34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Dynamic SVM Detection of Tremor and Dyskinesia During Unscripted and Unconstrained Activities*

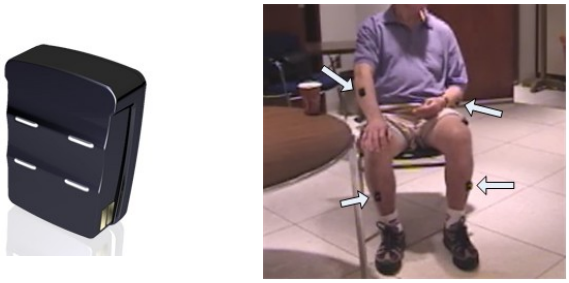
Bryan T. Cole, *Member, IEEE*, Pinar Ozdemir, *Student Member, IEEE* and S. Hamid Nawab, *Senior Member, IEEE*

***Abstract*— In this paper, we report an experimental comparison of dynamic support vector machines (SVMs) to dynamic neural networks (DNNs) in the context of a system for detecting dyskinesia and tremor in Parkinson's disease (PD) patients wearing accelerometer (ACC) and surface electromyographic (sEMG) sensors while performing unscripted and unconstrained activities of daily living. These results indicate that SVMs and DNNs of comparable computational complexities yield approximately identical performance levels when using an identical set of input features.**

I. Introduction

THE management of neuromuscular disorders such as Parkinson's disease (PD) requires individualized treatment plans adjusted by the clinician in response to the movement disorders experienced by the patient on a day-to-day basis. To that end, patients typically fill out diaries or questionnaires detailing the type and severity of movement disorders experienced. However, patient self-reporting of movement disorders typically does not correlate well with expert annotations in a clinical environment. It has been reported that the correlation of these diaries with expert annotations can be as low as 0.49 and as high as 0.74 [1]. The use of a system based on wearable sensors, therefore, has been proposed as an alternative to patient self-reporting that would allow for the objective monitoring of movement disorders experienced by PD patients in an unintrusive fashion.

In our previous work [2], we have developed a system to recognize multiple important movement disorders associated with PD — including tremor [3] and dyskinesia [4] — using dynamic neural networks (DNNs). The DNNs take as input features from a conveniently small number of hybrid sensors worn by the patient as he or she performs unscripted and unconstrained activities of daily living. Each sensor, depicted in Fig. 1, acquires and wirelessly transmits three channels of triaxial accelerometer (ACC) data and one channel of surface electromyographic (sEMG) data, all sampled at 1 kHz (with appropriate anti-aliasing filtering). In order to improve the practicality and convenience of our system, we have trained the DNNs used by our system to detect tremor and dyskinesia based on input features collected from only one ACC and one sEMG sensor placed on the limb of interest.



(a) (b)

Fig. 1. (a) One of the wireless sensors used by our system. The sensor collects three channels of data from a triaxial accelerometer and one channel of surface electromyographic data. (b) A PD patient wearing one sensor on each limb (as indicated by the arrows) during testing in an apartment-like setting.

The DNN-based system has many important qualities that make it superior to previously reported systems. Firstly, it can detect movement disorders in the presence of unscripted and unconstrained activities of daily living. In contrast, previously reported systems restricted the patients to the performance of standardized activities meant to elicit evidence of the desired movement disorders [5]. Additionally, in contrast to previous systems, it requires no subject-specific training – that is, no additional training is needed for our system to recognize movement disorders in data collected from new subjects. Furthermore, the temporal resolution of our system is significantly greater than that of previously reported systems [6]. However, the DNN-based system's performance degrades in certain instances, and it is of interest to determine if replacing DNN technology by SVM technology could lead to improved performance.

II. Previous Work

Our system decides, on a per-second basis, if the movement disorder of interest is present based on features calculated from the ACC and/or sEMG signals and fed as input into a dynamic neural network [7]. This decision is compared to expert annotation derived from video taken of the patient during the sensor data regions of interest. The annotation produces a "ground truth" as to the presence of

*Research supported in part by the Public Health Service/National Institute of Health/National Institute of Biomedical Imaging and Bioengineering (PHS/NIH/NIBIB) under Grant 5 R01 EB007163-05 and under Supplementary Grant 3 R01 EB007163-03S1 from the PHS/NIH/NIBIB.

