

# Evaluation of Infants with Spinal Muscular Atrophy Type-I Using Convolutional Neural Networks

Bilge Soran<sup>1</sup>(✉), Linda Lowes<sup>2</sup>, and Katherine M. Steele<sup>1</sup>

<sup>1</sup> Mechanical Engineering Department, University of Washington, Seattle, USA  
bilge@cs.washington.edu, kmsteele@uw.edu

<sup>2</sup> Clinical Therapies Department, Nationwide Children's Hospital, Columbus, USA  
linda.lowes@nationwidechildrens.org

**Abstract.** Spinal Muscular Atrophy is the most common genetic cause of infant death. Due to its severity, there is a need for methods for automated estimation of disease progression. In this paper we propose a Convolutional-Neural-Network (CNN) model to estimate disease progression during infants' natural behavior. With the proposed methodology, we were able to predict each child's score on current behavior-based clinical exams with an average per-subject error of 6.96 out of 72 points (<10 % difference), using 30-second videos in leave-one-subject-out-cross-validation setting. When simple statistics were used over 30-second video-segments to estimate a score for longer videos, we obtained an average error of 5.95 (~8 % error rate). By showing promising results on a small dataset ( $N = 70$ , 2-minute samples, which were handled as 1487, 30-second video segments), our methodology demonstrates that it is possible to benefit from CNNs on small datasets by proper design and data handling choices.

**Keywords:** Spinal muscular atrophy · Longitudinal analysis · Microsoft kinect · Convolutional neural networks · Regression

## 1 Introduction

Spinal Muscular Atrophy (SMA) is a neuromuscular disorder caused by a genetic defect in the SMN1 gene which affects the motor nerve cells in the spinal cord leading to progressive weakness. The incidence of SMA, which can be diagnosed with DNA tests, is approximately 14 out of every 100,000 live-born infants [10]. There is no cure for SMA and current treatments focus on supportive care, not an improvement in the disease.

There are four different types of SMA [35], which are classified genetically by the copy number of the SMN2 gene or clinically based on motor function capability and age of diagnosis [25]. SMA type-I is the most severe form and presents in infants younger than 6 months of age leading to severe muscle weakness, progressive respiratory insufficiency and typically death before the age of 2 years. Due to

the severity and rapid progression of this disorder, there is a need for new tools to track its progression. In this paper, we propose a method to evaluate infant movements in order to estimate disease progression in infants with SMA type-I. It is well known that clinical trials for infants with SMA type-I pose unique challenges due to the profound weakness, respiratory insufficiency and vulnerability to complications related to participation in trials, such as travel [31]. Clinical trials could be advanced by an outcome measure that is relevant to families and the FDA, reliably quantifies small changes, and minimizes stress on fragile infants with SMA and their families [6].

Until recently, disease progression in SMA was quantified by time until death or the need for 16 h of ventilator support. The current standard clinical evaluation is done based on the Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND) score, which was developed to evaluate motor function in infants with a variety of neuromuscular disorders. It evaluates a child’s ability to move his/her extremities, head and trunk by rating performance over 16 items using a 4-point scale [10]. A trained evaluator elicits movement from the child through a series of items such as encouraging the child to roll onto their stomach by pulling on the arm or leg. Additional items include supporting the child in sitting or in prone with the neck flexed and observing if the infant will attempt to lift their head. Unfortunately, these activities can be extremely taxing on fragile infants with compromised respiratory systems. The CHOP INTEND has been shown to track the decline in motor function over time in infants with SMA and is related to other electro-physiological indicators [4,7,8]; however, this evaluation is subjective and requires extensive training to be performed reliably. An affordable, easy-to-use, and accurate system for evaluating infant movement, assessing risk, and tracking disease progression would be a welcome addition to enhance patient care.

