2016 IEEE 16th International Conference on Data Mining Workshops

Applying Deep Learning to Stereotypical Motor Movement Detection in Autism Spectrum Disorders

Nastaran Mohammadian Rad * †, Cesare Furlanello*
* *Fondazione Bruno Kessler, Trento, Italy*
† *University of Trento, Trento, Italy*
Email:nastaran@fbk.eu

***Abstract*—Autism Spectrum Disorders (ASD) are often associated with specific atypical postural or motor behaviors, of which Stereotypical Motor Movements (SMMs) interfere with learning and social interaction. Wireless inertial sensing technology offers a valid infrastructure for real-time SMM detection, whose automation would provide support for tuned intervention and possibly early alert on the onset of meltdown events. The identification and the quantification of SMM patterns remains complex due to strong inter-subject and intra-subject variability, in particular when handcrafted features are considered. This study aims at developing automatic SMM detection systems in a real world setting, based on a deep learning architecture. Here, after a review of the current state of the art of automatic SMM detection, we propose to employ the deep learning paradigm in order to learn the discriminating features from multi-sensor accelerometer signals. Our results with convolutional neural networks provided the preliminary evidence that feature learning and transfer learning embedded in deep architectures can provide accurate SMM detectors in longitudinal scenarios.**

***Keywords*-Autism, Stereotypical Motor Movement, Accelerometery, Deep Learning, Convolutional Neural Network.**

I. Introduction

Autism spectrum disorders (ASD) are defined as a range of developmental disability conditions that affect at some degree the social interaction and communication abilities of patients. Prevalence of ASD is reported be 1 in 88 individuals [1]. ASD is generally characterized by restricted, repetitive, and stereotyped patterns of behavior in patients. Stereotypical Motor Movements (SMM) in autism (such as body rocking, mouthing, and complex hand movements [2]) can significantly restrict the learning and social interactions. Further, SMMs frequently increase in occasion of emotional or sensory overload which may lead to autistic meltdown events. Alleviating SMMs is thus one primary target of interventions on ASD, which requires accurate tools for recognizing and quantifying SMM patterns. In order to guide behavioral interventions [3] and possibly prevent SMM insurgence, it is worthwhile to consider the limitations of traditional methods for measuring SMM, e.g., paper-and-pencil rating scales, direct behavioral observation, and video-based coding. Measures by wireless accelerometer sensing technology and machine learning techniques provide an automatic, time-efficient, and accurate measure of SMM [3]–[12]. Research by Goodwin et al. [3] showed the potential of automatic SMM detection in real-life settings and the limits towards developing robust and adaptive real-time algorithms. Further, the rapid advance in Inertial Measurament Units (IMU) technology and of its miniaturization, lead to envision a feasible large-scale application of such algorithms [13].

As in many other signal processing applications, SMM detection is commonly based on extracting handcrafted features from the accelerometer signals. So far, a wide variety of feature extraction methods have been used in the literature. Generally two main types of features are extracted from the accelerometer signal [14]: i) time domain features, ii) frequency domain features. For time domain features, some statistical features such as mean, standard deviation, zero-crossing, energy, and correlation are extracted from the overlapping windows of signal. In the case of frequency features the discrete Fourier transform is used to estimate the power of different frequency bands. In addition for SMM classification, the Stockwell transform [15] had been proposed for feature extraction from 3-axis accelerometer signals [3], in order to provide better time-frequency resolution for non-stationary signals. Despite the popularity of handcrafted features in movement analysis, in general, manual feature extraction and selection suffer from two main limitations [16]: i) the feature extraction phase is more based on researchers' domain knowledge rather than information in the raw signal. Therefore, the extracted features may miss important task-related characteristics of the signal, and at the same time they are not robust enough to intra and inter-subject differences; ii) feature extraction and feature selection are computationally intensive steps in the processing pipeline and such computational cost limits the applicability of the movement classification task in real-time scenarios.

To overcome such limitations, we propose to use the deep learning paradigm in order to *learn* discriminating features for SMM pattern detection. In particular, we introduce a convolutional neural network (CNN) deep model [17] to bypass the commonly used feature extraction procedure. The idea of CNN is inspired from the visual sensory system of

2375-9259/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDMW.2016.184

IEEE computer society

living creatures. The initial idea of CNN was first introduced by Fukushima [18] where a hierarchical multi-layer artificial neural network based on local connectivity between neurons were used for visual pattern recognition. Following up on this idea, LeCun et al. [19], developed the CNN deep architecture to address several pattern recognition problems. In addition, the idea of shared weights in the framework of convolutional networks was used in phoneme [20] and spoken word recognition [21]. Having fewer connections and parameters due to weight sharing, CNNs are easier to train compared to other deep neural networks. Currently CNN solutions are among the best performing systems on pattern recognition systems specifically for handwritten character [19] and object recognition [22].

