IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING ALGORITHMS FOR INTERNET OF MEDICAL THINGS

Received February 5, 2020, accepted February 17, 2020, date of publication March 12, 2020, date of current version March 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980269

Abnormal Infant Movements Classification With Deep Learning on Pose-Based Features

KEVIN D. MCCAY[1], EDMOND S. L. HO[1], HUBERT P. H. SHUM[1], (Senior Member, IEEE), GERHARD FEHRINGER[1], CLAIRE MARCROFT[2], AND NICHOLAS D. EMBLETON[2]

[1]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.
[2]Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE7 7DN, U.K.

Corresponding author: Edmond S. L. Ho (e.ho@northumbria.ac.uk)

This work was supported in part by the Royal Society under Grant IES\R1\191147 and Grant IES\R2\181024, and in part by the NIHR Fellowship under Grant ICA-CDRF-2018-04-ST2-020.

ABSTRACT The pursuit of early diagnosis of cerebral palsy has been an active research area with some very promising results using tools such as the General Movements Assessment (GMA). In our previous work, we explored the feasibility of extracting pose-based features from video sequences to automatically classify infant body movement into two categories, normal and abnormal. The classification was based upon the GMA, which was carried out on the video data by an independent expert reviewer. In this paper we extend our previous work by extracting the normalised pose-based feature sets, Histograms of Joint Orientation 2D (HOJO2D) and Histograms of Joint Displacement 2D (HOJD2D), for use in new deep learning architectures. We explore the viability of using these pose-based feature sets for automated classification within a deep learning framework by carrying out extensive experiments on five new deep learning architectures. Experimental results show that the proposed fully connected neural network *FCNet* performed robustly across different feature sets. Furthermore, the proposed convolutional neural network architectures demonstrated excellent performance in handling features in higher dimensionality. We make the code, extracted features and associated GMA labels publicly available.

INDEX TERMS Deep learning, feature extraction, classification, infants, pose-based features.

I. INTRODUCTION

Automated human action recognition has been an active area of research for a number of years [2]. The ability to automatically recognise, analyse and reconstruct complicated motion, such as human activity, has wide ranging applications including content based video indexing, intelligent monitoring, surveillance, human-computer interaction and virtual reality [40]. Building upon our previous work [26], we propose that this technology could be applied to the healthcare domain, specifically in paediatrics, to aid with the early diagnosis of movement disorders, such as cerebral palsy.

Cerebral palsy is an umbrella term that covers a group of lifelong neurological conditions usually caused by a brain injury occurring before, during or shortly after birth [32]. It is a condition that primarily affects movement, posture and coordination, although it can manifest in a range of other complications, such as swallowing difficulties, speech problems, vision problems and learning disabilities. The severity of these symptoms can vary quite significantly, with some individuals presenting very minor symptoms, whilst others may be severely disabled. It is estimated that around 1 in every 400 babies born in the UK develop some form of cerebral palsy [15], suggesting that there may be as many as 1,800 new cases of cerebral palsy every year. Whilst the continual development and enhancement of neonatal care has provided a significant decline in infant mortality rates, studies suggest that this has also contributed towards an increase in the incidence and associated severity of cerebral palsy [29].

Early diagnosis is seen as key in providing the best possible outcome for individuals with cerebral palsy, as it can allow for early intervention care. However, in many cases, diagnosis is not confirmed until 18 months of age or later for those who present mild symptoms [25]. Early identification not only provides a framework whereby the patient can receive the best possible care, but it also allows for the targeting of resources and for the deployment of parental support systems [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

Additionally, access to health, social and educational services often rely upon a diagnosis [33].

The means of providing reliable early diagnosis of cerebral palsy has been investigated for a number of years, with some tools, such as the General Movements Assessment (GMA) [11], producing some very promising results. These tools evaluate the quality, complexity and spontaneity of the infant's movements at a specific window in their development, typically 12 to 20 weeks post-term.

Whilst tools like the GMA have the potential to provide an important understanding of early neurological development, they are not without their problems. In particular, the ability to apply these assessments depends upon the availability of fully trained clinicians. These clinicians require significant training, as well as years of practical assessment experience to achieve a suitable level of accuracy. Additionally, given the manual, time-consuming nature of the assessment, it is highly susceptible to observer fatigue. Additionally, the assessment is subjective, with no discernibly quantifiable features present in the current diagnostic method. The test is also heavily reliant upon the infant being in a suitable behavioural state [14]. Tests such as the GMA are typically only used where there are existing medical concerns, such as prematurity, stroke, lack of oxygen, or congenital heart disease; they are not currently used as a screening tool for healthy babies [4].

Given the nature of the problems found in tools such as the GMA, it is conceivable that the development of automated systems could help to alleviate some of these issues. It is highly likely that the production of an automated pipeline would help to reduce the time and cost associated with current manual diagnostic practices. Additionally, a system which is able to quantify an early diagnosis of cerebral palsy has the potential to aid healthcare professionals in relaying information to the patient's family more reliably. Furthermore, the development of a suitably reliable automated tool would mean that the analysis of all babies could be carried out, helping healthcare professionals establish any additional care requirements.

Several early works [1], [28], [36] propose the use of automated systems for cerebral palsy diagnosis. These systems typically make use of video-based optical flow methods, frequency analysis and background subtraction. However, each of these methods suffered from a lack of robustness in dealing with unnecessary information, illumination changes, body part dimensions, and external influences, such as parental interaction with the infants. Other methods include using wearable accelerometers [16], and whilst the tracking can provide accurate results, the logistics associated with assessing an infant using wearable sensors make them less suitable. Our focus therefore remains on a vision based approach, as such we have developed a pipeline based around pose extraction from 2D RGB footage.

In our pilot study [26], we evaluated the viability of analysing the pose and the joint specific movements of infants as a means of automatically diagnosing movement conditions. We suggest that pose-based analysis presents several advantages over other methods, such as lower dimensionality of features, the ability to deal with multiple people in frame, reduced ambiguity in classification, and the ability to remove superfluous information from the classification process.

Based upon our encouraging results in pose-based classification, we suggest that a deep learning method would be well suited to working with pose-based data in the healthcare domain. As such, we propose a deep learning classification framework which makes use of the extracted, normalised pose-based features. We extend this work by undertaking additional data pre-processing and normalisation, as well as carrying out extensive experiments on several deep learning pipelines in order to determine their feasibility. Our experiments examine the effectiveness of 3 separate types of neural network architecture in classification of the extracted pose-based feature sets. We carry out ablation tests to evaluate the effect that different dropout rates have upon the classification performance of our neural networks. We also perform a comparison with the results obtained using several traditional machine learning classifiers to assess the classification robustness.

We also suggest that by utilising anonymised, unidentifiable, pose-based features, we make the likelihood of collaborative working arrangements in the healthcare domain a more viable possibility. By ensuring that anonymised features are used in the classification process, our hope is that this approach has the potential to enhance deep learning frameworks, such as the one proposed here, through an ever greater abundance of data availability. As such, we make our code, the extracted features dataset and annotated labels publicly available.