In this paper, we are proposing method pipelines for the automatic evaluation of infants with SMA, the most common genetic cause of infant death, using depth camera technology. We used 2-minute-long videos of awake supine infants to demonstrate that motion of the arms and legs can be used to longitudinally track disease progression of infants with SMA with comparison to current clinical standard, the CHOP INTEND. The proposed pipelines could also be applied to analysis of other pediatric neuromuscular disorders, such as cerebral palsy. *Note that we are estimating CHOP INTEND extremity scores from videos of infants during natural behavior, however the groundtruth CHOP INTEND extremity scores are given by functional assessments.* This distinction increases the challenge in estimating disease progression and comparing novel methods to current clinical standards, which are limited by dependence on the subjective opinion of the assessor [4], and are fatiguing to the infant as they require the child to attempt activities and be placed in positions (head lifting, prone suspension) that can increase the work of breathing.

## 2 Related Work

Prechtl *et al.* described a type of spontaneous movements in infants 10–15 weeks of age, called fidgety or general movements [23], and characterized by small amplitude circular movements of the neck, trunk and limbs. General Movement Assessments (GMA) involves observing and rating these movements in infants [13], which is a reliable, non-invasive predictor of neurological problems. Compared to traditional methods GMA is a cost-effective technique for prediction and analysis of neurological impairments, including cerebral palsy and autism [2, 22]. On individuals with impaired nervous systems, general movements lack complex patterns seen in typically-developing infants. GMA has been shown to predict neurological outcome at 2 years more precisely than standard neurological examination [5]. Based on this work, we expect movements of the limbs during early infancy to be important predictors and hypothesize that analyzing infants while they are lying awake on their back will enable us to quantify motor abilities.

Although promising, adoption of GMA in standard practice has been extremely slow. The methodology is qualitative and classifications are made based on subjective judgments. Costly professional training by the General Movements Trust is required to implement this assessment and enrollment is very limited with roughly 6 courses worldwide each year allowing around 50 people each. This is yet another indication that a low-cost alternative would be extremely useful. Although there exists previous work on analysis of other neurological disorders from videos of infants (e.g [24, 30]), to the best of our knowledge there exists *no* research on automated evaluation of patients with SMA.

Depth cameras and depth estimation methods has been used for analysis and prediction of movement disorders caused by other neurological impairments. A Kinect based upper extremity motion analysis system was used to determine the spectrum of reachable workspace in facioscapulohumeral muscular dystrophy [14] and in byamyotrophic lateral sclerosis [21]. A Parkinson patient’s movements was automatically analyzed by [27, 28] showed a system that could reliably detect voice, postural and Parkinson’s tremors. A vision system for physical rehabilitation at home was proposed by [3]. Illness stage of patients with Alzheimer’s disease were determined in [16]. A system for cognitive assessment and rehabilitation of individuals with body scheme dysfunctions and left-right confusion was proposed by [12]. Disease progression in MS patients by analyzing selected movements performed by patients was measured by [17]. For a detailed review on the impact of Kinect for understanding neurological impairments please see [15].

For many years, traditional features have been widely used in all different types of vision problems. However, manually-designing those features required a lot of engineering effort and many times the designed features may not represent the problem domain properly especially for complex domains. Convolutional Neural Networks (CNNs) have the ability to learn the features (both low level and high level) directly from input data, which has been successfully used to address many computer vision problems, with state-of-the-art results (e.g., [18, 34]). Especially large problems that require intensive processing power to train

models have benefited from GPU technology and their success (e.g., [26,32]) spread the use of CNNs.

Despite the successful applications of CNNs to many computer vision tasks, the application of CNNs to regression problems are limited compared to recognition/detection/classification problems. One of the problems that CNNs successfully employed for regression analysis is age estimation. For example, [20] designed the problem as ordinal regression, [33] trained a complete multi-scale-network that can estimate age from pixels. Additionally, [9,19] handled age estimation as a classification problem by discretizing the scores and handling each one as a separate class. However, when handled as a classification problem, different classes have implicitly assumed to be independent from each other and the closeness of two class labels has no meaning to the model, which is actually informative. Therefore, we approached the score estimation task as a regression problem.