CNNs has been applied beyond audio and image recognition systems, successfully used on various types of signals. Mirowski et al. [23] applied CNN on EEG signals for seizure detection. In the domain of psychophysiology, first time Martinez et al. [16] proposed a model based on CNN to predict affective states of fun, excitement, anxiety, and relaxation. Their proposed model was tested on skin conductance and blood volume pulse signals. Their experimental results showed that automatically learned features outperform the ad-hoc features. Furthermore, this study demonstrated the feasibility of using CNNs on the other types of physiological signals including electroencephalograph (EEG) and electromyography (EMG). Recent studies show the feasibility of applying CNN on accelerometer signals for human activity recognition [24], [25]. Yang et al. [24] proposed a new CNN deep architecture for investigating the multichannel time series data in order to recognize human activity. This architecture mainly uses the convolution and pooling operations to identify the most important patterns of the sensor signals at different time scales, outperforming state-of-the art methods.

In this paper, we hypothesize that the feature learning and transfer learning capabilities of CNN provide more accurate SMM detectors, as well as a platform for learning more robust representation of inertial signals, thus giving the capability of effectively transforming this learned representation to a new dataset which is essential in longitudinal studies [3]. Our preliminary experiments with a well annotated reference dataset [3] support these hypotheses. To the best of the author knowledge, CNN has not been applied in SMM detection applications.

The rest of this paper is organized as follows. First, in Section II we go through the state of the art of automatic SMM detection in ASD children. Then in Section III we will discuss the existing gaps in real-time automatic SMM detection. In Section IV we introduce a hint to the solution based on 1D-CNN architecture. Then, Section V describes the experimental material and set up. In this section, we will benchmark the proposed method versus the state of the art. Finally, Section VI concludes our achievements and states the future directions.

II. State of the Art

Autism spectrum disorders (ASD) are a range of developmental disabilities that affect the social interaction and communication abilities of patients. It has some specific symptoms such as difficulties in social interactions, repetitive or restricted behaviors, verbal and nonverbal communication difficulties. While the majority of studies have mainly focused on social and communication problems of ASD children, the repetitive and restricted behaviors associated with ASD patients received less attention and are recently object of interest [26], [27].

SMMs are one major class of the atypical repetitive behaviors in ASD children. SMMs include hand flapping, body rocking, and mouthing which occur without evoking stimulus [28], [29]. SMMs have a negative effect on the quality of life of ASD children for many reasons. For example, these repetitive behaviors can decrease the performance of children during learning new skills and using the learned skills [30]. Furthermore, since these type of movements are socially abnormal, they are cause of difficulties in the interaction with pairs in the school or other social settings [31]. Finally, in some cases, the severity of SMM leading into meltdown event can cause the self-damaging behaviors [32]. Considering the high prevalence rate of autism in children [33] and the effect of SMM behaviors on the quality of their life, it is essential to develop automatic methods for SMM detection and accurate quantification.

There are three traditional approaches for measuring the SMMs: 1) paper-and-pencil rating, 2) direct behavioral observation, 3) video-based methods. Paper-and-pencil rating scale is an interview-based approach which suffers from the subjectivity in rating. Furthermore, it cannot accurately detect the intensity, amount, and duration of SMM [34]. In the direct behavioral observation approach, therapists can directly observe and record the sequences of SMMs. This method is also not a reliable approach due to the several reasons [35], [36]. First, in the high speed movements, it is impossible for therapists to accurately observe and document all SMMs sequences. Second, determining the start and end time of the SMM sequences is difficult. Third, it is impossible for therapists to concurrently record all environmental conditions and SMMs. Video-based approaches are based on video capturing, offline coding, and analysis of SMMs. Since the captured videos can be reexamined, this method is more accurate than two previous approaches. On the negative side, it is time consuming and therefore it is not applicable as a clinical tool [37].

Considering the limitations of existing methods for measuring SMMs, it is essential to develop time efficient and accurate methods for automatic SMM detection. In this direction, several studies have currently focused on employing accelerometer sensors to detect the stereotypical behavior

of ASD children. Accelerometers are electro-mechanical sensors for measuring the frequency, intensity, and duration of physical activities over a time period. Due to the small size and possibility of embedding in the mobile phones, accelerometers have been accepted as common, useful, and appropriate sensors for wearable devices to measure the physical activities in either constrained and free-living environments [38]. In the following, we go through the current state of the art of the SMM detection using accelerometer signals.