II. RELATED WORKS

In this section, we provide an overview of the GMA by looking at the origins of the test and how it is currently applied in clinical practice. We discuss several studies which have attempted to automate the GMA using different methods and briefly contextualise these studies in relation to our proposed work. We also examine some state-of-the-art computer vision pose estimation techniques and observe how these might be implemented in our proposed diagnostic pipeline.

A. THE GENERAL MOVEMENTS ASSESSMENT

General movements (GMs) are spontaneous movements which are present from early fetal life through to approximately twenty weeks post-term. GMs engage the whole body in a diverse range of arm, leg, neck, and trunk movements which vary in intensity over time. In a typically developing infant, GMs show complexity, variability, frequency, and have sufficient duration to be observed properly [12]. Prechtl's 'Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants' [10] is the foundation of the GMA and explores the specific makeup of infant GMs in detail. Prechtl suggests that, in the case of infants with an impaired nervous system, GMs lose their complex and variable character, becoming less fluid and smooth.

Additionally, it is proposed that these infants show a lack of 'fidgety movements'. The presence of these abnormal GM patterns is seen as a strong predictor that the infant will go on to develop cerebral palsy. Zlatanovic *et al.* [44] emphasize the importance of early diagnosis of cerebral palsy, suggesting that it enables neurodevelopmental treatment, which can contribute to improved motor-function ability at a later age, due to the ''brain plasticity'' found in developing infants.

B. AUTOMATING THE GMA

Whilst the GMA has proven to be an accurate, non-invasive and non-intrusive diagnostic tool, it requires a significant investment of both time and resources to train an assessor. As such, several studies have been carried out which attempt to assess the viability of automated GMA. One of the earliest examples of this is a preliminary study by Adde *et al.* [1]. In this work, they developed a method which made use of per frame background subtraction. By removing the background they created a simple representation which allowed for the identification and subsequent calculation of the difference between two frames in a video sequence. A point value per pixel of 0 or 1 was then assigned to represent the presence of movement.

Following on from this, Stahl *et al.* [36] produced a method using optical flow which predicted cerebral palsy based upon statistical pattern recognition of the infant's spontaneous movements. They incorporated Wavelet frequency decomposition analysis to determine the time dependent trajectory signals found in the optical flow data.

Similarly Orlandi *et al.* [28] utilised large displacement optical flow (LDOF) to track infant movements and obtain velocities. They calculated the displacement of each pixel over 10 frames before extracting features for classification. The extracted features were used in a binary classification to determine normal or abnormal GMs using several classifiers.

In [20], a similar LDOF model is used to track infant movements through a pixelwise representation. The centroid of motion, rather than the centre of mass or anticipated joint position, of these tracked movements is manually annotated and fed into a classification pipeline to determine the likelihood of cerebral palsy based upon the proportion of CP risk-related movements. This approach focuses upon a statistical analysis of the data rather than a predefined set of rules governing the classification.

C. POSE ESTIMATION TECHNIQUES

Due to the limitations inherent in traditional optical flow based methods researchers have recently started to evaluate the effectiveness of pose-based assessment. The automated estimation of human pose from 2D images is an active research area, with several significant recent contributions [6], [13], [17], [37]. With the continued progression in deep learning techniques, various robust frameworks have been proposed which can accurately estimate human poses from 2D images. One of the most widely known methods is OpenPose by Cao *et al.* [6], who present an approach to detect the 2D pose of multiple people from a single RGB image. This framework produces an output which provides both the joint positions and orientation of human limbs based upon a pre-determined set of keypoints. In our pilot study [26] we made use of the OpenPose framework to examine the viability of a pose-based approach. We were subsequently able to establish a robust set of pose-based features which could be used for classification by several different traditional machine learning classifiers.

Using a comparable pose-based approach, Chambers *et al.* [7] developed a framework to extract posture, kinematic variables, complexity and symmetry for further analysis. They suggest that combinations of the extracted features are indicative of heightened neuromotor risk.

Similarly, Moccia *et al.* [27] proposed a framework for limb-pose estimation of infants from depth images. They attempted to exploit spatio-temporal features in an effort to improve pose estimation performance. By using a detection and a regression convolutional neural network (CNN) they performed limb-pose estimation, they then used 3D convolution to encode connectivity in the temporal domain.

D. HISTOGRAMS FOR HUMAN ACTION RECOGNITION

Histogram-based approaches have seen wide use in several fields for a number of years, and have been found to perform well in visual recognition tasks. Histogram-based approaches, such as [9], have been successfully implemented in human action recognition tasks by condensing data into a lower dimensional range whilst also retaining the most useful information, providing a full but manageable impression of the associated data. The combined use of different kinds of histogram features [38] has improved action recognition accuracy significantly [23]. 3D histogram features are also proposed on 3D human video, with the purpose of aiding 2D recognition [41].

RGBD cameras such as Kinect provide estimated 3D joint positions using a random forest [34], which can be further enhanced by introducing human prior knowledge [42], allowing it to be used effectively for motion monitoring [30]. In [39] a histogram-based method was used, in conjunction with the joint positions extracted from Kinect depth maps, to undertake classification of human actions into one of ten indoor activities. This approach allows for the generation of feature descriptors, which can be used to examine the distribution of both the orientation and displacement of each of the joints over a period of time. Additionally, this method bypasses the need to solve the time misalignment and variations in speed between two frames, as well as being robust to differing video durations.

We prefer the use of RGB over RGBD when it comes to capturing infants' movement for two reasons, firstly RGB video is much more accessible as it requires no specialist equipment (a camera-phone is sufficient), and secondly, RGBD requires the emission of infrared light, which may have a health impact upon infants. Inspired by the successes of [39], we propose a technique to produce histogram

representations of infants movements from 2D RGB footage, and a deep learning algorithm for action classification.

E. DEEP LEARNING METHODS

Recently, researchers have been successfully applying deep learning frameworks to the task of human action recognition [22], [24], [43]. These approaches make use of large datasets to train deep learning models capable of achieving state-of-the-art classification accuracy. Several studies have attempted to make use of the general improvements to accuracy provided by deep learning by applying deep learning frameworks to similar movement related diagnostic activities [8], [21], [35]. Whilst the results are promising, the holistic application of deep learning in the healthcare domain faces several challenges, most notably the large amount of data required for suitable results, and the problem of understandable AI. Understanding how a framework arrives at a decision is particularly important in the healthcare domain, and this is often very difficult, if not impossible to do with an end-to-end deep learning framework as deep features are typically incomprehensible for human perception. With this in mind, our proposed deep learning framework acts simply to classify the hand-crafted features generated using our previous method.

III. METHODOLOGY

In this section, we first detail the infant dataset creation process. We then explain how we extract pose-based features from the video footage, followed by the deep learning algorithms we propose for classification.

