottom plot of **Figure 6**. All the force measurement signals and force patterns are predicted well, with

the predicted and measured data visibly following the same overall trends. The model even learned the involuntary negative forces, which was also observed by *Gijsberts et al. (2014)* and they attribute these forces to synergistic or compensatory mechanisms.

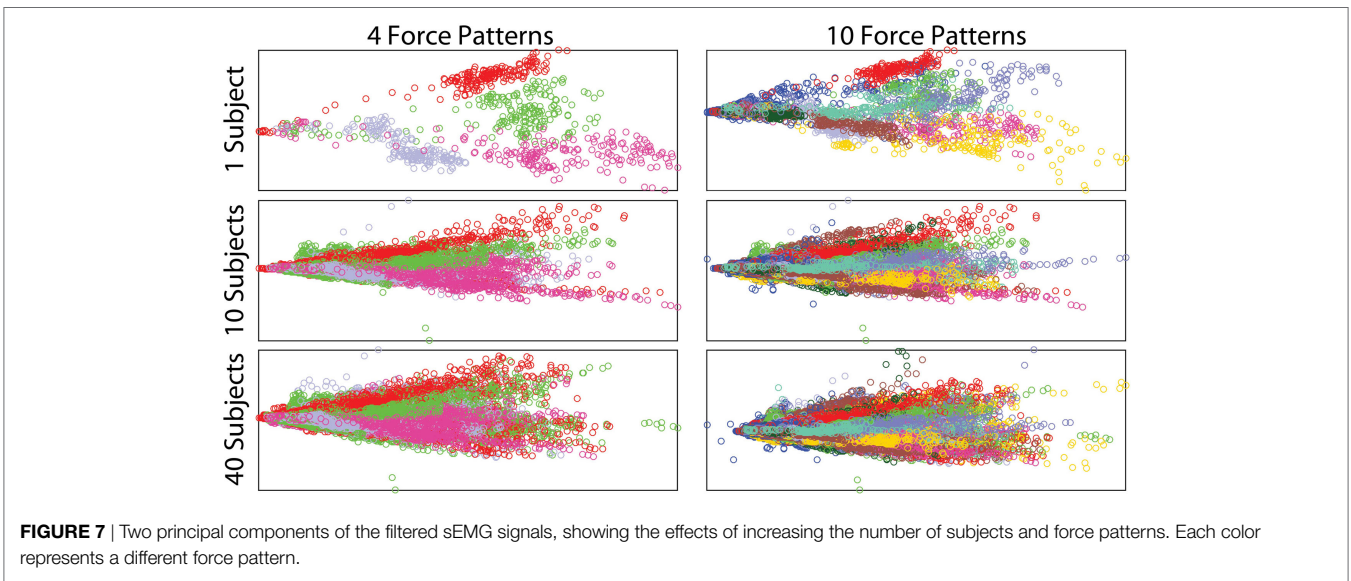
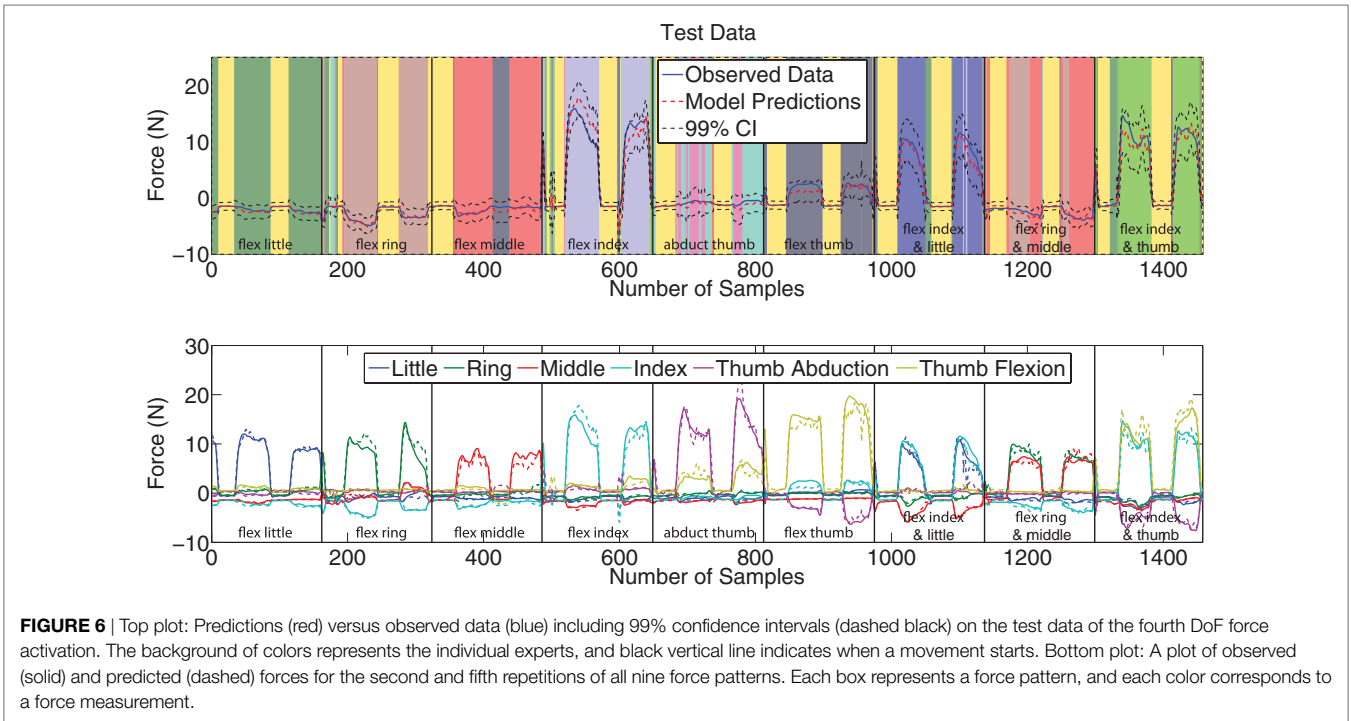
3.2. Finger Movement Classification