A. Automatic SMM Detection

In 2005, for the first time, Westeyn et al. [4] proposed a proof-of-concept system for data collection, modeling, and self-stimulatory behavior detection using on-body accelerometer sensors. Their preliminary results on the synthetic data showed the feasibility of automatic indexing system for abnormal behaviors. They used hidden Marcov model (HMM) to detect seven stimming behaviors including hand flapping and body rocking from the captured accelerometer data. The synthetic data is collected on a healthy neuro-typical adult mimicking abnormal behaviors. Their proposed method achieved accuracy rates of 90.95% and 92.86% in isolated and continuous settings, respectively. Despite high accuracy rates, their study suffered from the lack of generalization to the realistic data.

Inspired from the work by [4], researchers started using the wearable accelerometer sensors to detect SMMs in actual ASD children. In this direction, Min et al. [5] conducted an experiment in order to detect the optimal location of accelerometer sensors on ASD children's limbs. To achieve this goal, they collected data using 3-axis accelerometer sensors located on the back and wrist of two ASD patients. They used time and frequency domain features with a dictionary-learning algorithm, called K-SVD [39], for automatic SMM detection. Their experimental results showed that employing combination of recorded accelerometer signal placed on the wrist and back provides comprehensive information for SMMs and self-stimulatory behaviors in ASD children. In a similar effort, Goodwin et al. [40] conducted an experiment using Massachusetts Institute of Technology Environmental Sensors (MITes) on six ASD children. By applying J48 Decision Tree classifiers, they achieved accuracy rates of 82% to 97% for hand flapping and body rocking across subjects. These studies showed the feasibility of using accelerometer data and pattern recognition algorithms for reliable and accurate SMM recognition on ASD children in a laboratory environment.

Elsewhere, Goncalves et al. [8] compared the performance of accelerometer sensors with Microsoft Kinect sensor which allows capturing 3D-movements and gesture recognition. To do this, they captured data from four ASD children in multiple 10-minute periods while the Microsoft Kinect sensors were set on the shoulders of subjects. They concurrently used 3-axis accelerometer on the right arm of ASD children to capture their repetitive behaviors. Then two different methods for analyzing the collected data were considered. First, Dynamic Time Warping (DTW) algorithm [41] was applied on captured data from Microsoft Kinect sensor for hand flapping recognition. Alternatively, they analyzed the statistical features such as mean, variance, number of peaks, and root mean square from accelerometer data. Their experimental results demonstrated superiority of performance of accelerometer sensors compared to the Kinect sensor.

During the intervention or therapy of ASD children, one stereotypy may disappear and another stereotypy behavior may develop. Therefore, it is necessary to develop SMM detection systems with the capability of novel event detection. Min et al. [7] proposed Iterative Subspace Identification (ISI) algorithm to extract orthogonal subspaces from 2-axis accelerometer data in order to generate dictionaries for clustering and signal representation. Afterwards, they used a non-parametric method to learn the density function of the observed data. They extracted statistical features such as mean, variance, energy, and number of zero-crossings from the incoming accelerometer signal. By analyzing the generated histogram based on the density of extracted features, they could detect novel events and update the dictionary. Their proposed method presented an average accuracy rate of 83%, 90% and 93% for recognizing hand flapping, body rocking behaviors, and new behavioral pattern detection, respectively.

As a follow up to [7], to improve the result of dictionary-based system, Min et al. [6] proposed Linear Predictive Coding (LPC) and template matching approach to train and update the dictionary. Their experimental results showed the proposed method outperforms their previous study. In the following up study, Min et al. [42] proposed a supervised method for generating the dictionary atoms based on Higher Order Statistics (HOS) features and clustering. Furthermore to learn novel events, they proposed using a semi-supervised method in which the system was trained on a subset of the training data to generate models for known events. Then the other part of training data was used to learn new events in an unsupervised manner using the HOS and LPC. The proposed algorithm produced comparable results with their previous published results [6], [7] which were based on the supervised method for initial dictionary training. Despite promising accuracy rates, the method has high complexity which restricts its application for SMM detection in the real-time scenarios.

In order to provide a real dataset in multiple settings and with varying degrees of complexity, Albinali et al. [10]–[12] collected a comprehensive accelerometer dataset on six ASD individuals engaged in body rocking and hand flapping in the laboratory and classroom settings. The data were collected using three 3-axis accelerometer sensors on the left wrist and right wrist and torso. All the activities of the subjects

were recorded by a video camera and annotated by an expert. Five time and frequency domain features including mean, variance, entropy, correlation, and FFT peaks were extracted from the captured data. Afterwards, by applying C4.5 decision tree classifier they achieved the accuracy rate of 89.5% in the laboratory and 88.6% in the school setting in a single and across-subject classification scenarios.