A. DATASET

Given the sensitive nature of the video data required for the GMA, a significant challenge facing researchers attempting to automate the process is the availability of publicly accessible datasets. Since human pose estimation frameworks are almost exclusively trained and tested using images of adults, a dataset consisting of images of infants for research purposes can understandably be difficult to obtain. In an effort to help researchers in this area the Moving INfants In RGB-D (MINI-RGBD) [18] synthetic dataset was produced using the Skinned Multi-Infant Linear (SMIL) [19] model and made publicly available. The MINI-RGBD dataset maps real-world infant movements to a synthetic SMIL 3D model in order to generate anonymised, and subsequently shareable footage. As such, this paper makes use of the MINI-RGBD dataset, which, at the time of writing, consists of twelve different sequences. Each of the twelve video sequences was analysed by an experienced GMs assessor. The assessor classified the videos into one of two categories; 1) those who demonstrate movements indicative of typically developing infants (Normal); and 2) those who demonstrate some movements that may be of concern to clinicians (Abnormal).

B. POSE ESTIMATION AND DATA PRE-PROCESSING

In this paper, the OpenPose framework [6] is used to extract the 2D poses from each of the twelve videos in the MINI-RGBD dataset. Each returned pose is represented by the 2D (x and y) coordinates of 18 landmarks on the body, which include 14 joints on the body and 4 facial landmarks. In addition to the 2D coordinates a confidence score for each joint is also included in the OpenPose output. In our experiments we only make use of the 14 body joints for feature extraction, we decided that this was a suitable approach because the facial landmarks were not as reliable as the body landmarks due to self occlusion, and analysis of the facial landmarks did not play as key a role in the GMA on this synthetic dataset.

The OpenPose output data is then pre-processed prior to the generation of features for classification. The first stage in our pre-processing pipeline is to remove any anomalous joint positions caused by occlusion or inaccuracies in OpenPose's joint assignment process. To do this we use the confidence score as a threshold by which we can judge the accuracy of OpenPose's predicted joint position. We calculate the average confidence score per joint across each video sequence and subtract an additional 10% from this. In any frame where a joint's confidence score falls below this threshold, we interpolate between the neighbouring frames with confidence scores higher than the threshold using modified Akima interpolation [3]. Given a set of control points $X = [x_0, x_1, \ldots, x_n]$ where n is the number of points, the slope δ_i on interval between x_i and x_{i+1} can be determined. The derivative d_i at the sample point x_i, which will be used for modified Akima interpolation, can be calculated by:

$$d_i = \frac{w_1}{w_1 + w_2}\delta_{i-1} + \frac{w_2}{w_1 + w_2}\delta_i \tag{1}$$

and the weights w_1 and w_2 are determined by:

$$w_1 = |\delta_{i+1} - \delta_i| + \frac{|\delta_{i+1} + \delta_i|}{2} \tag{2}$$

$$w_2 = |\delta_{i-1} - \delta_{i-2}| + \frac{|\delta_{i-1} + \delta_{i-2}|}{2}. \tag{3}$$

Applying this approach means that frames in which joint confidence scores are particularly low are smoothed to create a more reliable movement between frames of higher confidence on a per joint basis.

In order to further normalise the data and ensure that the orientation and displacement are comparable between videos we calculate a point mid-way between the hip joints (Figure 1, Joints 9 and 12), which we refer to as the root. We then calculate the mid-line which runs from the neck joint (Figure 1, Joint 2) to the root. Whilst retaining their relative distance from one another, all joints are then re-positioned so that the root is centred on point 0,0 and the calculated mid-line is aligned with the Y-Axis. This re-positioning and rotation is carried out on all frames and on all sequences.

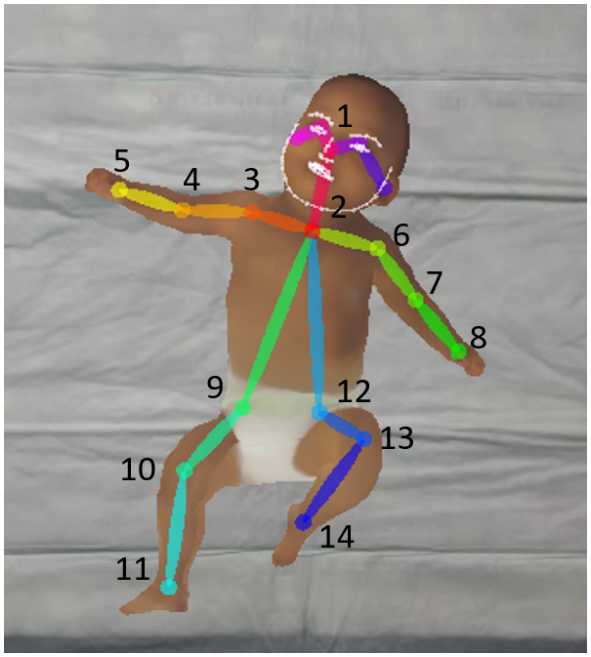


FIGURE 1. The OpenPose output skeleton and associated joint reference numbers, overlaid on an example input RGB image.

C. POSE-BASED FEATURE SETS

In our earlier work we proposed two new pose-based feature sets, Histograms of Joint Orientation 2D (HOJO2D) and Histograms of Joint Displacement 2D (HOJD2D). These feature sets consisted of histogram representations which described different aspects of the extracted pose based features. Building upon the successful implementation of these features we repeat our method to extract the same feature sets from the normalised, pre-processed data. Details of the generation of these feature sets are as follows:

1) HISTOGRAMS OF JOINT ORIENTATION 2D (HOJO2D)

In this representation the 2D space is segmented into n bins that denote the prevalent angle of joint orientation. The joint orientation is computed by calculating the alignment of the bone connecting a joint and parent joint:

$$bone = j_i - j_{i-parent}. \tag{4}$$

where j_i and $j_{i-parent}$ are the vectors containing the 2D coordinates of the i-th joint and its parent joint. We manually select the joint range to extract part specific information before a suitable bin is assigned for each joint per frame. As a result, the pose is represented by an n bin histogram of normalised data.

2) HISTOGRAMS OF JOINT DISPLACEMENT 2D (HOJD2D)

In this representation the displacement of each joint is extracted and recorded every five frames. The displacements are then associated with a relevant bin, each of which represents a regular incremental increase. Again, a range of joints is selected manually for part-based analysis. In this way the displacement can be represented by an n bin histogram of normalised data.

Fused feature sets are then exported for further evaluation in our classification experiments using the five separate deep learning architectures discussed in Section III-D.

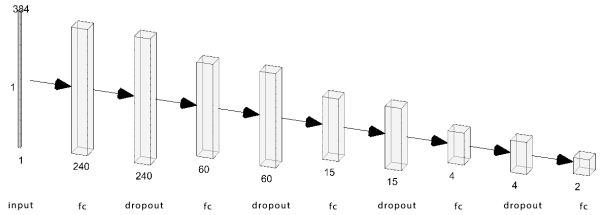


FIGURE 2. The proposed *FCNet* network architecture consisting of fully connected (fc) layers and dropout layers.

D. DEEP LEARNING FRAMEWORKS

In this section, we explain the neural network architectures proposed for classifying the pose-based features presented in Section III-C. In particular, 3 types of network architectures are proposed. We first introduce a fully connected network architecture in Section III-D.1 which serves as a basic classification framework. We further propose 1D (Section III-D.2) and 2D (Section III-D.3) convolutional neural network architectures.

