Automated System to Measure Static Balancing in Children to Assess Executive Function

Hamza Reza Pavel
hamzareza.pavel@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Enamul Karim
enamul.karim@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Mohammad Zaki Zadeh
mohammad.zakizadehgharie@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Ashish Jaiswal
ashish.jaiswal@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Rithik Kapoor
rithik.kapoor@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Fillia Makedon
makedon@uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

ABSTRACT

We present multiple methods based on computer vision and deep learning to automate the task of "Balancing on one foot". This is one of the Activate Test of Embodied Cognition (ATEC) tasks used to measure cognitive skills in children through physical activity. A dataset of 27 children performing the ATEC task is used to train and validate the deep learning models used to automate the task. As opposed to most balance identification systems that use sensors, our proposed approach relies only on computer vision which can be easily deployed at home or classroom environment, is portable, and cheap. Our proposed system automatically identifies the task and assigns an ATEC and an ergonomics score for the "Balancing on one foot" task. Our proposed system achieves an accuracy of 97% when calculating the raw score for the ATEC task and 86.5% for assigning the ergonomic score.

CCS CONCEPTS

• **Human-centered computing**; • **Computing methodologies**;

KEYWORDS

Computer Vision, Embodied Cognition, Cognitive Assessment, Human Activity Recognition, Ergonomics

ACM Reference Format:
Hamza Reza Pavel, Enamul Karim, Mohammad Zaki Zadeh, Ashish Jaiswal, Rithik Kapoor, and Fillia Makedon. 2022. Automated System to Measure Static Balancing in Children to Assess Executive Function. In *The15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3529190.3534750

PETRA '22, June 29-July 1, 2022, Corfu, Greece

ACM ISBN 978-1-4503-9631-8/22/06…$15.00
https://doi.org/10.1145/3529190.3534750

1 INTRODUCTION

Executive functions (EFs) are high-order cognitive processes that are necessary to regulate cognitive control of behavior. The cognitive control of behavior is required for us to attain our chosen goals and is essential for a person's positive development and the ability to make healthy life choices. According to this study, core EFs are based on inhibition, interference control, working memory, and cognitive flexibility [13]. People suffering from Attention Deficiency Hyper-Activity Disorder (ADHD), depression, or other mental or learning disorders often have issues with higher-order executive functions [26] [30]. Cognitive impairment in executive functions in children could lead to poor academic performance, and various other issues such as substance use disorder, impulse disorder, and other mental illnesses. It is important to identify cognitive problems in early childhood due to brain plasticity being highest in children. Early diagnosis also provides the chance for remedial interventions and overall better quality of life for the children affected with cognitive issues.

There are various tools for cognitive assessments of children, each with different goals and target age group [28] [22]. One such standard tool is the NIH toolbox [35]. The two tasks of the NIH toolbox, the Dimensional Change Card Sort (a measure of cognitive flexibility) and a flanker task (a measure of inhibitory control in the context of selective visual attention), are administered using computers or tablets. These tasks require little to no body movements and various study shows a strong correlation between cognitive and motor skills [10] [11]. Thus the need for an assessment method arises which is both mentally and physically challenging for children, resembles day-to-day activity, and is an objective measure. The ATEC is an assessment method designed to measure executive functions in children through physically and cognitively demanding tasks [5, 6, 14, 16, 32–34]. The ATEC aims to provide an automated system that is cost-effective, easy to deploy, can capture the movement of children accurately, and generate a reliable score for the tasks performed by children.

Some executive functions such as motivation, behavior organization, working memory, inhibition, etc., are part of the ergonomics/human factors behavioral spectrum [15]. Previously, the authors introduced two ATEC tasks named "Ball-Drop-to-the-Beat" and "Sailor Step" with an automated scoring system for the tasks [14] [25]. In this paper, we introduce a new ATEC task with an automated scoring system. Apart from the raw score to calculate the ATEC score, our system provides an ergonomics score for the task performed by the children. This ergonomics score is an indicator of the efficiency of the static balancing task performed by the children and correlates with their executive function. The main contributions of this work are:

- A new ATEC task named "Balancing on one foot" to assess cognition in children.
- Computer-vision and deep learning-based system to automatically calculate the ATEC score and ergonomics score for the introduced task.

The rest of the paper is organized as follows: Section 2 discusses related work, Section 3 provides a short introduction to ATEC and the assessment task, Section 4 provides the data collection and annotation method, Section 5 describes our proposed methods, Section 6 discusses experiments and results, section 7 discusses how to relate the raw score to ATEC score for cognitive measure followed by conclusion in Section 8.