As a follow up study, Goodwin et al. [3], in a longitudinal study, replicated the same experiment in [11] after three years on the same subjects. The aim of this study was to understand whether the trained classifier on the former dataset can accurately detect SMMs on the new dataset. Further, the authors were interested to evaluate the generalization power of SMM detector from a set of subjects to a new subject. With a data collection similar to previous study, five time and frequency domain features, and additionally, Stockwell transform features [15] were extracted from the captured data. However, high variability was observed in their experimental results across different subjects. The authors concluded that developing adaptive algorithms, that generalize across subjects and over long time intervals, is an emergent need for accurately and consistently SMM detection in real-time scenarios.

Our review on the state of the art shows extensive usage of handcrafted features from the accelerometer signals in time and frequency domains. For time domain features, statistical features such as mean, standard deviation, zero-crossing, energy, and correlation are extracted from overlapping windows of signal. In frequency domain features, the discrete Fourier transform is used to estimate the power of different frequency bands. Despite its popularity in movement analysis, manual feature extraction suffers from two main limitations [16]: a) the feature extraction phase is mainly based on generic researchers' domain knowledge rather than encoding movement information. Thus, characteristics of atypical movements may be missed, without coping with intra-subject and inter-subject variation; b) feature extraction is a computationally intensive step in the processing pipeline and such computational cost limits the applicability of atypical movement detection in real-time scenarios. To address the aforementioned problems of handcrafted features, in this study, we are interested in *learning* the discriminative features from the data.

III. Problem Statement

Developing a real-time SMM detection and quantification system would be advantageous for ASD researchers, caregivers, families, and therapists. Such a system would provide a powerful tool to evaluate the adaptation of subjects with ASD to diverse life context, measuring within an ecologic approach. In particular, it would give a chance of mitigating the onset of those meltdowns that are anticipated by an increase in atypical behaviors. Any automatic quantification of atypical movements would indeed help caregivers and teachers to defuse the mechanism leading to stereotyped behaviors by involving children in specific activities or social interactions. Such involvement decreases the frequency of SMMs and gradually alleviates the duration and severity of abnormal movements [43], [44]. A real-time implementation of SMM detection within a system would help therapists to evaluate efficacy of behavioral interventions. So far, there is no real-time tool for therapist, caregivers and families to accurately and reliably monitor SMMs.

One major challenge toward developing a real-time SMM detection system is personalization due to intra and inter-subject variability [3]. This challenge, despite its crucial importance, has been undervalued [3]. Intra-subject variability is mainly due to the high variability in the intensity, duration, frequency, and topography (type of movements) of SMMs in each individual ASD patient. Inter-subject differences are defined by the same variability issues across different individuals [3]. Existence of these two types of variability within and across ASD persons motivates the necessity of developing an adaptive SMM detection algorithm that is capable to adjust to new patterns of behaviors. The transfer learning paradigm [45] can be considered as a natural candidate solution to attack this challenge.

Hand-crafted features, such as time-domain and frequency-domain features, are vastly used for activity recognition from accelerometer signals [14]. Such features, even if expanded over gyro and magnetometer data, are too elementary to describe SMM patterns along different morphology and personalized patterns. Thus, they are not robust enough to deal with intra and inter subject differences. Further, feature extraction and selection are time consuming tasks [16]. Therefore, developing feature extraction methods, in which robust and informative specification of signal can be directly learned, could offer a significant step towards real-time systems.

IV. Proposed Solution

Let $\mathbf{S}_x^i, \mathbf{S}_y^i, \mathbf{S}_z^i \in \mathbb{R}^{n\times d}$ be n samples of recorded signal by ith $\in \{1, 2, ..., s\}$ acceleremeter sensor with d sampling rate at x, y, and z directions, respectively. Assume $Y^{n\times 1} \in \{-1, 1\}$ be the corresponding label vector for the recorded data where -1 and 1 represent *no-SMM* and *SMM* classes, respectively. Then let $\mathbf{X} \in \mathbb{R}^{n\times c\times d}$ be a 3-dimensional tensor matrix constructed by concatenating the signal of s accelerometer sensors along the sensor-direction dimension (see Figure 1) where $c = s*3$ (3 is the number of directions, i.e. x,y, and z). In other words, we consider each direction of a sensor as an input channel of data.