1) FULLY CONNECTED DEEP NETWORKS

Fully connected deep network architectures are considered a generic framework for handling different problems since they are robust to different kinds of inputs (such as text, extracted features, images, videos, etc). Our proposed fully connected network architecture (Figure 2), namely *FCNet*, is designed with gradually decreasing layer sizes. The input of the network is a 1D vector of the histogram-based features. The output of the last fully connected layer is fed into a softmax layer for classification. To reduce the negative impact of overfitting, we have constructed a system where each fully connected layer is followed by a dropout layer. We evaluate the classification accuracy with different dropout rate settings and the results are presented in an ablation study in Section IV-C.

2) 1D CONVOLUTIONAL NEURAL NETWORKS

In the proposed pose-based features, the neighboring values are actually capturing similar body postures (i.e. with body part orientation in HOJO2D) and movements (i.e. with body part displacement in HOJD2D). To exploit the spatial information from the features, we propose two 1D convolutional neural network architectures (Figure 3 and 4), namely *Conv1DNet-1* and *Conv1DNet-2*, to learn the deep representation for better performance. Due to the relatively low dimensionality of the input feature vector, both of the proposed architectures contain two 1D convolution layers. To further improve the performance, each 1D convolution layer is followed by a max pooling layer to down-sample the output, further feeding into a dropout layer to avoid

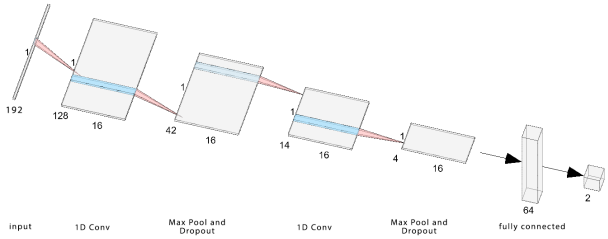


FIGURE 3. The proposed *Conv1DNet-1* network architecture which consists of 1D convolution, max pooling and dropout layers.

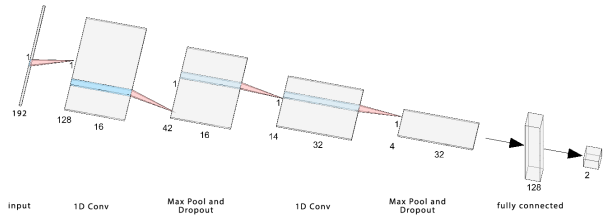


FIGURE 4. The proposed *Conv1DNet-2* network architecture which consists of 1D convolution, max pooling and dropout layers. Note the gradually increasing output channel sizes.

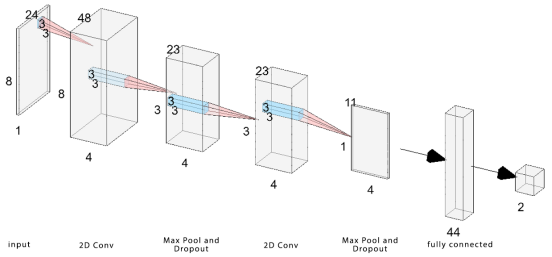


FIGURE 5. The proposed *Conv2DNet-1* network architecture which consists of 2D convolution, max pooling and dropout layers.

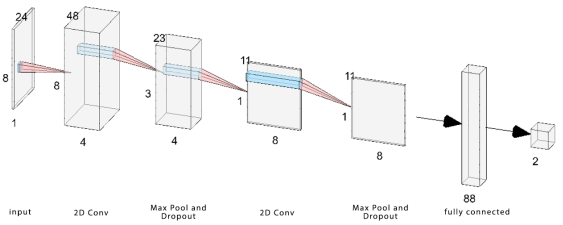


FIGURE 6. The proposed *Conv2DNet-2* network architecture which consists of 2D convolution, max pooling and dropout layers. Note the gradually increasing output channel sizes.

overfitting. Similar to *FCNet*, the input of our network is a 1D vector of the histogram-based features. The output of the last dropout layer is flattened into a 1D vector and the dimensionality is reduced by a fully connected layer before feeding into a softmax layer for classification. All the convolutional layers are using the same set of settings with $kernel\_size = 3$ and $stride = 3$. For the max pooling layers, $kernel\_size = 3$ and $stride = 3$ are used.

The main difference between the two networks is that *Conv1DNet-1* (Figure 3) has a constant output channel size while *Conv1DNet-2* (Figure 3) increases the output channel sizes gradually. We evaluate the difference in performance between these two architectures in Section IV-B.

3) 2D CONVOLUTIONAL NEURAL NETWORKS

To further exploit the spatial information among different body parts in the motion, we further propose two 2D convolutional neural network architectures. Recall that the limb-level and fused features are created by appending the histogram features of individual body parts resulting in a long 1D vector:

$$hist_{combined1D} = [hist_{part_1}, hist_{part_2}, \ldots, hist_{part_n}] \quad (5)$$

where $hist_{combined1D}$ is the final feature vector concatenated from the histogram features extracted from individual body parts and n is the number of body parts included in this feature.

To learn the spatial correlation within the 2D convolutional neural network, the input vector has to be converted into a 2D matrix shape. This is done by reshaping the 1D feature vector to a 2D matrix with each row containing the histogram features extracted from a single body part:

$$hist_{combined2D} = \begin{bmatrix} hist_{part_1} \\ hist_{part_2} \\ \vdots \\ hist_{part_n} \end{bmatrix} \quad (6)$$

The two 2D convolutional neural network architectures (Figure 5 and 6) we propose, namely *Conv2DNet-1* and *Conv2DNet-2*, share a common design with two 2D convolution layers. Similar to the proposed 1D convolutional neural networks, each 2D convolution layer is followed by a max pooling layer to down-sample the output and further feed into a dropout layer to avoid overfitting. The output of the last dropout layer is flattened into a 1D vector and the dimensionality is reduced by a fully connected layer before feeding into a softmax layer for classification. All the convolutional layers are using the same set of settings with $kernel\_size = 3$ and $stride = 1$. For the max pooling layers, $kernel\_size = 3$ and $stride = 2$ are used.

The main difference between the two networks is that *Conv2DNet-1* (Figure 5) has a constant output channel size while *Conv2DNet-2* (Figure 5) increases the output channel sizes gradually. We evaluate the difference in performance between these two architectures in Section IV-B.

IV. EXPERIMENTAL RESULTS

In this section, we present the experimental results in this study to evaluate the performance of the proposed motion classification framework with different deep neural network architectures. We first compare the classification accuracy (Section IV-B) obtained from the proposed methods and baseline approaches as in [26]. Next, we justify the selection of the hyper-parameters in the proposed network architectures by conducting a series of ablation studies (Section IV-C).

A. EXPERIMENTAL SETTINGS AND IMPLEMENTATION DETAILS

The MINI-RGBD dataset [19] is used in all experiments, we collected the annotations of all videos from an experienced GMs assessor in our pilot study [26]. There are

TABLE 1. HOJO2D feature set: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods.