### 3 Dataset



**Fig. 1.** Camera setting used for data collection with Kinect

In this research, we used a dataset of the Microsoft Kinect V2, 70 two-minute recordings of the natural movements of 15 infants with SMA type-I between the ages 0–790 days. During the recordings infants wore colored coflex tape on their hands and feet (which were later used as color markers for limb tracking), and were positioned supine on a neutral-colored sheet. The camera was positioned above the infant centered over the sheet, which can be seen in Fig. 1. To confirm the movement quality output derived by our analysis we compared our generated motor score to the CHOP INTEND score obtained on the same day. As the CHOP INTEND has items evaluating head control, trunk control, and extremity movement and our generated motor score only includes the extremities, we separated out

only those CHOP items pertaining to extremity movement and call this the CHOP INTEND extremity score. This included 9 items that were scored for both the left and right side on a 0–4 scale [10], resulting in a maximum score of 72 points for the self-described “extremities” subset of the test, which was composed of items such as spontaneous arm movement, spontaneous leg movement, hand grip, hip strength, shoulder movement, reaching, kicking, leg movement and arm pulling.

## 4 Method

The main goal of this research is to make an assessment method for the progression of SMA type-I on infants using convolutional-neural-networks. We will describe a possible network architecture, together with the parameters used for training.

We propose a system that is capable of analyzing infant movements from videos recorded by a depth camera for disease progression assessment. A color tracker system is used to track each limb over time. For training, we use the current clinical standard, the CHOP INTEND scores for each limb assigned by trained physical therapists to evaluate neuromuscular function.

### 4.1 Data Preparation

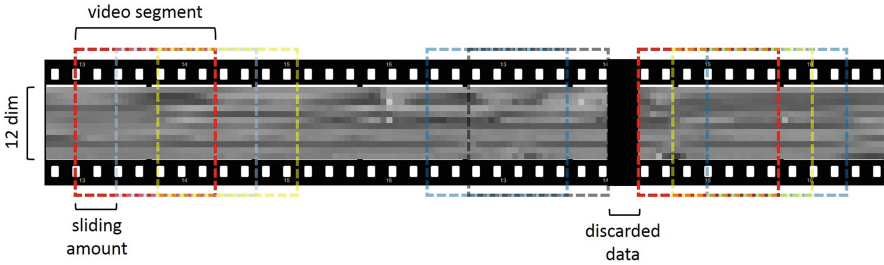
For estimating a score, we used the tracking data of four limbs available for each recording. We employed a simple color tracking technique to track limbs'  $x, y, z$  positions over time. For this, markers were manually selected by clicking to the corresponding color locations in the beginning of each video and the center of the blobs with similar hue value in proximity to the clicked location was tracked throughout the video. The output of the tracking module was the  $x, y, z$  coordinates of each limb, namely left arm, left foot, right arm and right foot for each unit time (1/30 s). The  $x$  axis corresponds to a line that would run vertically from head to toe on the child, the  $y$  axis corresponds to movement across the body or side-to-side and the  $z$  axis corresponds to movement up off the surface toward the Kinect camera. We ignore frames when tracking of one of the limbs fails, such as when an infant's hand moves behind his/her head.

### 4.2 Regression Using Convolutional Neural Networks

Since we have recordings of infants over time and the corresponding CHOP INTEND extremity scores from the same day, we compare the predicted scores with the groundtruth. We evaluate the performance of the proposed system by calculating the average error in the predicted scores over-time in leave-one-infant-out-cross-validation, which involves using one subject's data as the test set and the remaining subjects' data as the training set. Our primary measure is the average error, calculated over the data of all infants when training and testing is repeated for each infant separately.

Since our purpose is to estimate a score per sample we used an Euclidean loss ( $\frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2$ ) layer in the final layer. However, it is more difficult to optimize for Euclidean loss compared to Softmax, which is more stable. Euclidean loss is required to output exactly one correct value for each input while, for example, in Softmax the magnitude of the score is more important instead of the exact value. Also the effect of outliers are more important because of the squared-term [1]. For these reasons, the parameter selections were more critical in our case, compared to more stable loss functions.