A. Feature Learning via Convolutional Neural Network

Convolutional Neural Networks (CNN) benefit from invariant local receptive fields, shared weights, and spatio-temporal sub-sampling features to provide robustness over shift and distortion of the input space [17]. CNN has a

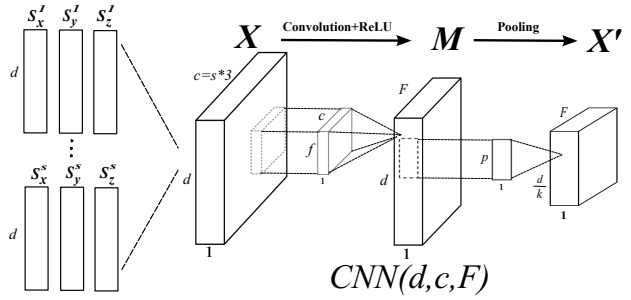

Figure 1: One layer of CNN, $CNN(d, c, F)$, where d is the length of input signal, c is the number of input channels, and F is the number of filters.

hierarchical architecture that alternates convolutional and pooling layers in order to summarize large input spaces with spatio-temporal relations into a lower dimensional feature space. A 1-dimensional convolutional layer $\mathcal{C}^i$ contains a set of F filters, i.e., receptive fields, $\Phi^i = \left\{\mathcal{F}_j^i \in \mathbb{R}^f \mid j \in \{1, 2, \ldots, F\}\right\}$ which learn different patterns on a time window of time-series, where f represents the size of each filter. In fact, the aim is to learn these filters from the input data. Each filter is convolved sequentially with the input signal across channels. Then, the output of the convolution operator is passed through an activation function to compute the feature maps $\mathbf{M} \in \mathbb{R}^{n \times F \times d}$. Generally a rectified linear unit (ReLU) is used as an activation function in a deep architecture where $ReLU(a) = max\{0, a\}$. To reduce the sensitivity of the output to shifts and distortions, feature maps are fed to an additional layer, called pooling layer, which performs a local averaging or sub-sampling. In fact, a pooling layer reduces the resolution of a feature map by factor of $\frac{1}{k}$ where k is the stride size. Max-pooling and average-pooling are two commonly used pooling functions which compute maximum or average value among the values in a pooling window, respectively. This aggregation is separately performed inside each feature map and provides $\mathbf{X'} \in \mathbb{R}^{n \times F \times \frac{d}{k}}$ as the output of pooling layer. $\mathbf{X'}$ can be used as an input to another convolutional layer $\mathcal{C}^{i+1}$ in a multi-layer architecture. Figure 1 illustrates one layer of a 1-dimensional convolutional layer.

B. Network Architecture

Here we propose to use a three-layer CNN to transform the time-series of multiple accelerometer sensors to a new feature space. The proposed architecture is shown in Figure 2. Three convolutional layers $\mathcal{C}^1, \mathcal{C}^2, \mathcal{C}^3$ have 4, 4, and 8 filters with length of 9 samples (i.e., 0.1 second), respectively; therefore $\Phi^1, \Phi^2 = \left\{\mathcal{F}_j^{1,2} \in \mathbb{R}^9 \mid j \in \{1, \ldots, 4\}\right\}$ and $\Phi^3 = \left\{\mathcal{F}_j^3 \in \mathbb{R}^9 \mid j \in \{1, \ldots, 8\}\right\}$. The length of pooling window and the pooling stride are fixed to 3 ($p = 3$) and 2, respectively. Pooling stride of 2 reduces the length of feature maps by factor of 0.5 after each pooling process. The

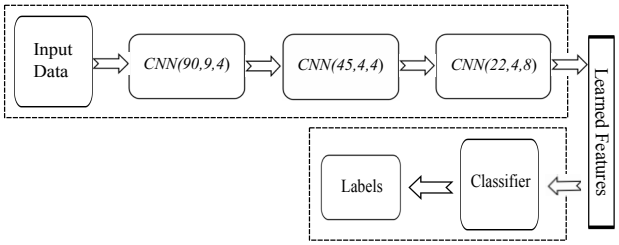

Figure 2: The proposed architecture.

output of the third convolutional layer after flattening provides the learned feature vector. The feature vector is fed to a classifier to predict the labels. This generic architecture can be implemented within existing deep learning toolboxes. We used Deeppy library[1] to configure and train CNN networks. For the sake of fair comparison, in all of our experiments to compare with [3], we used a support vector machine (SVM) with c=1 for classification, i.e., without parameter tuning.

V. Experiments

A. Data and Preprocessing

In our first study, we use the data presented in [3] wherein the accelerometer data were collected over 6 subjects with autism in a longitudinal study[2]. The data is collected in the laboratory and classroom environments while the subjects wearing three 3-axis wireless accelerometers are engaged in body rocking, hand flapping, or simultaneous body rocking and hand flapping. The sensors were worn on the left wrist and right wrist using wristbands, and on the torso using a thin strip of comfortable fabric tied around the chest. To annotate the data, all the activities of the subjects are simultaneously recorded via a video camera and analyzed by an expert. The first data collection, here we call it *Study1*, is recorded by MITes sensors at $60Hz$ sampling frequency [11]. The second dataset, here we call it *Study2*, is collected three years after the first dataset on the same subjects using Wockets sensors with sampling frequency of $90Hz$. To equalize sampling frequencies between two datasets, the data of Study1 is resampled by a linear interpolation to $90Hz$. Furthermore, to remove DC components of signal a $0.1Hz$ cut-off high pass filter is applied. Then, similar to [3], the signal is segmented to 1-second long (i.e., 90 time points) using a sliding window. The sliding window is moved along the time dimension with 10 time-steps resulting in 0.87 overlap between consecutive windows. Considering the data is collected using 3 sensors, $\mathbf{X}$ is a $n \times 9 \times 90$ matrix, where n denotes number of samples. Due to the skewness of classes, same as [3], the training data are balanced based on the number of samples in the