Histograms of Joint Orientation 2D (HOJO2D)							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**100.00%**	75.00%	**100.00%**	75.00%	41.67%	75.00%	77.78%
SVM	66.67%	66.67%	66.67%	66.70%	66.70%	66.70%	66.70%
Tree	75.00%	0.00%	75.00%	75.00%	33.33%	75.00%	55.56%
kNN (k=1)	83.33%	25.00%	58.33%	**83.33%**	8.33%	33.33%	48.61%
kNN (k=3)	83.33%	25.00%	41.67%	**83.33%**	41.67%	58.33%	55.56%
Ensemble	75.00%	25.00%	75.00%	75.00%	8.33%	75.00%	55.56%
FCNet	83.33%	**83.33%**	83.33%	**83.33%**	**83.33%**	**83.33%**	**83.33%**
Conv1D-1	83.33%	75.00%	83.33%	**83.33%**	75.00%	75.00%	79.17%
Conv1D-2	83.33%	**83.33%**	75.00%	75.00%	**83.33%**	**83.33%**	80.55%
Conv2D-1	75.00%	**83.33%**	75.00%	**83.33%**	**83.33%**	75.00%	79.17%
Conv2D-2	83.33%	**83.33%**	83.33%	**83.33%**	75.00%	**83.33%**	81.94%

TABLE 2. HOJD2D feature set: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods.

Histograms of Joint Displacement 2D (HOJD2D)							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**91.67%**	41.67%	50.00%	**100.00%**	50.00%	75.00%	68.06%
SVM	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%
Tree	83.30%	50.00%	83.33%	66.67%	41.67%	66.67%	65.28%
kNN (k=1)	**91.67%**	50.00%	75.00%	**100.00%**	41.67%	75.00%	72.22%
kNN (k=3)	66.67%	41.67%	83.33%	58.33%	58.33%	66.67%	62.50%
Ensemble	83.33%	58.33%	83.33%	66.67%	58.33%	66.67%	69.44%
FCNet	83.33%	**83.33%**	**91.67%**	83.33%	**83.33%**	**91.67%**	**86.11%**
Conv1D-1	83.33%	75.00%	83.33%	83.33%	**83.33%**	83.33%	81.94%
Conv1D-2	83.33%	75.00%	83.33%	83.33%	**83.33%**	83.33%	81.94%
Conv2D-1	75.00%	**83.33%**	83.33%	91.67%	**83.33%**	75.00%	81.94%
Conv2D-2	83.33%	**83.33%**	83.33%	75.00%	75.00%	83.33%	80.55%

12 videos in the dataset and we employ a leave-one-out cross-validation approach in all of the experiments in this study. The averaged classification accuracy is then reported.

The proposed deep neural network architectures are implemented in the PyTorch framework. All experiments were run on a desktop computer with a single NVIDIA TITAN Xp graphics card. Additional parameters such as $epochs = 4000$, $learningrate = 0.0005$ and $batchsize = 3$ are used in all tests.

B. CLASSIFICATION ACCURACY - COMPARING WITH BASELINE APPROACHES

In this section, we compare the performance of our proposed deep learning frameworks with the baseline approaches. We obtained the classification accuracy of all methods in 3 types of input features: 1) HOJO2D, 2) HOJD2D, and 3) fusing (i.e. concatenating) HOJO2D and HOJD2D. Due to the random initialization of our newly proposed deep learning frameworks, the performance of the classifier may vary in different trials. In this section, we report the best performance of classifiers.

1) HOJO2D

The results are presented in Table 1. In general, the newly proposed deep learning classification frameworks perform better, as most of the highest accuracies (highlighted in bold) are obtained using our methods. In particular, *FCNet* performs well consistently achieving 83.33% across all of the different features. This highlights the generality of the fully connected neural network. The proposed *Conv2D-2* and *Conv1D-2* with gradually increasing output channel size in the convolutional layers also demonstrated high performance with most of the features having the same classification accuracy as *FCNet*. Accuracy obtained using *Conv1D-1* and *Conv2D-1* are lower than the other proposed frameworks, but they are more consistent and robust than the baseline approaches. For the baselines, the results are highly inconsistent. While some of the classification accuracies are high (such as the 8-bin Arms and 8-bin Limbs features with LDA), classifying some other features can result in very low accuracy (such as Legs with 16 bins). In summary, the results demonstrated the high performance and robustness of the proposed deep learning frameworks.

2) HOJD2D

The results are presented in Table 2. Again, the newly proposed deep learning frameworks are more robust and performed more consistently. Whilst kNN (k=1) and LDA achieved some of the best accuracies with 91.67% on the 8-bin Arms feature and 100% on 16-bin Arms feature, the accuracy on other features are much lower (such as 50.00% and 41.67% on both Legs features). For our approaches, *FCNet* performed well and obtained 91.67% accuracy in Limbs features whilst achieving 83.33% in the

TABLE 3. Fusing the HOJO2D and HOJD2D feature sets: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods.

Fused features - HOJO2D + HOJD2D							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**100.00%**	66.67%	**100.00%**	**91.67%**	58.33%	83.33%	83.33%
SVM	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%
Decision Tree	75.00%	50.00%	75.00%	75.00%	25.00%	75.00%	62.50%
kNN (k=1)	91.67%	33.33%	83.33%	**91.67%**	50.00%	75.00%	70.83%
kNN (k=3)	91.67%	33.33%	83.33%	66.67%	58.33%	66.67%	66.67%
Ensemble	75.00%	58.33%	75.00%	75.00%	33.33%	75.00%	65.28%
FCNet	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
Conv1D-1	83.33%	**91.67%**	83.33%	83.33%	83.33%	83.33%	84.72%
Conv1D-2	83.33%	**91.67%**	91.67%	**91.67%**	**91.67%**	**91.67%**	**90.28%**
Conv2D-1	83.33%	83.33%	83.33%	75.00%	75.00%	75.00%	79.17%
Conv2D-2	83.33%	75.00%	83.33%	83.33%	75.00%	83.33%	80.55%

rest of the features. The other deep learning frameworks are performing in a predictable manner with a smaller range in classification accuracy variations between 75.00% to 91.67%. This again highlights the robustness of the proposed deep learning frameworks.

3) FUSED FEATURES - HOJO2D + HOJD2D

The results are presented in Table 3. Since the input feature size is doubled in this experiment, deep learning frameworks generally demonstrated a large advantage in processing features in higher dimensionality. In particular, *Conv1D-2* achieved an excellent performance by having 91.67% classification accuracy in 5 out of 6 feature types. Whilst the LDA method obtained excellent classification accuracy on the 8-bin Arms and Limbs features, the results obtained in other features are significantly lower highlighting the inconsistency of the baseline methods, with accuracies ranging from 58.33% to 100%. *Conv1D-1* demonstrated a solid performance in achieving 91.67% in 8-bin Legs feature and 83.33% in the rest of the features. *FCNet* showed a robust performance again by obtaining 83.33% classification accuracy in all feature types. For the 2D convolutional neural networks *Conv2D-1* and *Conv2D-2*, the performance is once again consistent, with a small range of accuracy from 75.00% to 83.33%. We also observe that, in most cases, applying feature fusion achieved a better classification performance than the individual histogram features.

In summary, the experimental results on different feature types highlight the performance gain in both accuracy and robustness with the use of the proposed deep learning frameworks over the baseline approaches. The results also show that *FCNet* performed in a highly predictable manner with a relatively simple network architecture.