B. T. Cole was with Dept. of Electrical and Computer Engineering (ECE), Boston University, Boston, MA 02215 USA. He is now with Dräger Medical Systems, Inc., Andover, MA 01810 USA (phone: 978-379-8168; e-mail: bryan.cole@draeger.com).

P. Ozdemir is with ECE at Boston University (email: pozdemir@bu.edu)

S. H. Nawab is with ECE, Dept. of Biomedical Engineering (BME), and NeuroMuscular Research Center (NMRC) at Boston University (email: hamid@bu.edu).

each disorder on a per-second basis, and was created under the supervision of an expert clinician.

The most basic way to describe the rate of discrepancies between the annotation and the system output is in terms of the sensitivity and specificity. Sensitivity depends on the number of true positives (intervals where the system correctly declares the disorder present) and false negatives (intervals where the system declares the disorder absent while the annotation declares it present). In contrast, specificity depends on the number of true negatives (intervals where the system correctly declares the disorder absent) and false positives (intervals where the system declares the disorder present while the annotation declares it absent).

From these basic statistics, we have developed [8] additional criteria to measure the performance of our movement disorder recognition solutions. The global error rate (GER) is defined as the average of sensitivity and specificity, and can be thought of as the error rate computed over a normalized set of testing data such that the number of seconds where the disorder is present equal to the number of seconds where the disorder is absent.

$$\mathrm{GER} = 1 - \frac{\mathrm{Sens} + \mathrm{Spec}}{2} \qquad (1)$$

However, achieving an acceptable error rate overall does not mean that all decisions made by the system are credible; poor performance in the presence of a specific class of voluntary movement will manifest in dense groupings of errors (i.e., false alarms or missed recognitions) over a certain interval. Thus we need an additional error rate — which we call the local error rate (LER) — to measure the performance of our recognition solutions in such "worst-case scenario" intervals. Because we are primarily interested in reducing the number of regions with dense errors, we will use as our figure of merit the proportion of 30-second intervals with an error rate of at least 50%. This proportion is what we refer to as the LER.

The performance of our DNN-based system is given in Table I in terms of both GER and LER for both tremor and dyskinesia as detected in the dominant arm. The performance is measured both overall and conditionally based on the presence of certain movement states (e.g., standing, walking). We can see that, although the system produces acceptable global and local error rates overall, it does not perform equally well across all movement states. For example, the recognition of dyskinesia in the presence of walking has a conditional GER of 12.6%, due largely to a high number of false alarms. Furthermore, nearly 11% of the 30-second intervals in which the subject walks have a high error density, as denoted by the LER. This suggests that during walking our DNN-based system sometimes has difficulty distinguishing intervals with dyskinesia from intervals without dyskinesia.

TABLE I
DNN-BASED TREMOR AND DYSKINESIA DETECTION IN THE ARM

Movement State	Tremor		Dyskinesia	
	GER	LER	GER	LER
Sitting	4.8%	0.23%	6.4%	1.6%
Standing	8.9%	0.19%	6.7%	3.9%
Walking	4.9%	0.18%	12.5%	10.9%
Total	6.2%	0.20%	8.7%	5.2%

GER and LER by movement state achieved by our DNN-based recognition system over the entire testing database. The LER column represents the percentage of all 30-second intervals in the testing database with an error rate greater than or equal to 50%.

In addition to our previous work, several other research groups have proposed systems to monitor individual disorders, relying in part on the patient's performance of scripted activities or standardized tests. However, these tests may interfere with the patient's ability to carry out activities of daily living. Our system, by entirely avoiding the use of scripted activities and standardized tests, represents an important improvement over the existing state of the art.