Classification can be seen as a natural by-product of the VBEM MoE algorithm presented in this article. The learning algorithm consists of soft competition among the experts, such that only one expert dominates in different regions of the input space. The dominant expert is probabilistically chosen by assessing the gates using (2).

Insight into the relation between the different force movements and the sEMG signals is obtained *via* principal component analysis (PCA). Using PCA on the sEMG signals, the first two principal components can be plotted against each other, with the different force patterns each taking on a unique color, as shown in **Figure 7**. In this plot, the columns indicate 4 and 10 (9 movements plus rest) force patterns, respectively, while the rows show the effect of increasing the number of subjects. Even when just one subject is considered, increasing the number of movements results in the force patterns having lots of overlap, hence making it harder to distinguish between the different movements. Similarly, increasing the number of subjects also results in a large overlap. This figure also highlights the high variability in the recorded sEMG signals between subjects, and it is expected that this variability will be reflected in large deviations in classification performance. High variability among subjects was also observed by *Atzori et al. (2012)* for different hand movements.

A common metric for assessing the quality of the classifier is the classification accuracy, which is the proportion of correctly classified instances to the total number of instances. The left-hand plot in **Figure 8** shows the average classification accuracy for each force pattern where F0 is the rest position, for all the FEs. All FEs achieve similar performance, and multifinger movements have a lower accuracy; more information can be obtained by analyzing the confusion matrix given in the top plot of **Figure 9**.

The confusion matrix profiles the misclassification of movements since correct predictions would result in a distinct black diagonal (left to right), while any non-clear off-diagonal cells are indicative of misclassifications. The top plot in **Figure 9** shows that the VBEM MoE framework is consistent in correctly classifying movements, since the diagonal is prominent. The most conspicuous misclassification is given by the non-clear first column indicating that all force patterns are sometimes mistaken as rest, that is, the absence of movement. This phenomenon was also observed by *Kuzborskij et al. (2012)*, and they attribute this to: 1/the data contains rest-to-movement segments which causes ambiguity and these segments are technically neither rest nor non-rest, and 2/when windowing is used, some windows include sections of rest and non-rest samples and so these samples are again neither rest nor non-rest. For the FILT FE only point 1 is valid (since no windowing is used), and this is enough to confuse the algorithm into thinking that these regions of the data are associated with no movement. In *Chan and Englehart (2003)*, it was reported that this type of error is treated as being acceptable since it is expected that due to mechanical inertia, a prosthesis