[1] http://andersbll.github.io/deeppy-website/index.html

[2] The dataset and a full description of data is publicly available at https://bitbucket.org/mhealthresearchgroup/stereotypypublicdataset-sourcecodes/downloads.

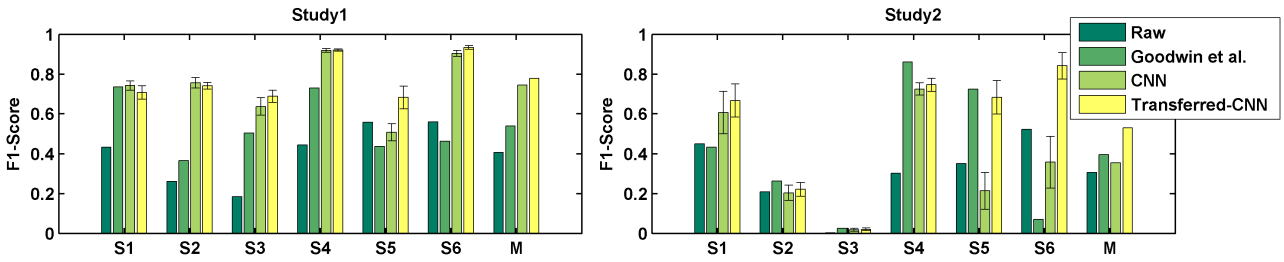

Figure 3: Comparison between results of four experiments.

minority class. The zero component analysis (ZCA) is used for normalizing the data.

B. Experimental Setup

To investigate the effect of feature learning and transfer learning via CNN in SMM recognition, we conducted four experiments. In all experiments the one-subject-leave-out scheme is used for model evaluation.

1) Experiment 1 (Raw Data): The aim of this experiment is to evaluate the effect of both feature extraction and feature learning on the classification performance. Therefore, without any feature extraction, samples of raw data are used as the input to the SVM classifier for SMM detection. In this case, all data channels of each sample in $\mathbf{X}$ are collapsed into a vector and a $n \times 810$ ($810 = 9 \times 90$) input matrix is constructed. We will refer to this experiment as *"Raw"*.

2) Experiment 2 (Goodwin et al.): In this setting, we replicated the third experiment in [3] using exactly the same implementation provided by the authors. All extracted features mentioned in [3] including time, frequency, and Stockwell transform features are used for the classification. We will refer to this experiment as *"Goodwin et al."*.

3) Experiment 3 (CNN): The main aim of this experiment is to investigate the superiority of learning robust features over handcrafted features in the across-subject classification setting. To this end, CNN is used to learn a middle representation of the signal, i.e., to learn the features. In the training phase one layer of 8 hidden neurons followed by two softmax neurons (since it is a binary classification problem) are attached densely to the last layer of CNN. All parameters of CNN are initialized by drawing small random numbers from the normal distribution. The stochastic gradient descent with momentum (the momentum is fixed to 0.9) is used for training the CNN. After training CNN, an SVM classifier is used for classifying the new learned feature space to target labels. All these steps are performed only on the training data to ensure unbiased error estimation. Due to the random initialization of weights and employing stochastic gradient descent algorithm for optimization, results can be different from one run to another. Therefore, we repeated the whole procedure of learning and classification 15 times, reporting the errorbars. This experiment is performed separately on Study1 and Study2 data and will be referred as *"CNN"*.

4) Experiment 4 (Transferred-CNN): In this experiment, we investigate the possibility of transferring learned knowledge from one dataset to another. To this end, we firstly trained the CNN on one dataset, e.g., Study1, and then we used the learned parameters, i.e., filters and weights, for initializing the parameters of CNN in another dataset, e.g., Study2. After retraining the pre-initialized CNN on Study2 dataset, the learned features are again fed to an SVM model for classification. In fact we tried to transform the learned representation from one study to another in a longitudinal study. We refer to this experiment as *"Transferred-CNN"*.