We also observe that, in general, the 16-bin variant is better for the proposed deep methods whilst the 8-bin version is better in the non-deep baseline methods. This is due to the fact that deep networks can handle features in higher dimensionality than non-deep methods. This also suggests that the 16-bin features are more discriminative, particularly in the case of joint displacement, where the magnitude of the joint displacement appears to be more consistent for classification than the joint orientation.

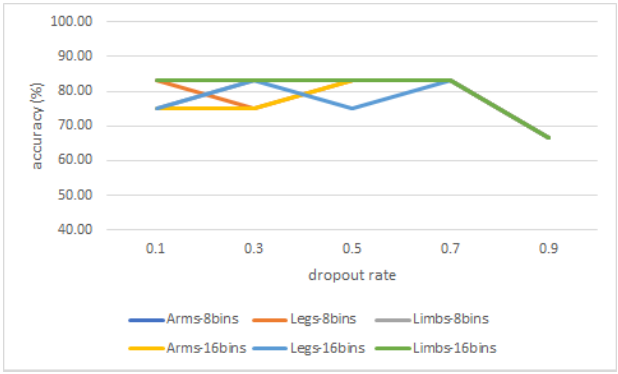


FIGURE 7. *FCNet* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

We also note that when the dimiensionality of input features becomes higher, the benefits of using convolutional neural networks can be observed, as seen when using *Conv1D-1* and *Conv2D-1* to classify fused features. This can be explained by the abstraction power of the convolutional layers in the network. We believe the performance gain of 2D convolutional networks will be even greater when the input features have even higher dimensionality (e.g. by incorporating time-series movement data).

C. ABLATION STUDIES

We conducted an ablation study to investigate the impact of the hyper-parameters on the classification performance. Since we have already compared the effect different layer sizes have on the proposed 1D (i.e. *Conv1D-1*, and *Conv1D-2*) and 2D (i.e. *Conv2D-1*, and *Conv2D-2*) network architectures in Section IV-B, in this section, we focus on another hyper-parameter, namely the dropout rate. We picked the fused features setting in this ablation study, while training the networks with different dropout rates (i.e. 0.1, 0.3, 0.5, 0.7 and 0.9). The results are plotted in Figures 7 to 11.

The results show that most of the different dropout settings result in similar classification accuracy. In most cases,

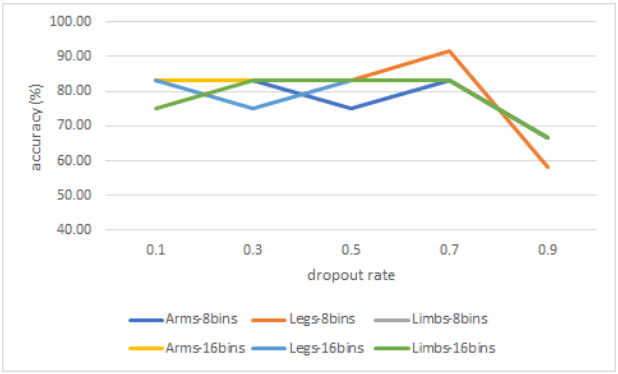


FIGURE 8. *Conv1D-1* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

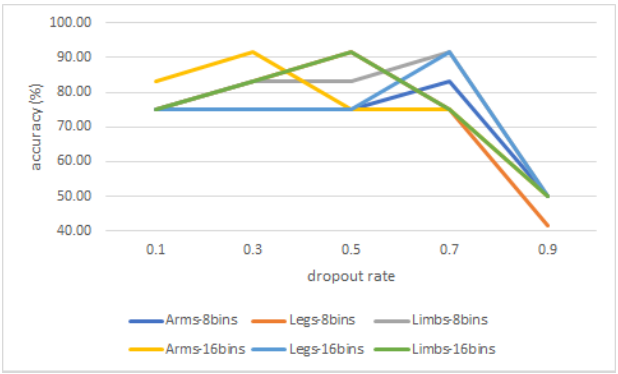


FIGURE 9. *Conv1D-2* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

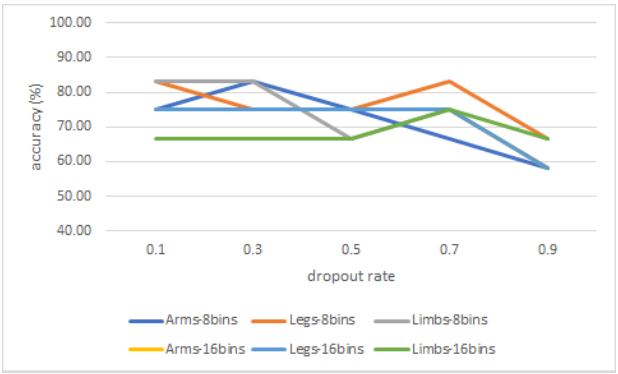


FIGURE 10. *Conv2D-1* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

the best performance occurs when the dropout rate equals 0.5 or 0.7, while some good performance can be obtained when the dropout rate equals 0.3. For the more extreme values we see a drop in accuracy, with 0.1 being unlikely to produce the best performance, and 0.9 being likely to produce the worst performance.

In summary, while there are some variations in the classification accuracy across different dropout rate settings, the range of accuracy is relatively small when compared with

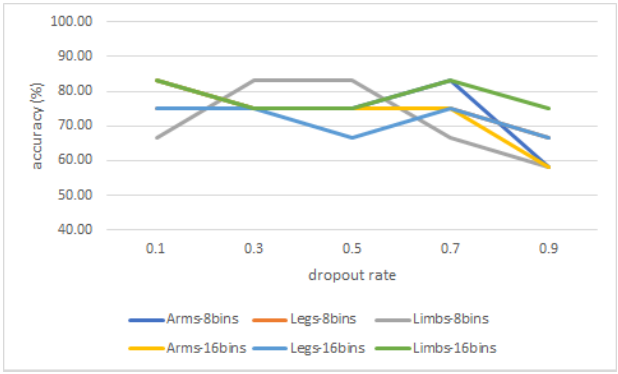


FIGURE 11. *Conv2D-2* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

the inconsistent performance from baseline approaches presented in Section IV-B. This also highlights that our proposed deep learning frameworks are less sensitive to changes in hyper-parameters.

V. CONCLUSION

In this work, we proposed five deep learning based frameworks to classify infant body movement based upon the pose-based features in our pilot study [26]. We further extend and enhance the feature extraction and pre-processing pipeline to facilitate the classification task. The proposed frameworks are evaluated and compared with the baseline approaches. Experimental results show that the proposed fully connected neural network *FCNet* performed robustly across different feature sets. Furthermore, the proposed 1D convolutional neural network architectures demonstrated an excellent performance in handling features in higher dimensionality. Finally, we conducted an ablation study to justify the selection of the hyper-parameters in the proposed frameworks. To stimulate the research in this area, the annotated dataset and the implementation of the deep learning frameworks will be available to the public as an open-source project.