Support vector machines (SVMs) were used by Patel et al. [6] to recognize both tremor and dyskinesia throughout the whole body from uniaxial ACC sensors. In this study, patients performed standardized tests specifically used by clinicians to elicit and assess movement disorders; the SVMs declared whether the disorders were present on the basis of features calculated over the entire 30-second duration of the test. They reported error rates of 2% for both tremor and dyskinesia. This study relied on the use of subject-dependent training to develop the SVMs; using subject-independent training would simplify the process of adapting the system to new patients.

While several previous papers have focused on the use of ACC sensors to recognize movement disorders, the applicability of sEMG sensors to the recognition of movement disorders is a more recent development. Recent research by Palmes et al. [5] utilizes sEMG sensors to recognize tremor in PD patients. Through the use of an ensemble of several cooperating SVMs, Palmes' group was able to distinguish scripted activities performed by PD patients from those performed by healthy controls with an error rate of 2% using subject-independent training. However, this algorithm does not produce per-second resolution of tremor, producing one decision as to the presence of tremor per scripted activity. The algorithm also does not distinguish dyskinesia from tremor, having been applied only to patients experiencing tremor.

III. SUPPORT VECTOR MACHINES

First developed in 1995, SVMs [9] are a relatively new and increasingly popular machine learning approach. As with neural networks, support vector machines aim to divide the given feature space through the use of generalized decision boundaries known as hyperplanes. However, whereas a neural network attempts to find a suitable hyperplane through iterative procedures such as the gradient descent method [10], the support vector machine defines a linear programming problem to find the optimal hyperplane.

In order to ensure the best possible generalization to new data, we would like to maximize the distance between the hyperplane and any training patterns. This distance is known

as the margin, and is represented in the diagram by d. Our optimal hyperplane $H(\vec{x})$, defined according to $H(\vec{x}) = \vec{w}^T \cdot \vec{x}_k + b$, satisfies the inequality

$$\frac{y_k H(\vec{x}_k)}{|\vec{w}|} \geq d, k = 1,2, \dots, N \qquad (2)$$

over all training patterns N for the largest possible value of the margin d. Here $y_k = \pm 1$, depending on the class of the kth training pattern. The training patterns for which $\frac{y_k H(\vec{x}_k)}{|\vec{w}|} = d$ are known as the support vectors.

But even an optimal hyperplane will not be able to perfectly solve a non-linearly separable problem. Whereas neural networks avoid this problem through the creation of multiple hyperplanes, support vector machines use nonlinear mappings to represent the training patterns in a higher dimensional feature space through the use of a transform function $\varphi(\vec{x})$. Different transform functions, including polynomials, sigmoids, and radial basis functions, can be used according to the distribution of the training patterns in the feature space.

It can be shown [11] that appropriate selection of this transform function can produce a linearly separable feature space for any data set. Once this has been achieved, the optimal hyperplane $H(\vec{x})$ is now defined as

$$H(\vec{x}) = \sum \alpha_k y_k K(\vec{x}_k \vec{x}_l), \qquad (3)$$

where $y_k = \pm 1$ and $K(\vec{x}_k \vec{x}_l)$ is the inner product kernel of the transform function $\varphi(\vec{x})$, $K(\vec{x}_k \vec{x}_l) = \varphi(\vec{x}_k)^T \varphi(\vec{x}_l)$.

The selection of the SVM kernel function, as well as its associated parameters, is dependent on the feature space of the problem and can be found through cross-validation. The kernel function and parameter values that perform best on the cross-validation data are then used in training. Once this selection has been performed, the goal of training is to find the weights $\vec{w}$ and bias b that correctly classify all of the training sequences x_i while maximizing the margin d between the hyperplane $H(\vec{x}) = \vec{w}^T \cdot \vec{x} + b$ and the support vectors s_i. We consider a segment to have been correctly classified if $|\vec{w}^T \cdot \vec{x}_i + b| \geq |y_i|$,where $y_i \in \{-1, +1\}$ is the label associated with the ith training sequence.

Once the hyperplane has been established, it can be used to classify any given sequence $\vec{x}$ according to the decision function

$$sgn\left(\sum_{i=1}^{M} \alpha_i y_i K(\vec{x}, \vec{s}_i) + b\right) \qquad (4)$$

where $K(\cdot)$ is the kernel function, and M is the number of support vectors $\vec{s}_i$ [12]. The support vectors are defined as the training sequences which satisfy the relationship $y_i(\vec{w}^T \cdot \vec{x}_i + b) - 1 = 0$.