C. Results and Discussions

Figure 3 summarizes the results of four experiments. Due to highly unbalanced classes, the evaluation is performed by computing F1-scores. The bar diagrams are representing the F1-score of four different methods on Study1 and Study2 datasets. The x-axis represents the subjects' ID (S1-S6) and the mean results over all subjects (M). In the case of the first and the second experiments, the result of experiments is deterministic in the one-subject-leave-out scenario. Therefore, no errorbar is reported. Examination of the mean performance on two datasets highlights the following preliminary results:

1) The higher classification performance achieved by handcrafted and learned features with respect to the classification on the raw data illustrates the importance of feature extraction/learning for predicting SMM.
2) Comparison between the results achieved by Goodwin et al. and CNN/transferred-CNN demonstrates the efficacy of feature learning over the manual feature extraction in SMM prediction.
3) Finally, our result shows that transferring knowledge from one dataset to another, by pre-initializing CNN can improve the classification performance in longitudinal studies.

VI. Conclusions and Future Work

In this study, we proposed an original application of deep learning for SMM prediction in ASD children using accelerometer, and in general IMU, sensors. To the best of our knowledge, this is the first effort toward applying deep

learning paradigm for detecting SMMs. Our experimental results showed that convolutional neural network outperforms the traditional classification on the handcrafted features. This observation supports our initial hypotheses about effectiveness of embedded feature learning and transfer learning capabilities of deep neural networks in providing more accurate SMM detection systems. This study is an early effort toward developing real-time SMM detectors. Such a system can be embedded in a mobile-based application to provide the possibility of ubiquitous SMM detection. Our future plans in this direction can be summarized as follows:

1) Improving the learning capability of the CNN by customizing the convolution layers for the heterogeneous multi-channel signals such as accelerometer signals.
2) Testing the proposed system on the collected IMU data using medical quality IMUs, smart-watch or sensorized garments technology.
3) Investigating the possibility of incorporating other physiological signals (such as electrocardiogram and galvanic skin response) in order to predict the SMMs rather than just detecting them.

REFERENCES

[1] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states, 2008. morbidity and mortality weekly report. surveillance summaries. volume 61, number 3." *Centers for Disease Control and Prevention*, 2012.

[2] S. J. LaGrow and A. C. Repp, "Stereotypic responding: a review of intervention research." *American Journal of Mental Deficiency*, 1984.

[3] M. S. Goodwin, M. Haghighi, Q. Tang, M. Akcakaya, D. Erdogmus, and S. Intille, "Moving towards a real-time system for automatically recognizing stereotypical motor movements in individuals on the autism spectrum using wireless accelerometry," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 861–872.

[4] T. Westeyn, K. Vadas, X. Bian, T. Starner, and G. D. Abowd, "Recognizing mimicked autistic self-stimulatory behaviors using hmms," in *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*. IEEE, 2005, pp. 164–167.

[5] C.-H. Min, A. H. Tewfik, Y. Kim, and R. Menard, "Optimal sensor location for body sensor network to detect self-stimulatory behaviors of children with autism spectrum disorder," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3489–3492.

[6] C.-H. Min and A. H. Tewfik, "Automatic characterization and detection of behavioral patterns using linear predictive coding of accelerometer sensor data," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 220–223.

[7] C. H. Min and A. H. Tewfik, "Novel pattern detection in children with autism spectrum disorder using iterative subspace identification," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2266–2269.

[8] N. Gonçalves, J. L. Rodrigues, S. Costa, and F. Soares, "Automatic detection of stereotyped hand flapping movements: two different approaches," in *RO-MAN, 2012 IEEE*. IEEE, 2012, pp. 392–397.

[9] N. Goncalves, J. L. Rodrigues, S. Costa, and F. Soares, "Automatic detection of stereotypical motor movements," *Procedia Engineering*, vol. 47, pp. 590–593, 2012.

[10] F. Albinali, M. S. Goodwin, and S. Intille, "Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 103–114, 2012.

[11] F. Albinali, M. S. Goodwin, and S. S. Intille, "Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum," in *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 2009, pp. 71–80.

[12] M. S. Goodwin, S. S. Intille, F. Albinali, and W. F. Velicer, "Automated detection of stereotypical motor movements," *Journal of autism and developmental disorders*, vol. 41, no. 6, pp. 770–782, 2011.

[13] F. Casamassima, A. Ferrari, B. Milosevic, P. Ginis, E. Farella, and L. Rocchi, "A wearable system for gait training in subjects with parkinsons disease," *Sensors*, vol. 14, no. 4, pp. 6229–6246, 2014.

[14] N. F. Ince, C.-H. Min, A. Tewfik, and D. Vanderpool, "Detection of early morning daily activities with static home and wearable wireless sensors," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 31, 2008.

[15] R. G. Stockwell, L. Mansinha, and R. Lowe, "Localization of the complex spectrum: the s transform," *Signal Processing, IEEE Transactions on*, vol. 44, no. 4, pp. 998–1001, 1996.