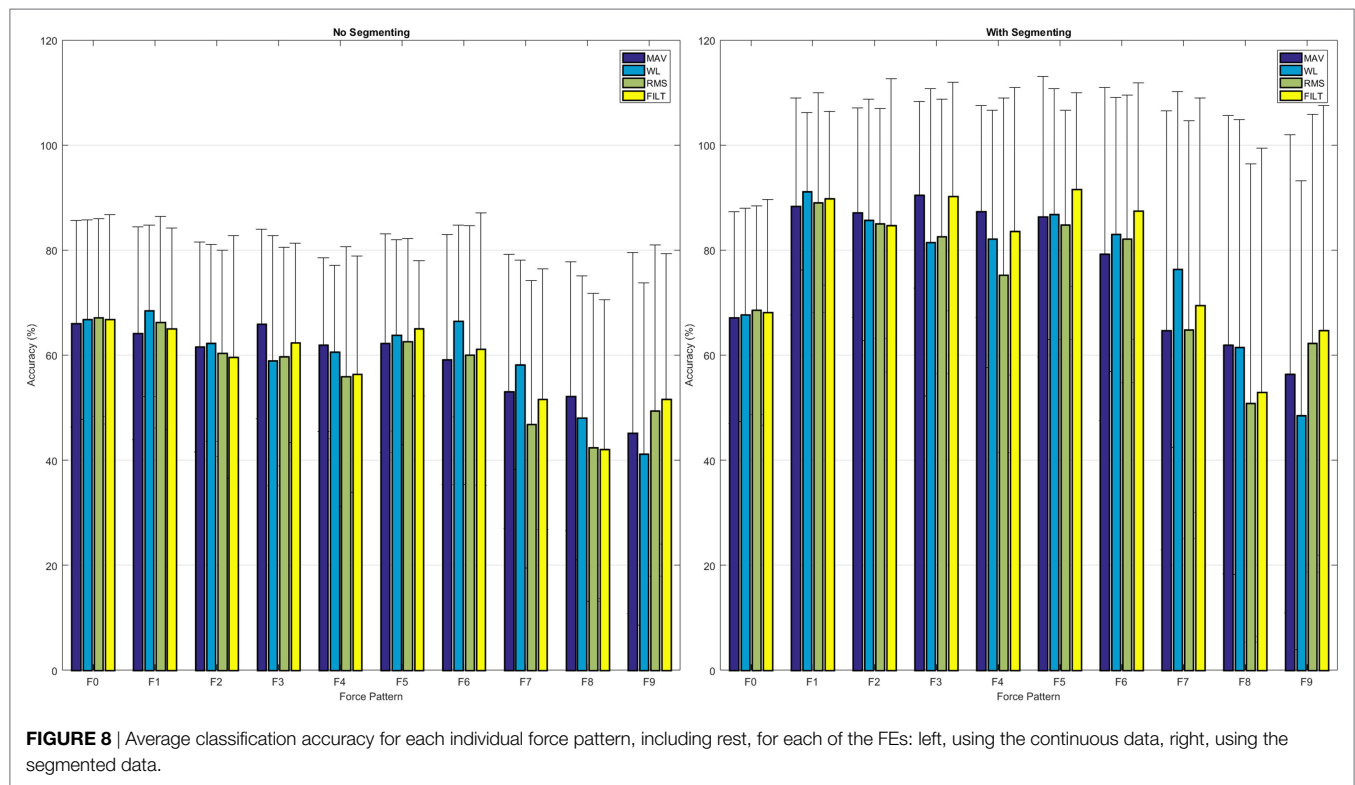


would not be able to respond to transitory misclassifications. In the literature regarding finger movements classification, the rest position is either ignored or the sEMG data stream is segmented in such a way that regions of ambiguity are removed from the training set.

The right hand plot of **Figure 8** and the bottom plot of **Figure 9** show the results when only sections of the data set are considered when calculating classification accuracy, that is, each movement is divided into three equally sized segments and only labels from the center segment were retained (no retraining was performed). A higher accuracy is now attainable: single finger movements and

multi-finger movements have an average accuracy of 87.89% and 62.39%, respectively.

Referring back to the confusion matrix plots (**Figure 9**) gives useful insight into why multi-finger movement achieves lower accuracy than single finger movement. Movement F8 (flexion of ring and middle fingers) has the lowest classification accuracy from all movements. The algorithm mostly misclassified F8 for the corresponding individual finger movements of F2 (flex ring) and F3 (flex middle). The same trend is observed with movements F7 (flexion of index and little fingers) and F9 (flexion of index finger and thumb).



3.3. Reducing the Number of Electrodes via SA

The SA procedure outlined in Section 2.6 is applied to the dataset using the final MoE model obtained with the FILT FE (similar results were obtained using the other models, so only the results of one FE are shown here). All 12 sEMG signals are treated as individual inputs, and the SA results revealed that the contributions of all the sEMG input signals to the variance of the force outputs are mainly due to the correlations with other inputs, since all inputs had a low *SU* index compared to *SC* index.

The SA is performed again but this time the sEMG input signals are grouped according to their electrode location. Thus the groups comprise of electrodes 1–8 (forearm), 9–10 (finger extensor and flexor), and 10–11 (biceps and triceps)—refer to **Figure 1**. The three sensitivity indices, averaged over all 40 subjects, for all the output force signals for these groups of inputs are shown in **Figure 10**. The *SU* index for electrodes 1–8 is now significant (and comparable to the *SC* index) for most outputs. Hence, the sEMG electrodes located around the forearm are necessary for describing the relationship between force at the fingertips and the sEMG signals because this group of inputs' individual contribution is important. The other 2 groups of inputs are almost entirely dominated by their correlated effect, and so their individual contribution is low.

The multivariate VBEM MoE algorithm was run again but this time only the 8 electrodes around the forearm were used as inputs to the model. The plot shown in **Figure 11** shows the performance of the models on the reduced input set (solid lines) versus that of the full input set (dashed lines) for both regression and classification over the different force patterns. The

top plot shows the quality of the models for the force regression given by the average NRMSE for the different force activations considered. A slight degradation in NRMSE is observed for the reduced input set (across all FEs), however, the models are still performing exceptionally well. The bottom plot demonstrates the quality of classifying the different finger movements including the rest position using the average classification accuracy. In this case, the reduced set and full set models have similar performance, with no clear winner.

4. DISCUSSION

4.1. Regression

The results and observations obtained for the force regression using the multivariate VBEM MoE are comparable to those reported in the literature, for example (Castellini and Kõiva, 2012; Gijbsberts et al., 2014). Predictions of thumb movements are worse than for the other fingers due to no sEMG activity being recorded from the majority of the thumb muscles since these are located at the wrist (and hence are not usually available on the amputee's stump) (Kõiva et al., 2012; Gijbsberts et al., 2014).