The number of recordings in our dataset is very limited ( $N=70$ ) to train a robust convolutional neural network if tracking data for each recording is used as one sample. For this reason, we used a sliding window approach to sample the data used in both training and testing, which reduces the problem to score estimation using only  $K$  frames (instead of  $M$  frames of approximately  $120\text{ s} * 30\text{ fps} = 3600$ ). The sliding amount  $S$  effects the number of samples obtained from one recording as  $((M - K)/S) + 1$ . In our experiments, we used a window size of 900 frames ( $\approx 30\text{ s}$  of video in 30 fps) and a sliding amount of  $S = 50$  frames. Therefore each sample used had  $12 \times 900$  dimensions to be inputted to our network. Our dataset contains many samples, where the tracking of one or more limb is lost for a period of time (e.g. an infant might place his/her hand under his/her head) besides dropping frames. In order to overcome the inconsistencies between different trackings we ignored the portions of data when the tracking of all four limbs are not available, or the difference of timestamps between two frames is bigger than a preset threshold. Figure 2 represents the described sampling process.



**Fig. 2.** The sampling process of tracking data. Our data has 12 dimensions per frame; each three consecutive component of which represent the  $x - y - z$  coordinates of one particular limb: right hand, left hand, right foot and left foot.

Before sliding window sampling, we standardized each recording to zero-mean and one standard deviation for each tracking feature separately. Since the infants in our dataset are pose normalized before tracking, the positions of the limbs of infants are similar between infants. Therefore, we did not apply a standardization among different recordings. Considering the outliers might dominate the optimization when  $L2$  loss is used, we tried normalizing the regression labels but that did not produce any better results on our dataset.

Our network consists of three convolutional layers, each followed by an activation layer and a pooling layer. Rectified Linear Units (ReLU) is used for the activation layers. We have two fully connected layers connecting the convolutional layers to the loss layer. Between these two fully connected layers, our network has another activation layer followed by a dropout layer to avoid overfitting [29]. Table 1 summarizes the network architecture we used.

The fully connected layers are prone to overfitting more than other layers, because of the large number of parameters. Therefore we introduced an increased

**Table 1.** Convolutional Neural Network architecture.

Layer	Type	Top shape	Filter size	Stride	Padding
0	Input	$1 \times 1 \times 12 \times 900$	—	—	—
1	Convolutional	$1 \times 48 \times 4 \times 300$	$3 \times 3$	$3 \times 3$	0
2	ReLU	$1 \times 48 \times 4 \times 300$	—	—	—
3	Max-pooling	$1 \times 48 \times 4 \times 150$	$1 \times 3$	$1 \times 2$	0
4	Convolutional	$1 \times 128 \times 4 \times 150$	$3 \times 3$	$1 \times 1$	1
5	ReLU	$1 \times 128 \times 4 \times 150$	—	—	—
6	Max-pooling	$1 \times 128 \times 4 \times 75$	$1 \times 3$	$1 \times 2$	0
7	Convolutional	$1 \times 256 \times 4 \times 69$	$1 \times 7$	$1 \times 1$	0
8	ReLU	$1 \times 256 \times 4 \times 69$	—	—	—
9	Max-pooling	$1 \times 256 \times 4 \times 34$	$1 \times 3$	$1 \times 2$	0
10	Fully-connected	$1 \times 1024$	—	—	—
11	ReLU	$1 \times 1024$	—	—	—
12	Drop-out	$1 \times 1024$	—	—	—
13	Fully-connected	$1 \times 1$	—	—	—
14	L2 Loss	—	—	—	—

weight-decay while learning the fully connected layers. We initialized all convolutional and fully-connected layers randomly as suggested by [11] and used a constant bias. In our final model, all of our pooling layers perform max-pooling operation. We experimented with different batch sizes, and decided to use 400 samples for estimating stochastic gradients on our dataset. At each iteration we shuffled the samples since our dataset is not large enough for compensating sequential data access. We used a fixed learning-rate of  $10^{-4}$  throughout the iterations. To avoid local-minimum we used a momentum of 0.95 and to avoid over-fitting we used  $L2$  regularization with a factor of 0.9. We trained our model for 1000 iterations (approximately 286 epochs).