The resulting SVM is static, in that the decision for any time n relies only on the features of the data centered at time n. As previously discussed, we desire the use of dynamic classifiers in order to improve recognition of the time-varying movement disorders. To that end, we have designed and implemented dynamic support vector machines [13]. In a DSVM, input features are taken not only from the window centered at the decision point n but also from adjacent decision points. Thus, the DSVM is dynamic in terms of the time-dependency of its input features; the SVM itself, once trained, remains constant across all data. Training and testing of the dynamic algorithm is performed in the same way as training and testing of the static algorithm.

IV. Experiments

This section describes the experiment we designed to compare the performance of our previous DNN-based system to that of the DSVM-based system. As with our DNN-based systems, we have designed two separate DSVMs, one to detect the absence or presence of tremor on a per-second basis, and the other to detect the absence or presence of dyskinesia on a per-second basis. Each DSVM was kept comparable to the DNN designed to detect the same movement disorder insofar as the same training and testing data were used in development. In addition, the same number and type of feature transformations were used by both the DNN and DSVM.

A. Feature Selection

The sEMG and ACC signals collected from the wearable sensors are first passed through a two-second rectangular window. Various features are then extracted from the windowed sections of the sEMG and ACC sensor signals [2]. In all, eight features are calculated, briefly defined as follows: (1) energy of ACC signal after lowpass filtering (LPF) with a cutoff frequency of 1 Hz, (2) energy of ACC signal after highpass filtering (HPF) with a cutoff frequency of 1 Hz, (3) energy of ACC signal after additional HPF with a cutoff frequency of 15 Hz, (4) lag of first peak (not at origin) in autocorrelation of highpass ACC signal, (5) ratio of height of first peak (not at origin) to height of peak at origin in autocorrelation of highpass ACC signal, (6) energy of sEMG signal, (7) lag of first peak (not at origin) in autocorrelation of sEMG signal, provided significant peaks also exist at integer multiples of that lag, and (8) ratio of height of first peak (not at origin) to height of peak at origin in autocorrelation of sEMG signal, provided significant peaks also exist at integer multiples of the first peak's lag.

These features were chosen because of their ability to capture specific qualities of the signal that can be used to differentiate between the movement disorders. Subsets of these eight features are used as inputs to the machine learning classifiers (i.e., DSVMs and DNNs) that recognize the presence of each disorder. The outputs of these classifiers are in turn used in the signal understanding sub-system to identify the presence of tremor and dyskinesia.

B. Training and Cross-Validation

Our development of the DSVM solutions began with the selection of appropriate kernel functions and optimal parameters for the detection of both tremor and dyskinesia. The parameters to be optimized vary according to the choice

of kernel function; in the case of a radial basis function, parameters to be selected include the width of a scale factor γ, and the misclassification trade-off factor C that controls the trade-off between maximum margin and minimum training error.

We trained our DSVM solutions on the identical training data used to develop our DNN solutions with a variety of different kernel functions, specifically linear, polynomial, sigmoid, and radial basis functions. For each kernel function, we adjusted the associated parameters, and determined the global error rate over the training dataset.

To determine which of these will best generalize over the testing dataset, we performed cross-validation over a dataset containing approximately one hour each of tremor, dyskinesia, and disorder-free data from both PD patients and controls. As a result of the training and validation process, we selected a DSVM that used a sigmoid kernel function with a trade-off coefficient (C) of 0.125 and scale factor (γ) of 0.5 to detect dyskinesia. For the detection of tremor, we selected a DSVM that used a radial basis function with a trade-off coefficient (C) of 1 and scale factor (γ) of 0.25.