In the benchmark set by Gijbsberts et al. (2014), overall R^2 values of 91.74 and 88.93% for marginal discrete wavelet transform (mDWT) and RMS, respectively, for a $\exp\text{-}\chi^2$ kernel ridge regression were reported. The results of **Table 4** indicate that the VBEM MoE algorithm achieves similar performance to that given in Gijbsberts et al. (2014) but with the use of a much simpler FE representation for sEMG signals. The mDWT has a considerably higher feature dimensionality (36 inputs) than the FEs considered in this article (12 inputs). The use of a low dimensional FE allows

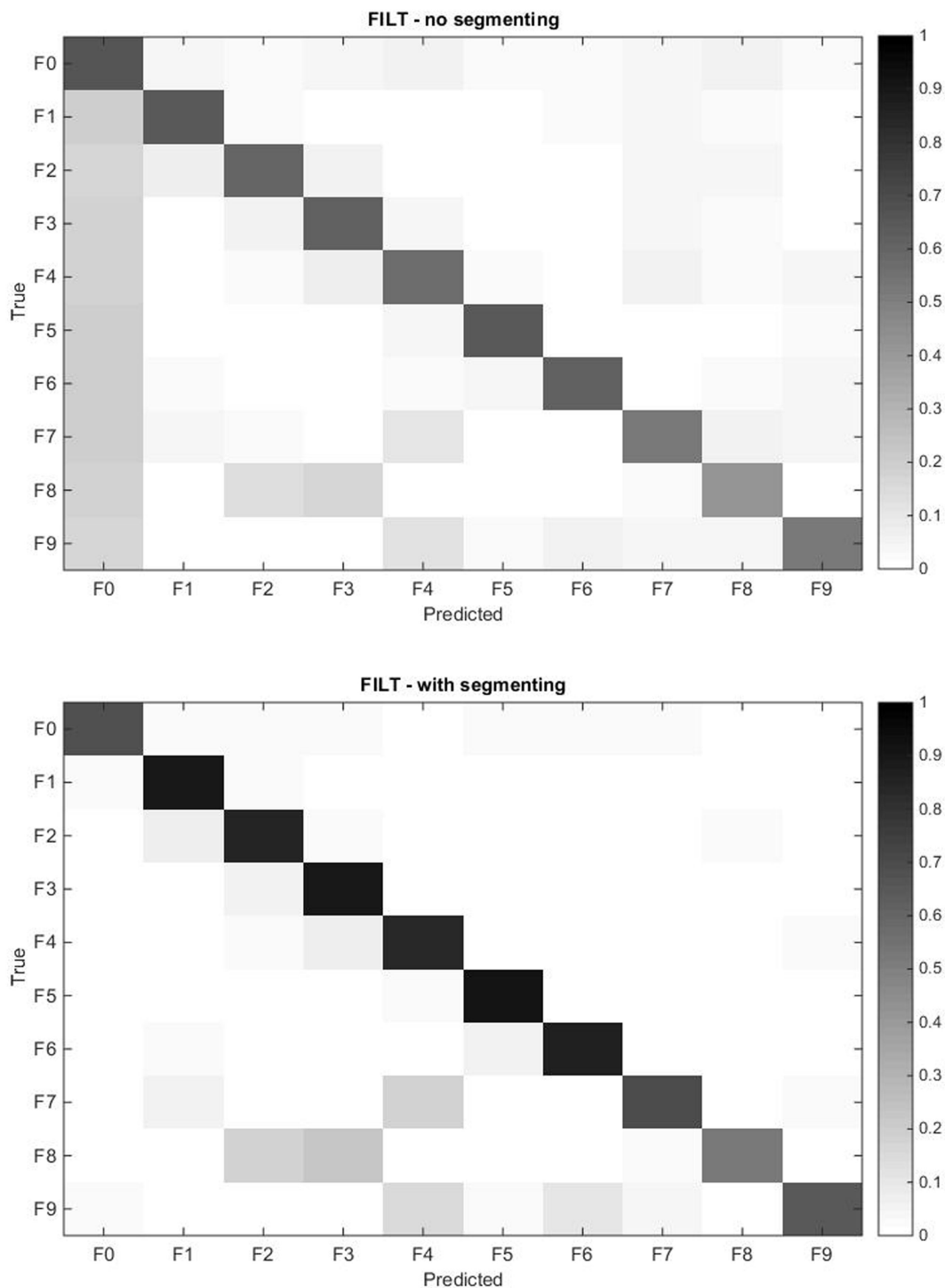


FIGURE 9 | Confusion matrix for the models obtained with the FILT FE averaged over all 40 subjects for all force patterns including rest. The rows represent the true class, and each cell corresponds to the prediction accuracy of each class. The top plot corresponds to the continuous data, while the bottom plot corresponds to the segmented data.

for faster feature computation of the windowed sEMG signals compared to the mDWT in the preprocessing stage. Faster training times are also achieved when a smaller input dimension is used in any algorithm. It has also been shown that TD features

performed better than frequency domain and WTs in real-time applications (Englehart and Hudgins, 2003).