2 RELATED WORK

The identification of the "Balancing on one foot" task is a problem of the Human Activity Recognition (HAR) domain. Human Activity Recognition is the task of classifying or predicting action performed by a single or group of individuals. HAR classification tasks require a series of data points which makes them different from image classification or object detection tasks. In this study [14], authors used OpenPose to extract the key body points from video frames. These extracted key body points were used in tandem with machine learning techniques such as K-Nearest Neighbor, Random Forest, Decision Tree, and Multi-layer Perceptron classifiers to identify the tasks performed by children. In another study [25], authors proposed a multi-modal system that employs an attention-based fusion mechanism to combine multiple modalities such as optical flow, human poses, and objects in the scene to predict a child's action. This multi-modal approach has a higher accuracy than the approach mentioned in [14].

In most cases, combining deep learning methods with traditional classifiers such as SVM is used to identify human activities [7]. This study presents a convolutional neural network-based approach for activity recognition by combining multiple vision cues [17]. The method presented in this study [29] uses a convolutional neural network (CNN) and deep bidirectional LSTM (DB-LSTM) network to preprocess the video data as a whole. The extracted features from the frames are learned using a DB-LSTM network, where multiple layers are stacked together in both forward pass and backward pass of DB-LSTM to increase its depth and learn long term sequence [29]. This method performs better than single-frame-based activity recognition models.

Another popular method of human activity recognition is by using the key body points extracted using Openpose [9] or other methods that extract key body points [21]. Deep neural network or other classifiers are used to identify and classify poses using key body points extracted using these methods [18] [23] [31]. Similar to these methods, our proposed system uses extracted key body points to identify the static balancing state of the children.

3 THE ACTIVATE TEST FOR EMBODIED COGNITION AND TASK DESCRIPTION

The ATEC [8] measures cognitive function in children. Unlike traditional standard measures for executive function such as the NIH toolbox, the tasks in ATEC are both physically and cognitively demanding. The test is administered using a system that scores the participant automatically with motion and video capture technologies. The ATEC tasks measure executive and motor functions which are then converted into a final ATEC score to describe the level of cognitive development.

There are 17 physical tasks in ATEC to measure balance, rhythm, response inhibition, coordination, attention, memory, working memory, motor speed, etc. Previously, automated scoring solutions for three of the ATEC tasks, "Tandem Gait" [33], "Ball-Drop-to-the-Beat", and "Sailor Step" [14] were developed. In this paper, we provide multiple solutions to automatically identify and score the "Standing on One Foot" task performed by children. This task belongs to the *gait and balance* domain of the ATEC tasks.

Even though it might appear simple, balancing on one foot requires focus, coordination between muscles and works as a building block for more advanced motor skills. This study [27] shows that dyslexic children are less stable when balancing on one foot. On the contrary, this study finds a relationship between poor balancing and attention deficit hyperactivity disorder (ADHD) [24]. Poor balancing is related to early childhood cognitive deficiencies, and "Balancing on One Foot" is an effective way to identify it.

4 DATA COLLECTION AND ANNOTATION

A dataset is created for the "Balancing on one foot" task from participants ranging from 6 to 11 years old of age. The Figure 1 shows the setup used for the data collection step. An RGB camera is used to capture the front view of the children performing the task of static balancing on one foot. The cameras used to collect data are connected to a GUI model that the administrator can use to monitor the flow of the assessment. A display in front of the children shows interactive videos to make the data collection process more engaging for the children. The dataset consists of 27 videos from 20 subjects. After parental consent and screening processes required by the study protocol were satisfied, data was collected in the classroom and home environment. Before each session, the participating children were instructed on the task.

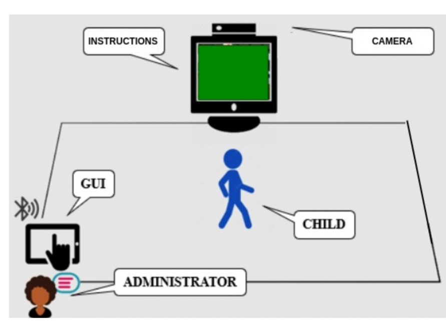


Figure 1: Data collection setup

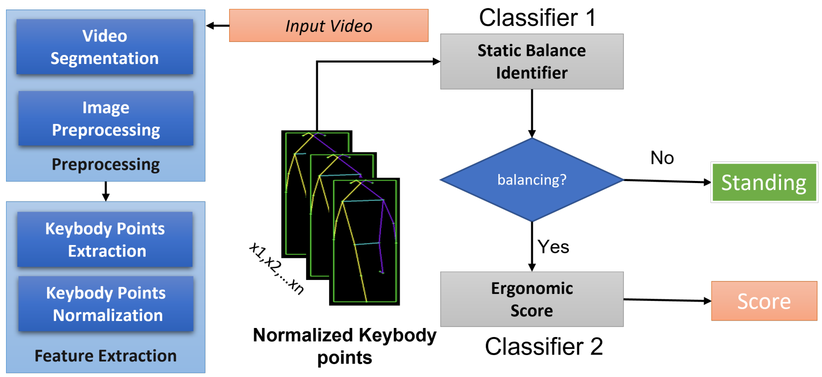


Figure 2: Complete system overview.