## 5 Results

### 5.1 Score Estimation for Extremities Using 30 s Videos

Using CNNs we could estimate the CHOP INTEND extremity scores (0–72 points) with an average error of less than 7 points per sliding window of length 900 frames. That means we could estimate the severity of SMA type-I condition on an infant with an average of less than 10% error rate using 30-seconds video recorded while the infant is behaving naturally. Note that, the ground-truth CHOP INTEND scores are assigned through a medical evaluation using distractors and response measures to these distractors. We postulate an automated solution to evaluation of severity of SMA type-I on infants while the infant

is acting naturally, which otherwise requires intensive expertise and interaction with fragile infants.

We shuffle the data at every iteration and use a batch size of 400 samples. Considering average training set size is about 1388 samples, (15 folds, leave-one-subject-out cross validation, dataset size is 1487, 30-second video-segments) we reach an epoch in approximately 3.5 iterations. Since we don't have a validation set for each fold, we observed the error on a random subset of training set.

Table 2 shows the results of score estimation experiments and standard baselines. As standard baselines we use linear regression and L2 regularized form of it: LASSO. As infants with SMA show decreased motor ability [10], we expect the correlation of different limb movements to be an indicator of severity of the disease. Therefore, for the other regression models, we used average correlation between coordinates of limb positions in each x, y, and z domain separately as our feature representation. For the LASSO regression, we used a geometric sequence of regularization coefficients, and used the largest coefficient that gives the non-null model.

**Table 2.** Average errors for 30-second tracking data, obtained from CNN, Linear Regression and LASSO

Method	Average error per subject (Fold)	Overall average error
Proposed CNN model	6.9645	8.1287
Linear regression	17.6026	15.7261
LASSO	18.3830	14.8281

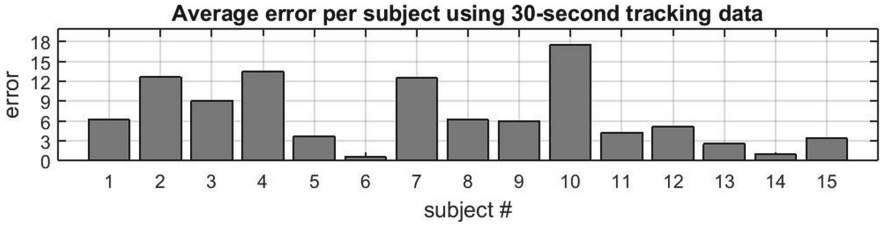
Note that, we do not incorporate the time dimension during the score estimation. Therefore, the score of each 30-second video segment (sliding window) is estimated individually, in leave-one-subject-out-cross-validation setting, which during training requires not including any video of the test subject. Average error and average per subject (fold) error differs as the number of 30-second samples in each fold are not equal. Figure 3 shows the error in score estimation per 30-second video-segment using Convolutional Neural Networks.

## 5.2 Score Estimation for the Whole Video

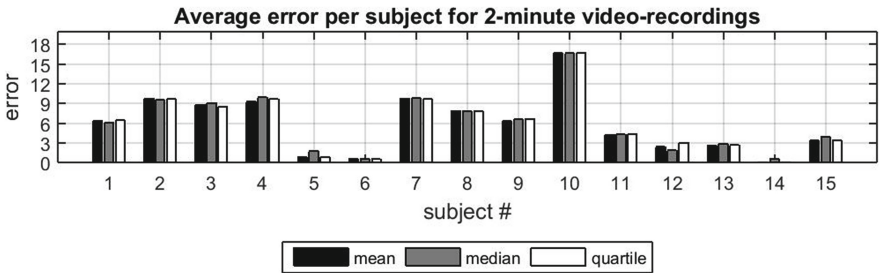
Although our intention was to predict the CHOP INTEND extremity scores from limb-tracking data of a short video (which is 30 s in our case), we also estimated a final score per 2-minute recoding using different simple statistics. Note that, in our dataset every video recording has an assigned score, but during regression we sample each video to video-segments using the sliding-window approach and therefore, every video-segment coming from the same video recording share the same ground-truth label.

In order to estimate a score per video recording, first we took the average of estimated scores of all 30-second video-segments of the corresponding video.





**Fig. 3.** Error in CHOP INTEND extremity score estimation for each 30-second video-segment sampled from 2-minute videos, which recorded over multiple visits, for each infant.