The NRMSE values for each finger force pattern are comparable to those reported in the literature, such as (Castellini and Kōiva,

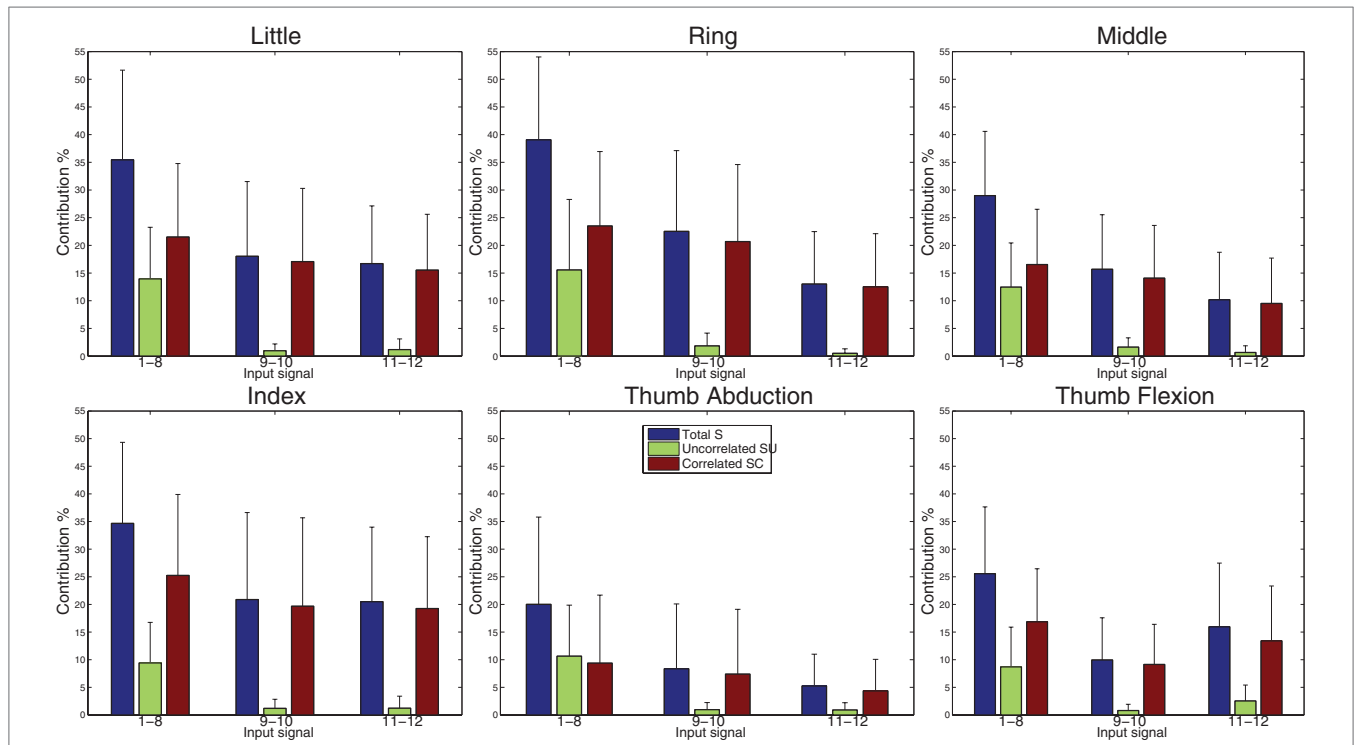


FIGURE 10 | Variance contribution for the six force signals using regression-based sensitivity analysis for correlated inputs (using the MoE model obtained with FILT FE) for grouped sEMG input signals, averaged over all 40 subjects. The large SU values (green) indicate that the group of inputs 1–8 is influential.