C. Testing

After the training and validation were completed, we tested our DSVM-based systems on the same 29 hours of data from PD patients and 15 hours of data from healthy controls used to establish the performance level of our DNN-based systems. Table II contains a comparison of the GER for tremor and dyskinesia recognition between the DNN-based system and the DSVM-based system. We can see that the DSVM-based system outperforms the DNN-based system over intervals in which the patient is sitting. However, the DNN-based system has a lower GER overall. Overall, we see that there is no evidence of any significant performance improvement by substituting our DNN-based solution with our DSVM-based solution, at least within the parameters of keeping the same set of input features.

V. Conclusion

Despite the overall performance level reached by our previous DNN-based movement disorder recognition solution [2], we have seen that the misclassifications produced by this solution are not randomly distributed, performing not as well in the presence of certain movement states (e.g., walking). We attempted in this research to resolve these errors by replacing the original dynamic neural network transformations with dynamic support vector machines. However, this alternate signal processing approach had little impact on the performance of movement disorder recognition, both overall and in the presence of those troublesome movement states.

TABLE II
Comparison of Disorder Recognition by DNN and DSVM

Movement Disorder	DNN			DSVM		
	Sens	Spec	GER	Sens	Spec	GER
Tremor	93.4%	94.2%	6.2%	87.8%	97.8%	7.2%
Dyskinesia	93.6%	89.0%	8.7%	90.0%	89.6%	10.2%

Comparison of performance of the DNN and DSVM systems for tremor and dyskinesia over the 44-hour testing database. Performance is measured in terms of sensitivity, specificity, and GER. The overall performance rates are based on the average of the conditional GERs found for each movement state.

Acknowledgment

We would like to thank Dr. Carlo J. De Luca and and Dr. Serge H. Roy for providing their expertise and guidance in the formulation of the overall solution to the problem of detecting movement disorders in PD. We would also like to thank Fahmil Shah for his assistance in carrying out the DSVM experiments reported in this paper. Our thanks to L. Don Gilmore for his assistance in carrying out the data collection from PD subjects, and to Marie Saint-Hilaire, MD, FRCP and Cathi Thomas, RN, MS of the Department of Neurology, Boston University Medical Center, for their clinical assistance in carrying out this project.

References

[1] J. Reimer, M. Grabowski, O. Lindvall, and P. Hagell, "Use and interpretation of on/off diaries in Parkinson's disease," *J. Neurology, Neurosurgery, and Psychiatry*, vol. 75, pp. 396-400, 2004.

[2] B.T. Cole, S.H. Roy, C.J. De Luca, and S.H. Nawab, "Dynamic neural network detection of tremor and dyskinesia from wearable sensor data," in *Proc. of the 32nd Ann. Intl. Conf. of the IEEE EMBS*, Buenos Aires, Argentina, Sept. 1-4, 2010, pp. 6062-6065.

[3] R.J. Elble, "Tremor: Clinical features, pathophysiology, and treatment," *Neurologic Clinics*, vol. 27, no. 3, pp. 679-695, 2009.

[4] J. Jankovic, "Motor fluctuations and dyskinesias in Parkinson's disease: Clinical manifestations," *Movement Disorders*, vol. 20, no. 11, pp. S11-S16, 2005.

[5] P. Palmes, W. T. Ang, F. Widjaja, L. C. S. Tan, and W. L. Au, "Pattern mining of multichannel sEMG for tremor classification," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2795-2805, Dec. 2010.

[6] S. Patel *et al*, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 864-873, Nov. 2009.

[7] E. Wan, "Discrete time neural networks," *Journal of Applied Intelligence*, vol. 3, pp. 91-105, 1993.

[8] B. T. Cole, "Integrated machine learning and signal understanding for movement disorder recognition," Ph.D. dissertation, Dept. Elect. and Comp. Eng., Boston Univ., Boston, MA, 2011.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[10] A.K. Jain, J. Mao, K.M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31-44, 1996.

[11] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2001.

[12] C.-C. Chang and C.-J. Lin. (2010). "LIBSVM (version 3.0)." [Online] Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[13] Q. Li, "Short-time traffic flow prediction based on support vector machine with time-dependent structure," in *IEEE Intl. Instrumentation and Measurement Technology Conference*, Singapore, 2009, pp. 1730-1733.