Since the video sequences used in this paper are synthetic, the appearance of the images used as an input for the OpenPose framework differ slightly from that of real-world video data. As such, evaluating the frameworks using video data captured from patients is one of our anticipated future directions. Also, given that the quantity of video sequences in the MINI-RGBD dataset is relatively small, we hope to extend this work by classifying a larger dataset. We have been working closely with local hospitals in an effort to produce a real-world dataset for future evaluation. The assessment of our system using a larger dataset will also allow us to undertake additional quantitative analysis and verification, enabling calculation of supplementary evaluation metrics such as sensitivity, specificity, statistical significance levels, and permutation-based p-values, to measure the competence of the proposed classifiers.

In the future, we also will further investigate the feasibility of modelling the temporal information of the input posture sequence by incorporating Recurrent Neural Networks (RNN) in the proposed framework. We also intend to compare our method with some commonly implemented handcrafted feature extraction methods and fully non-handcrafted feature extraction methods as a means of establishing the accuracy, robustness and comparative interpretability of our proposed method. Our future work will also incorporate comparisons with other methods by re-implementing proposed approaches for use on our real-world dataset, in this way more accurate assessments can be made than by simply comparing with reported accuracies. Finally, another interesting future direction could be to evaluate the proposed network architectures with other advanced pose-based features [31].

ACKNOWLEDGMENT

The authors wish to gratefully acknowledge the support of NVIDIA Corporation, who donated the Titan Xp GPU used for this research. The project was supported in part by the Royal Society (Ref: IES\R1\191147 and IES\R2\181024) and NIHR Fellowship (Ref: ICA-CDRF-2018-04-ST2-020).

REFERENCES

[1] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, and R. Støen, "Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study," *Develop. Med. Child Neurol.*, vol. 52, no. 8, pp. 773–778, 2010.

[2] J. K. Aggarwal and L. Xia, "Human activity recognition from 3D data: A review," *Pattern Recognit. Lett.*, vol. 48, pp. 70–80, Oct. 2014.

[3] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, Oct. 1970.

[4] *What is the General Movements Assessment?* Cerebral Palsy Alliance, Allambie Heights, Australia. [Online]. Available: https://cerebralpalsy.org.au/our-research/about-cerebral-palsy/what-is-cerebral-palsy/signs-and-symptoms-of-cp/general-movements-assessment/

[5] A. P. Basu and G. Clowry, "Improving outcomes in cerebral palsy with early intervention: New translational approaches," *Frontiers Neurol.*, vol. 6, p. 24, Feb. 2015.

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1302–1310.

[7] C. Chambers, N. Seethapathi, R. Saluja, H. Loeb, S. Pierce, D. Bogen, L. Prosser, M. J. Johnson, and K. P. Kording, "Computer vision to automatically assess infant neuromotor risk," *BioRxiv*, Sep. 2019, Art. no. 756262.

[8] R. Cunningham, M. B. Sánchez, P. B. Butler, M. J. Southgate, and I. D. Loram, "Fully automated image-based estimation of postural point-features in children with cerebral palsy using deep learning," *Roy. Soc. Open Sci.*, vol. 6, no. 11, Nov. 2019, Art. no. 191011.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[10] C. Einspieler and F. R. H. Prechtl, *Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants*. London, U.K.: Mac Keith Press, 2004.

[11] C. Einspieler and H. F. R. Prechtl, "Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system," *Mental Retardation Developmental Disabilities Res. Rev.*, vol. 11, no. 1, pp. 61–67, Feb. 2005.

[12] C. Einspieler, H. F. R. Prechtl, F. Ferrari, G. Cioni, and A. F. Bos, "The qualitative assessment of general movements in preterm, term and young infants—Review of the methodology," *Early Hum. Develop.*, vol. 50, no. 1, pp. 47–60, Nov. 1997.

[13] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2353–2362.

[14] T. Fjørtoft, C. Einspieler, L. Adde, and L. I. Strand, "Inter-observer reliability of the 'assessment of motor repertoire—3 to 5 months' based on video recordings of infants," *Early Hum. Develop.*, vol. 85, no. 5, pp. 297–302, 2009.

[15] *NICE Seeks to Improve Diagnosis and Treatment of Cerebral Palsy*, National Institute for Health and Care Excellence, London, U.K., Jan. 2017.

[16] Y. Gao, Y. Long, Y. Guan, A. Basu, J. Baggaley, and T. Ploetz, "Towards reliable, automated general movement assessment for perinatal stroke screening in infants using wearable accelerometers," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 1, pp. 1–22, Mar. 2019.

[17] R. A. Guler, N. Neverova, and I. Kokkinos, "DensePose: Dense human pose estimation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7297–7306.

[18] N. Hesse, C. Bodensteiner, M. Arens, G. U. Hofmann, R. Weinberger, and A. S. Schroeder, "Computer vision for medical infant motion analysis: State of the art and RGB-D data set," in *Proc. Comput. Vis.—ECCV Workshops*, L. Leal-Taixé and S. Roth. Cham, Switzerland: Springer, 2019, pp. 32–49, doi: 10.1007/978-3-030-11024-6_3.

[19] N. Hesse, S. Pujades, J. Romero, J. M. Black, C. Bodensteiner, M. Arens, G. U. Hofmann, U. Tacke, M. Hadders-Algra, R. Weinberger, W. Müller-Felber, A. S. Schroeder, "Learning an infant body model from RGB-D data for accurate full body motion analysis," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, F. A. Frangi, A. J. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger. Cham, Switzerland: Springer, 2018, pp. 792–800, doi: 10.1007/978-3-030-00928-1_89.

[20] E. A. F. Ihlen, R. Støen, L. Boswell, R.-A. de Regnier, T. Fjørtoft, D. Gaebler-Spira, C. Labori, M. C. Loennecken, M. E. Msall, U. I. Möinichen, C. Peyton, M. D. Schreiber, I. E. Silberg, N. T. Songstad, R. T. Vågen, G. K. Øberg, and L. Adde, "Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study," *J. Clin. Med.*, vol. 9, no. 1, p. 5, 2020.

[21] M. Lempereur, F. Rousseau, O. Rémy-Néris, C. Pons, L. Houx, G. Quellec, and S. Brochard, "A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity," *J. Biomech.*, vol. 98, Jan. 2020, Art. no. 109490.

[22] J. Liu, H. Rahmani, N. Akhtar, and A. Mian, "Learning human pose models from synthesized data for robust RGB-D action recognition," *Int. J. Comput. Vis.*, vol. 127, no. 10, pp. 1545–1564, Oct. 2019.

[23] S.-L. Lo and A.-C. Tsoi, "Human action recognition: A dense trajectory and similarity constrained latent support vector machine approach," in *Proc. 2nd IAPR Asian Conf. Pattern Recognit.*, Nov. 2013, pp. 230–235.

[24] D. C. Luvizon, D. Picard, and H. Tabia, "2D/3D pose estimation and action recognition using multitask deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5137–5146.

[25] C. Marcroft, A. Khan, N. D. Embleton, M. Trenell, and T. Plötz, "Movement recognition technology as a method of assessing spontaneous general movements in high risk infants," *Frontiers Neurol.*, vol. 5, p. 284, Jan. 2015.

[26] K. D. McCay, E. S. L. Ho, C. Marcroft, and N. D. Embleton, "Establishing pose based features using histograms for the detection of abnormal infant movements," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5469–5472.