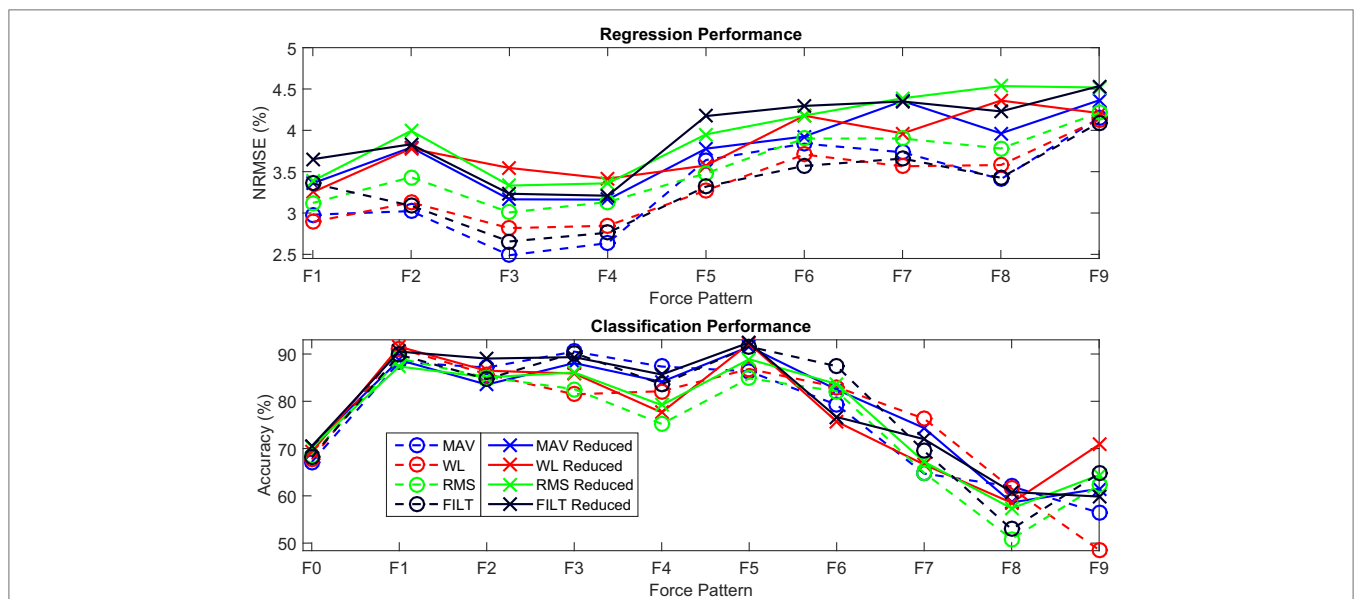


FIGURE 11 | Comparison between the models obtained using all the electrodes (dashed) and using the 8 electrodes around the forearm only (solid) for 4 FEs (MAV, WL, RMS, and FILT). Top: Average NRMSE for each force pattern for assessing the quality of the force regression. Bottom: Average classification accuracy for each force pattern, including rest (F0), for assessing the quality of classification.

2012; Gijsberts et al., 2014). The R^2 values are not reported for the individual force patterns since the sum of error over each force pattern is not necessarily 0, and so the R^2 value can go negative although the model provides a good fit. Therefore, it

is not possible to perform a direct comparison to the R^2 values reported in Gijsberts et al. (2014) since it would result in an erroneous interpretation. However, the NRMSE values reported here highlight what the authors in Gijsberts et al. (2014) and others

have established: force signals from single finger movements are easier to predict than from multiple finger movements and thumb movements.

4.2. Classification

The misclassification of multiple finger movements reported in this article tends to be represented by the corresponding individual finger movements. This result seems to suggest that multiple finger movements can be accurately represented by their corresponding individual finger movements implying that the sEMG signals of the multiple finger movements can be decomposed into the individual sEMG signals. This outcome appears to give weight to the hypothesis presented in Castellini and Kõiva (2012) whereby the authors suggested that for example flexion of ring + middle would be statistically similar to flexion of little + ring + middle. To the best of the authors' knowledge, this phenomenon of sEMG signals of some multiple finger movements being similar to the sEMG signals of the individual fingers has not been analyzed in the literature. Possible explanations for this could be that similar finger movements have not been tackled (some articles do not specify which multiple finger movements are being analyzed), and confusion matrices are not reported hence by-passing the opportunity to assess in detail how the classifier is misclassifying movements.

The classification accuracy reported in this article is similar to that found in the literature, refer to **Table 1**. The algorithm presented here is capable of classifying finger movements which is the result of a by-product from regression analysis—no extra computational expense is incurred. Another important point to note is that this is the first instance of simultaneous regression and classification of finger movements *via* sEMG signals using MoE models.

4.3. SA

The initial SA investigation indicated that the contribution to the force variance was a result of correlated contributions of all the sEMG signals. This result seems to suggest that the VBEM MoE learning algorithm is not sensitive to the nominal placement of electrodes since the uncorrelated contribution of all the electrodes was low suggesting that individual contribution of the electrodes themselves are not important. This result supports the research performed in Hargrove et al. (2008), where the authors report that the pattern recognition framework is insensitive to nominal electrode placements, however, it is sensitive to electrode displacement during the training/testing phase. Hence, the same locations need to be used for training and testing in order for a classifier/regressor to have a high accuracy.

Good force regression and movement classification was achieved by the VBEM MoE model on a reduced input set which consists of a dense sampling of electrodes around the forearm. This setup in which no relevant muscles are targeted was considered by Castellini and Kõiva (2012) such that 9 electrodes were uniformly positioned around the forearm. The regression results obtained here are comparable to those reported by Castellini and Kõiva (2012). Reducing the number of electrodes reduces the overall cost in terms of both physical cost (less electrodes) and

computational cost (training the model). With a training dataset of around 3,000 samples, an average training time of 114 s for the full input set and an average of 93 s for the reduced input set was achieved. Therefore, SA techniques provide a structured methodology for reducing the number of electrodes needed for clinical application.

5. CONCLUSION

Following on from preliminary work, this article provides an in-depth analysis of the use of multivariate Bayesian mixture of experts models using sEMG signals for refined control of prosthetics with multiple DoF. Bayesian inference is a novel concept within the sEMG community and it allows uncertainties to be naturally incorporated into the model structure, supporting a fuller description of the model. The use of a MoE model provides simultaneous finger force regression and finger movement classification at no extra computational cost. The MoE model favors a more natural interpretation between the sEMG and force signals by automatically separating out the data into individual finger movements. This feature relates the model with the underlying biological and physical properties of the data. The MoE model is mathematically and (potentially) computationally demanding compared to other regression and classification methods; however, this is balanced by the extra capability that is obtained.