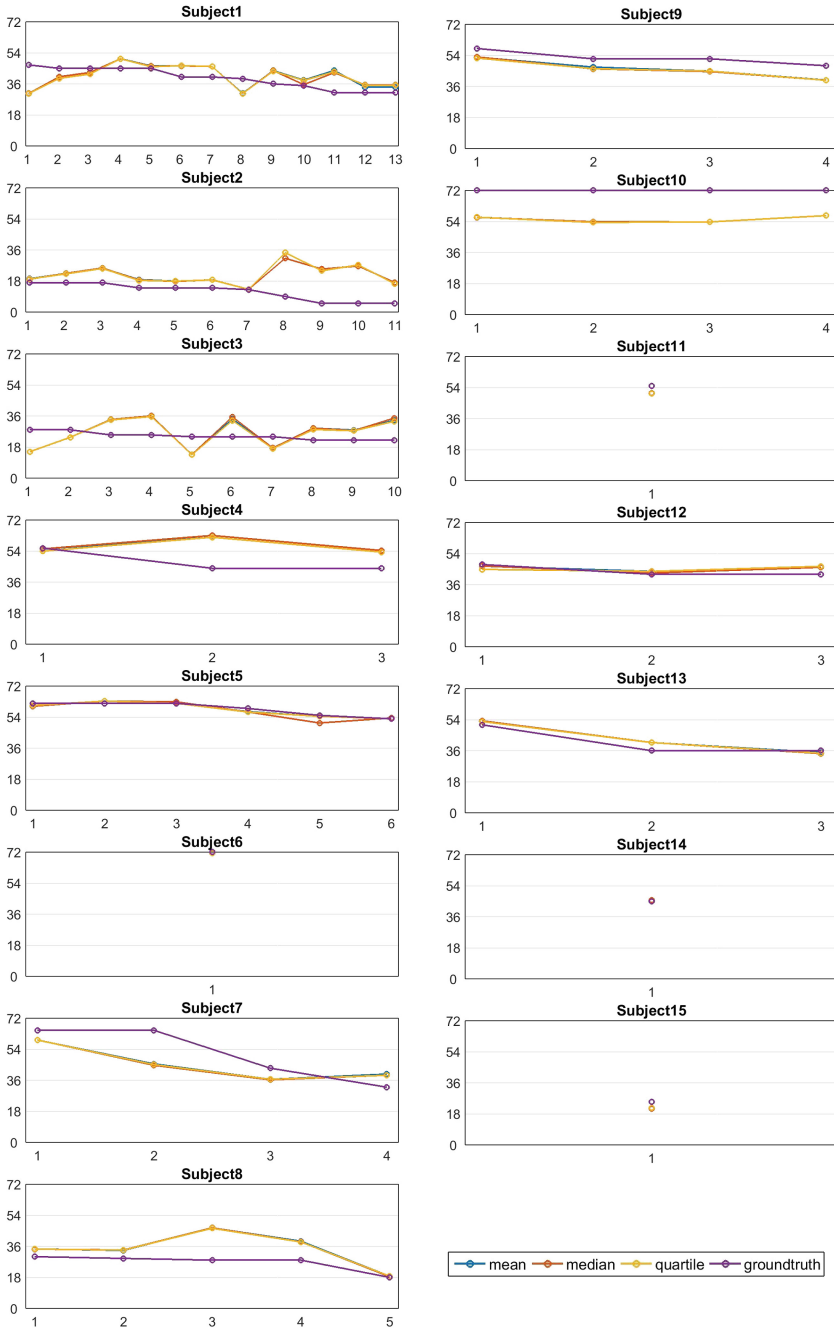


**Fig. 4.** Average errors per subject using mean, median and quartile analysis over video-segments of the corresponding 2-minute videos.

As a second method, we assigned the final score per video as the median of all estimated scores of related video-segments. Finally, we used quartile analysis for estimating a score per recording. For this, first we calculated the outliers from score estimations of 30-second video-segments per corresponding 2-minute video. We used interpolation to find upper and lower quartiles. Then, for calculating a score per 2-minute video out of the estimated scores of related video-segments, we used the average of the predictions between lower and upper fence defined as  $Q1 - 1.5 * IQR$  and  $Q3 + 1.5 * IQR$ , respectively, where  $Q1$  and  $Q3$  represent the lower and upper quartiles and  $IQR$  represents the inter-quartile range defined as the difference between upper and lower quartiles.