[27] S. Moccia, L. Migliorelli, V. Carnielli, and E. Frontoni, "Preterm infants' pose estimation with spatio-temporal features," *IEEE Trans. Biomed. Eng.*, to be published.

[28] S. Orlandi, K. Raghuram, C. R. Smith, D. Mansueto, P. Church, V. Shah, M. Luther, and T. Chau, "Detection of atypical and typical infant movements using computer-based video analysis," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 3598–3601.

[29] P. C. Panteliadis, *Cerebral Palsy: A Multidisciplinary Approach*, 3rd ed. Springer, 2018.

[30] P. Plantard, A. Müller, C. Pontonnier, G. Dumont, H. P. H. Shum, and F. Multon, "Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics?" *Int. J. Ind. Ergonom.*, vol. 61, pp. 71–80, Sep. 017.

[31] W. Rueangsirarak, J. Zhang, N. Aslam, E. S. L. Ho, and H. P. H. Shum, "Automatic musculoskeletal and neurological disorder diagnosis with relative joint displacement from human gait," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2387–2396, Dec. 2018.

[32] *NHS-Cerebral Palsy Overview*, National Health Service, London, U.K., Mar. 2015.
[33] A. H. Shevell and M. Shevell, "Doing the 'talk' disclosure of a diagnosis of cerebral palsy," *J. Child Neurol.*, vol. 28, no. 2, pp. 230–235, 2013.
[34] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, Jun. 2011, pp. 1297–1304.
[35] P. Shukla, T. Gupta, A. Saini, P. Singh, and R. Balasubramanian, "A deep learning frame-work for recognizing developmental disorders," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 705–714.
[36] A. Stahl, C. Schellewald, O. Stavdahl, O. M. Aamo, L. Adde, and H. Kirkerod, "An optical flow-based method to predict infantile cerebral palsy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 605–614, Jul. 2012.
[37] S. Suwajanakorn, N. Snavely, J. J. Tompson, and M. Norouzi, "Discovery of latent 3D keypoints via end-to-end geometric reasoning," *Adv. Neural Inf. Process. Syst.*, Dec. 2018, pp. 2059–2070.
[38] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," in *Proc. CVPR*, Jun. 2011, pp. 3169–3176.
[39] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 20–27.
[40] L. Yao, Y. Liu, and S. Huang, "Spatio-temporal information for human action recognition," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 39, Dec. 2016.
[41] J. Zhang, H. P. H. Shum, J. Han, and L. Shao, "Action recognition from arbitrary views using transferable dictionary learning," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4709–4723, Oct. 2018.
[42] L. Zhou, Z. Liu, H. Leung, and H. P. H. Shum, "Posture reconstruction using Kinect with a probabilistic model," in *Proc. 20th ACM Symp. Virtual Reality Softw. Technol. (VRST)*. New York, NY, USA: ACM, Nov. 2014, pp. 117–125.
[43] Y. Zhou, X. Sun, Z.-J. Zha, and W. Zeng, "MiCT: Mixed 3D/2D convolutional tube for human action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 449–458.
[44] D. Zlatanović, H. Čolović, V. Zivković, M. Kocić, A. Stanković, J. Vučić, N. Bojović, M. Raičević, M. Golubović, L. Dinić, and T. Stanković, "The importance of the Prechtl method for ultra-early prediction of neurological abnormalities in newborns and infants," *Acta Medica Medianae*, vol. 58, no. 3, pp. 111–115, Sep. 2019.

KEVIN D. MCCAY received the B.Sc. degree (Hons.) in computer animation and digital SFX and the M.A. degree in animation from Northumbria University, Newcastle upon Tyne, U.K., in 2015 and 2016, respectively, where he is currently pursuing the Ph.D. degree in computer science.

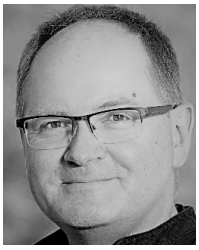
His research interests include human motion analysis, computer vision, machine learning, and the application of automated analysis within the healthcare domain.

EDMOND S. L. HO received the B.Sc. degree (Hons.) in computer science from Hong Kong Baptist University, in 2003, the M.Phil. degree in computer science from the City University of Hong Kong, in 2006, and the Ph.D. degree from The University of Edinburgh, in 2010.

He was a Research Assistant Professor with the Department of Computer Science, Hong Kong Baptist University. He is currently a Senior Lecturer with the Department of Computer and Information Sciences, Northumbria University, Newcastle, U.K. His research interests include computer graphics, computer vision, robotics, motion analysis, and machine learning.

HUBERT P. H. SHUM (Senior Member, IEEE) received the bachelor's and master's degrees from the City University of Hong Kong and the Ph.D. degree from the University of Edinburgh.

He was a Senior Lecturer with Northumbria University, a Lecturer with the University of Worcester, and a Postdoctoral Researcher with RIKEN, Japan. He is currently an Associate Professor of computer science with Northumbria University, U.K., and the Director of Research and Innovation of the Computer and Information Sciences Department. He has over 100 publications in the areas of computer graphics, computer vision, human motion analysis, and machine learning. He serves as an Associate Editor of *Computer Graphics Forum*.



GERHARD FEHRINGER is currently working as a Principal Lecturer with the Department of Computer and Information Sciences, Northumbria University, U.K. Prior to joining Northumbria University, he worked as a Systems Analyst at Procter and Gamble. He has research experience in the IoT, computer networks, and cyber security.

CLAIRE MARCROFT received the B.Sc. degree (Hons.) in physiotherapy from the University of Huddersfield, Huddersfield, U.K., in 2004.

She is currently a Neonatal Physiotherapist with the Newcastle upon Tyne Hospitals, NHS Foundation Trust/Newcastle University, and holds a National Institute of Health Research (NIHR) Integrated Clinical Academic Doctoral Research Fellowship.

Ms. Marcroft is a member of the Health Care and Professions Council (HCPC), the Chartered Society of Physiotherapy (CSP) and the Association of Paediatric Chartered Physiotherapists (APCP).

NICHOLAS D. EMBLETON received the B.Sc. degree in environmental studies and computer science from the University of East Anglia, Norwich, U.K., in 1984, and the M.B.B.S. degree (Hons.) and the M.D. degree (commendation) from Newcastle University, Newcastle upon Tyne, U.K., in 1990 and 2002, respectively.

He is a Consultant Neonatal Paediatrician and a Professor of neonatal medicine, having completed paediatric and neonatal training in U.K. and Vancouver, Canada. He helps to lead a broad portfolio of research coordinated by the Newcastle Neonatal Research Team, which includes large-scale NIHR nutrition trials, along with mechanistic microbiomic and metabolomic studies. He coordinates the Newcastle Preterm Birth Growth study that has tracked the growth and metabolic outcomes of children who were born preterm into late adolescence. He leads a series of qualitative studies exploring the experiences of parents who suffered a reproductive or neonatal loss. He has more than 140 peer-reviewed publications in addition to numerous educational articles and book chapters.

Dr. Embleton is an elected member of the ESPGHAN Committee of Nutrition, coordinates the U.K.-based neonatal nutrition network.