The method described in this work achieves high performance using low-dimensional feature extraction techniques for the sEMG signal. Accurate force predictions and movement classification were obtained for several finger movements, across all 6 DoF force activations and all 40 subjects. Examination into the effects of a reduced set of sEMG inputs using SA techniques enabled a structured investigation into the influence of the sEMG inputs on the force regression/movement classification performance. It was concluded that the dense sampling of electrodes around the forearm had the greatest influence on the output force DoF. Retraining the models on a reduced input set resulted in similar performance for both the force regression and movement classification when compared to the full input set models.

The next phase of research is to apply this framework to finger force data collected from transradial amputees, and compare performance to that obtained by healthy subjects. It is foreseen that the algorithm will achieve high accuracy, thus providing an exciting novel method of controlling state-of-the-art dexterous myoprosthetics.

AUTHOR CONTRIBUTIONS

TB ran simulations and wrote up most of the article. WJ and SA were involved with the literature review and introduced TB to the NinaPro database. WJ, SA, KW, and JR were involved in discussions regarding machine learning algorithm. JR was PI.

ACKNOWLEDGMENTS

TB was funded by a Leverhulme Trust Research Project Grant. WJ was financially supported by an EPSRC scholarship (EP/K503149/1).

REFERENCES

- Al-Timemy, A. H., Bugmann, G., Escudero, J., and Outram, N. (2013). Classification of finger movements for the dexterous hand prosthesis control with surface electromyography. *IEEE J. Biomed. Health Inform.* 17, 608–618. doi:10.1109/JBHI.2013.2249590
- Artemiadis, P. K., and Kyriakopoulos, K. J. (2011). A switching regime model for the EMG-based control of a robot arm. *IEEE Trans. Syst. Man Cybernet. B* 41, 53–63. doi:10.1109/TSMCB.2010.2045120
- Atzori, M., Gijssberts, A., Heynen, S., Hager, A.-G. M., Deriaz, O., van der Smagt, P., et al. (2012). “Building the NINAPRO database: A resource for the biorobotics community,” in *The Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, Roma.
- Baldacchino, T., Cross, E. J., Worden, K., and Rowson, J. (2016). Variational Bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems. *Mech. Syst. Sig. Process.* 66, 178–200. doi:10.1016/j.ymsp.2015.05.009
- Baldacchino, T., Jacobs, W., Anderson, S. R., Worden, K., and Rowson, J. (2015). “Estimation of fingertip force from surface EMG: a multivariate Bayesian mixture of experts approach,” in *BIO SIGNALS 2015*, Lisbon.
- Batterbee, D. C., Sims, N. D., Becker, W., Worden, K., and Rowson, J. (2011). Computational model of an infant brain subjected to periodic motion simplified modelling and Bayesian sensitivity analysis. *Proc. Inst. Mech. Eng. H* 225, 1036–1049. doi:10.1177/0954411911420002
- Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London.
- Beal, M. J., and Ghahramani, Z. (2003). “The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures,” in *Bayesian Statistics 7*, eds J. M. Bernardo, M. J. Bayarri, A. P. Dawid, J. O. Berger, D. Heckerman, A. F. M. Smith, and M. West (New York: Oxford University Press), 453–464.
- Becker, W., Rowson, J., Oakley, J. E., Yoxall, A., Manson, G., and Worden, K. (2011). Bayesian sensitivity analysis of a model of the aortic valve. *J. Biomech.* 44, 1499–1506. doi:10.1016/j.jbiomech.2011.03.008
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.
- Castellini, C., and Köiva, R. (2012). “Using surface electromyography to predict single finger forces,” in *The Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, Roma, 1266–1272.
- Castellini, C., and van der Smagt, P. (2009). Surface EMG in advanced hand prosthetics. *Biol. Cybern.* 100, 35–47. doi:10.1007/s00422-008-0278-1
- Chan, A., and Englehart, K. (2003). “Continuous classification of myoelectric signals for powered prostheses using Gaussian mixture models,” in *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, Cancun, 2841–2844.
- Corduneanu, A., and Bishop, C. M. (2001). “Variational Bayesian model selection for mixture distributions,” in *Proceedings of the Eighth International Conference on Artificial Intelligence and Statistics*, eds T. Jaakkola, and T. Richardson (Morgan Kaufmann), 27–34.
- Englehart, K., and Hudgins, B. (2003). A robust, real time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 50, 848–854. doi:10.1109/TBME.2003.813539
- Farrell, T. R., and Weir, R. F. (2008). A comparison of the effects of electrode implantation and targeting on pattern classification accuracy for prosthesis control. *IEEE Trans. Biomed. Electron.* 55, 2198–2211. doi:10.1109/TBME.2008.923917
- Ferguson, S., and Dunlop, G. (2002). “Grasp recognition from myoelectric signals,” in *Proceedings of the Australasian Conference on Robotics and Automation*, Auckland, NZ, 83–87.