**Table 3.** Average errors for 2 min videos, obtained from mean, median and quartile analysis.

Method	Average of average error per subject	Overall average error
Mean	7.9648	5.9495
Median	8.0499	6.1175
Quartile analysis	8.0066	6.0291



**Fig. 5.** Estimated regression scores per 2-minute video-recording. The horizontal axis shows 2-minute video-recording number for each infant, and the vertical axis shows the scores obtained using mean, median and quartile analysis over 30-second video segments of the corresponding 2-minute video.

Figure 4 shows the average error in the estimation of CHOP INTEND extremity scores per subject using mean, median and quartile analysis over video-segments of the corresponding 2-minute videos of each subject. The average errors of using these statistics for the whole 2-minute video dataset is given in Table 3. The lowest overall average error was 5.94 points, obtained by assigning a score for each 2-minute video as the mean of its video-segments' scores. The average of average error per subject differs from overall average error since the number of samples for each subject are not equal.

Figure 5 shows CHOP INTEND extremity score estimation for each 2-minute video of all subjects, obtained by using mean, median and quartile analysis over the 30-second video segments of the corresponding 2-minute video together with the related groundtruth scores.

## 6 Conclusions and Future Work

In this paper we proposed a model for the problem of estimating the progression of SMA on naturally behaving infants. This is a very important problem due to the severity and rapid progression of the disease. CHOP INTEND is the current gold standard for evaluating SMA, which requires a lot of time for assessment, extensive clinician training, and requires interaction with the fragile infant using distractors. Our model relies on the tracking data, and does not make any domain related assumptions, therefore can be applied to progression estimation of other neuromuscular disorders.

Our experiments show that it is possible to evaluate the severity of SMA during infants' natural behavior using automated analysis of limb tracking data obtained from very short videos recorded by Kinect V2. The average per-subject error we got is  $< 7$  out of 72 points metric in leave-one-subject-out cross-validation setting for 30-second videos. When the scores estimated from 30-second video-segments of a longer video is averaged to produce a final score, we got an average error of 5.95 points.

Future work includes applying ordinal-regression that can benefit from the order of the recordings and their corresponding scores. Our results suggest that it is possible to benefit from CNNs even when the dataset size is small with the proper design and parameter selection. However, the data collection for the dataset used in this paper is still in progress. With the addition of more data, deeper and wider networks can be trained. Finally in order to estimate a score for a longer video from its segments we used simple statistical methods like mean, median, quartile analysis over the estimated scores of the shorter video-segments. As a next step, with the help of a second layer model, the estimated scores of video-segments can be combined more intelligently to produce a score for the corresponding longer video.