[16] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *Computational Intelligence Magazine, IEEE*, vol. 8, no. 2, pp. 20–33, 2013.

[17] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, 1995.

[18] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE transactions on acoustics, speech, and signal processing*, vol. 37, no. 3, pp. 328–339, 1989.

[21] L. Bottou, F. F. Soulie, P. Blanchet, and J.-S. Liénard, "Speaker-independent isolated digit recognition: multilayer perceptrons vs. dynamic time warping," *Neural Networks*, vol. 3, no. 4, pp. 453–465, 1990.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[23] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg," in *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*. IEEE, 2008, pp. 244–249.

[24] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proceedings of the 24th International Conference on Artificial Intelligence*. AAAI Press, 2015, pp. 3995–4001.

[25] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*. IEEE, 2014, pp. 197–205.

[26] E. Hedman, L. Miller, S. Schoen, D. Nielsen, M. Goodwin, and R. Picard, "Measuring autonomic arousal during therapy," in *Proc. of Design and Emotion*, 2012, pp. 11–14.

[27] J. Hernandez, I. Riobo, A. Rozga, G. D. Abowd, and R. W. Picard, "Using electrodermal activity to recognize ease of engagement in children during social interactions," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 307–317.

[28] G. Berkson and R. Davenport, "Stereotyped movements of mental defectives: I. initial survey." *American Journal of Mental Deficiency*, 1962.

[29] R. M. Ridley and H. F. Baker, "Stereotypy in monkeys and humans," *Psychological medicine*, vol. 12, no. 01, pp. 61–72, 1982.

[30] J. W. Varni, O. I. Lovaas, R. L. Koegel, and N. L. Everett, "An analysis of observational learning in autistic and normal children," *Journal of Abnormal Child Psychology*, vol. 7, no. 1, pp. 31–43, 1979.

[31] R. Jones, D. Wint, and N. Ellis, "The social effects of stereotyped behaviour," *Journal of Intellectual Disability Research*, vol. 34, no. 3, pp. 261–268, 1990.

[32] C. H. Kennedy, "Evolution of stereotypy into self-injury." 2002.

[33] S. Goldman, C. Wang, M. W. Salgado, P. E. Greene, M. Kim, and I. Rapin, "Motor stereotypies in children with autism and other developmental disorders," *Developmental Medicine & Child Neurology*, vol. 51, no. 1, pp. 30–38, 2009.

[34] D. A. Pyles, M. M. Riordan, and J. S. Bailey, "The stereotypy analysis: An instrument for examining environmental variables associated with differential rates of stereotypic behavior," *Research in Developmental Disabilities*, vol. 18, no. 1, pp. 11–38, 1997.

[35] R. L. Sprague and K. M. Newell, *Stereotyped movements: Brain and behavior relationships.* American Psychological Association, 1996.

[36] N. C. Gardenier, R. MacDonald, and G. Green, "Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders," *Research in Developmental Disabilities*, vol. 25, no. 2, pp. 99–118, 2004.

[37] J. L. Matson and M. Nebel-Schwalm, "Assessing challenging behaviors in children with autism spectrum disorders: A review," *Research in Developmental Disabilities*, vol. 28, no. 6, pp. 567–579, 2007.

[38] M. J. Mathie, A. C. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiological measurement*, vol. 25, no. 2, p. R1, 2004.

[39] M. Elad, M. Aharon, and A. Bruckstein, "The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representations," *IEEE Trans. Image Process*, vol. 15, no. 12, pp. 3736–3745, 2006.

[40] M. S. Goodwin, S. S. Intille, W. F. Velicer, and J. Groden, "Sensor-enabled detection of stereotypical motor movements in persons with autism spectrum disorder," in *Proceedings of the 7th international conference on Interaction design and children*. ACM, 2008, pp. 109–112.

[41] F. A. Kondori, S. Yousefi, H. Li, S. Sonning, and S. Sonning, "3d head pose estimation using the kinect," in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*. IEEE, 2011, pp. 1–4.

[42] C.-H. Min and A. H. Tewfik, "Semi-supervised event detection using higher order statistics for multidimensional time series accelerometer data," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 365–368.

[43] S. Lee, S. L. Odom, and R. Loftin, "Social engagement with peers and stereotypic behavior of children with autism," *Journal of Positive Behavior Interventions*, vol. 9, no. 2, pp. 67–79, 2007.

[44] R. L. Loftin, S. L. Odom, and J. F. Lantz, "Social interaction and repetitive motor behaviors," *Journal of Autism and Developmental Disorders*, vol. 38, no. 6, pp. 1124–1135, 2008.

[45] S. J. Pan and Q. Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359, 2010.