- Gijssberts, A., Atzori, M., Castellini, C., Müller, H., and Caputo, B. (2014). The movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 735–744. doi:10.1109/TNSRE.2014.2303394
- Hahne, J. M., Mann, F. B., Jiang, N., Rehbaum, H., Farina, D., Müller, F. C. M. K.-R., et al. (2014). Linear and non-linear regression techniques for simultaneous and proportional myoelectric control. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 269–279. doi:10.1109/TNSRE.2014.2305520
- Hargrove, L., Englehart, K., and Hudgins, B. (2008). A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control. *Biomed. Signal Process. Control* 3, 175–180. doi:10.1016/j.bspc.2007.11.005
- Hudgins, B., Parker, P., and Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 40, 82–94. doi:10.1109/10.204774
- Iman, R., and Conover, W. (1982). A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. Simul. Comput.* 11, 311–334. doi:10.1080/03610918208812265
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural. Comput.* 3, 79–87. doi:10.1162/neco.1991.3.1.79
- Jarrassé, N., Nicol, C., Touillet, A., Richer, F., Martinet, N., Paysant, J., et al. (2017). Classification of phantom finger, hand, wrist, and elbow voluntary gestures in transhumeral amputees with sEMG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 68–77. doi:10.1109/TNSRE.2016.2563222
- Jiang, M. W., Wang, R. C., Wang, J. Z., and Jin, D. W. (2005). “A method of recognizing finger motion using wavelet transform of surface EMG signal,” in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, 2672–2674.
- Kanitz, G. R., Antfolk, C., Cipriani, C., Sebelius, F., and Carrozza, M. C. (2011). “Decoding of individuated finger movements using surface EMG and input optimization applying a genetic algorithm,” in *33rd Annual International Conference of the IEEE EMBS*, Boston, MA, 1608–1611.
- Köiva, R., Hilsenbeck, B., and Castellini, C. (2012). “FELS: an accurate linear device for measuring synergistic finger contractions,” in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, San Diego, CA, 531–534.
- Kumar, D. K., Arjunan, S. P., and Singh, V. P. (2013). Towards identification of finger flexions using single channel surface electromyography – able bodied and amputee subjects. *J. Neuroeng. Rehabil.* 10, 50. doi:10.1186/1743-0003-10-50
- Kuzborskij, I., Gijssberts, A., and Caputo, B. (2012). “On the challenge of classifying 52 hand movements from surface electromyography,” in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, Osaka, 4931–4937.
- Lenzi, T., Rossi, S. M. M. D., Vitiello, N., and Carrozza, M. C. (2012). Intention-based EMG control for powered exoskeletons. *IEEE Trans. Biomed. Eng.* 59, 2180–2190. doi:10.1109/TBME.2012.2198821
- Li, X., Fu, J., Xiong, L., Shi, Y., Davoodi, R., and Li, Y. (2015). “Identification of finger force and motion from forearm surface electromyography,” in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)* (San Diego, CA, USA).
- Luca, C. J. D. (1997). The use of surface electromyography in biomechanics. *J. Appl. Biomech.* 13, 135–163. doi:10.1123/jab.13.2.135
- MacKay, D. J. C. (1995). Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks. *Network* 6, 469–505. doi:10.1088/0954-898X_6_3_011
- Naik, G. R., Kumar, D. K., and Arjunan, S. (2009). “Use of sEMG in identification of low level muscle activities: features based on ICA and fractal dimension,” in *31st Annual International Conference of the IEEE EMBS*, Minneapolis, MA, 364–367.
- Naik, G. R., and Nguyen, H. T. (2015). Nonnegative matrix factorization for the identification of EMG finger movements: Evaluation using matrix analysis. *IEEE J. Biomed. Health Inform.* 19, 478–485. doi:10.1109/JBHI.2014.2326660
- Nin. (2014). *Ninapro – Non-Invasive Adaptive Hand Prosthetics*. Available from: <http://www.idiap.ch/en>
- Saltelli, A., Chan, K., and Scott, E. M. (2000). *Sensitivity Analysis*. Chichester, UK: Wiley.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global Sensitivity Analysis. The Primer*. Chichester, UK: Wiley.
- Saridis, G. N., and Gootee, T. P. (1982). EMG pattern analysis and classification for a prosthetic arm. *IEEE Trans. Biomed. Eng.* 29, 403–412. doi:10.1109/TBME.1982.324954
- Tenore, F. V. G., Ramos, A., Fahmy, A., Acharya, S., Etienne-Cummings, R., and Thakor, N. V. (2009). Decoding of individuated finger movements using surface electromyography. *IEEE Trans. Biomed. Eng.* 56, 1427–1434. doi:10.1109/TBME.2008.2005485
- Ueda, N., and Ghahramani, Z. (2002). Bayesian model search for mixture models based on optimizing variational bounds. *Neural Networks* 15, 1223–1241. doi:10.1016/S0893-6080(02)00040-0
- Visweswaran, S., Angus, D. C., Hsieh, M., Weissfeld, L., Yealy, D., and Cooper, G. F. (2010). Learning patient-specific predictive models from clinical data. *J. Biomed. Inform.* 43, 669–685. doi:10.1016/j.jbi.2010.04.009

Xu, C., and Gertner, G. Z. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliab. Eng. Syst. Safety* 93, 1563–1573. doi:10.1016/j.res.2007.06.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Baldacchino, Jacobs, Anderson, Worden and Rowson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.