## References

1. <http://cs231n.github.io/neural-networks> 2. Accessed 25 June 2016
2. Adde, L., et al.: General movement assessment: predicting cerebral palsy in clinical practise. *Early Hum. Dev.* **83**(1), 13–18 (2007)
3. Benettazzo, F., Iarlori, S., Ferracuti, F., Giantomassi, A., Ortenzi, D., Freddi, A., Monteriù, A., Innocenzi, S., Capecci, M., Ceravolo, M.G., Longhi, S.: Low cost RGB-D vision based system to support motor disabilities rehabilitation at home. In: Andó, B., Siciliano, P., Marletta, V., Monteriù, A. (eds.) *Ambient Assisted Living. Biosystems & Biorobotics*, pp. 449–461. Springer, Cham (2015)
4. Cano, S., et al.: Rasch analysis of clinical outcome measures in spinal muscular atrophy. *Muscle Nerve* **49**(3), 422–430 (2014)
5. Cioni, G., et al.: Which better predicts later outcome in fullterm infants: quality of general movements or neurological examination? *Early Hum. Dev.* **50**(1), 71–85 (1997)
6. Crawford, T.: Concerns about the design of clinical trials for spinal muscular atrophy. *Neuromuscul. Disord.* **14**(8), 456–460 (2004)
7. Finkel, R.: Electrophysiological and motor function scale association in a pre-symptomatic infant with spinal muscular atrophy type I. *Neuromuscul. Disord.* **23**(2), 112–115 (2013)
8. Finkel, R., et al.: Observational study of spinal muscular atrophy type I and implications for clinical trials. *Neurology* **83**(9), 810–817 (2014)
9. Geng, X., et al.: Automatic age estimation based on facial aging patterns. *PAMI* **29**(12), 2234–2240 (2007)
10. Glanzman, A., et al.: The Children’s hospital of philadelphia infant test of neuromuscular disorders (CHOP INTEND): test development and reliability. *Neuromuscul. Disord.* **20**(3), 155–161 (2010)
11. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: *AISTATS* (2010)
12. González-Ortega, D., et al.: A kinect-based system for cognitive rehabilitation exercises monitoring. *Comput. Methods Programs Biomed.* **113**(2), 620–631 (2014)
13. Hadders-Algra, M.: Evaluation of motor function in young infants by means of the assessment of general movements: a review. *Ped. Phys. Ther.* **13**(1), 27–36 (2001)
14. Han, J., et al.: Reachable workspace in facioscapulohumeral muscular dystrophy (FSHD) by kinect. *Muscle Nerve* **51**(2), 168–175 (2015)
15. Hondori, H., et al.: A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *J. Med. Eng.* (2014)
16. Iarlori, S., et al.: RGBD camera monitoring system for alzheimers disease assessment using recurrent neural networks with parametric bias action recognition. In: *ICCHP* (2014)
17. Kontschieder, P., et al.: Quantifying progression of multiple sclerosis via classification of depth videos. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014, Part II. LNCS*, vol. 8674, pp. 429–437. Springer, Heidelberg (2014)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *NIPS* (2012)
19. Mu, G., et al.: Human age estimation using bio-inspired features. In: *CVPR* (2009)
20. Niu, Z., et al.: Ordinal regression with multiple output CNN for age estimation. In: *CVPR* (2016)

21. Oskarsson, B., et al.: Upper extremity 3-dimensional reachable workspace assessment in amyotrophic lateral sclerosis by kinect sensor. *Muscle Nerve* **53**(2), 234–241 (2016)
22. Phagava, H., et al.: General movements in infants with autism spectrum disorders. In: *Georgian Medical News* (2008)
23. Prechtl, H., Hopkins, B.: Developmental transformations of spontaneous movements in early infancy. *Early Hum. Dev.* **14**(3), 233–238 (1986)
24. Rahmati, H., et al.: Frequency-based features for early cerebral palsy prediction. In: *EMBC* (2015)
25. Russman, B.S.: Spinal muscular atrophy: clinical classification and disease heterogeneity. *J. Child Neurol.* **22**(8), 946–951 (2007)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR* (2014)
27. Sooklal, S., et al.: Using the kinect for detecting tremors: challenges and opportunities. In: *BHI* (2014)
28. Spasojević, S., et al.: A vision-based system for movement analysis in medical applications: the example of parkinson disease. In: *Computer Vision Systems* (2015)
29. Srivastava, N., et al.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
30. Stahl, A., et al.: An optical flow-based method to predict infantile cerebral palsy. *Neural Syst. Rehabil. Eng.* **20**(4), 605–614 (2012)
31. Swoboda, K.J., et al.: Perspectives on clinical trials in spinal muscular atrophy. *J. Child Neurol.* **22**(8), 957–966 (2007)
32. Szegedy, C., et al.: Going deeper with convolutions. In: *CVPR* (2015)
33. Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) *ACCV 2014*. LNCS, vol. 9005, pp. 144–158. Springer, Heidelberg (2015)
34. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*, Part I. LNCS, vol. 8689, pp. 818–833. Springer, Heidelberg (2014)
35. Zerres, K., Rudnik-Schoneborn, S.: Natural history in proximal spinal muscular atrophy: clinical analysis of 445 patients and suggestions for a modification of existing classifications. *Arch. Neurol.* **52**(5), 518–523 (1995)