Each video session in the dataset ranged between 14 seconds to 17 seconds. The videos are captured with equipment that provides 30 frames-per-second. For annotation purposes, five frames per second are sampled from the videos. As the subjects moved relatively slowly, the impact of sampling images at five frames-per-second for training the models is sufficient.

The sampled images are cropped to get rid of unnecessary objects and to reduce computational costs. Then each image is assigned two labels. The labeling is done by expert annotators by manually going over each of the sampled image. The first label is the static balancing state, which is either "Standing" or "Balancing". The label "Standing" means children are standing on both feet and the label "Balancing" means children are performing the static balancing task of balancing on one foot. A second label between 0 to 3 based on the subject's balancing posture is assigned to the images. This second label serves as the ergonomic score of the children while performing the task. If the subject is standing on two feet, a label of 0 is assigned. If the subject is balancing on one foot, a label between 1 to 3 is assigned based on the posture and stability. An image having the label 3 means the subject performing the balancing task has good stability and posture. A label of 1 means poor balance and stability of the subject. This labeled dataset is used to train the models in the system to calculate the raw score of the ATEC balancing task and ergonomics score for the task.

5 PROPOSED METHODS

Unlike methods that use specialized sensors or devices to detect balance or lack of balance [4] [12], our solution is based on computer vision. In our proposed solutions, we use MoveNet [1] with Tensorflow [3] to extract the key points of the human body and to train our models. Our proposed solutions can be deployed on smartphones or tablets and do not require any specialized sensors or devices. As no specialized sensors or devices are required, our proposed solution is portable, cheap, and easy to deploy in a home or classroom environment.

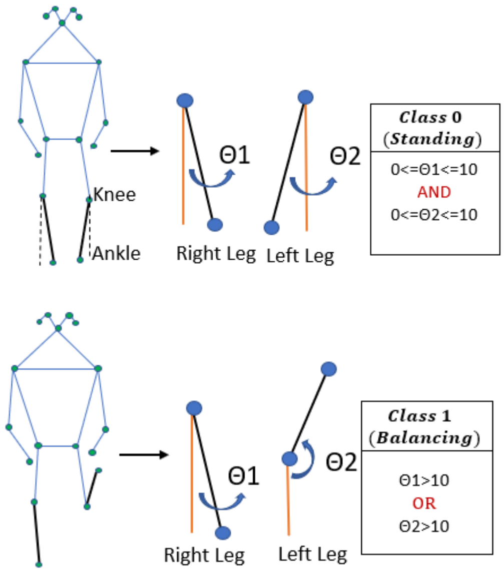
5.1 Static Balance Identification

The Figure 2 shows an overview of our proposed system. Our system has a preprocessing module, a feature extraction module, and two classifiers. The preprocessing module segments the video into images. As the movements of the children are slow, we obtain every sixth frame from the videos to reduce the number of similar images and computational costs. The images are preprocessed to move the subjects to the center and resized to reduce computational complexity and remove irrelevant subjects from the background. The feature extraction module uses MoveNet [1] to extract the 17 2D key body points along with their confidence score from each of the preprocessed images. The extracted features are then used by the two classifiers of the system. The accuracy of the entire system depends on the two classifiers. The first classifier is used to identify the static balancing state of the children and the second classifier is used to score the ergonomics of their posture if they are performing the static balancing task. The number of frames where the children are performing the static balancing task is identified and used to calculate the total time they are performing the balancing task. The ergonomics score of each frame where they are performing the balancing task is obtained from the second classifier and averaged to obtain the overall ergonomics score of the task. The first task is to determine whether the subject under consideration is standing or balancing. This step is critical since the ergonomic ratings of the balancing state can only be evaluated once the classification has been done successfully.

5.1.1 Range Based Classification. For static balance identification, we developed heuristics to see how it can classify the subject under consideration from the test image to one of the two available classes: Standing and Balancing. The angle formed by the line segment joining the knee and the ankle with a line drawn perpendicular to the horizon is used as the classification metric.

Let us denote the perpendicular to the horizon with P. Based on observation by the domain experts, it is assumed that this angle formed by both the legs in standing posture is between 0 to 10 degrees on either side of P. To summarize:

- Classify the test image to Class 0 **(Standing)** if the angle formed by the line segment joining the ankle and knee with P is between 0 to 10 degrees for *both the legs.*
- Classify the test image to Class 1 **(Balancing)** if the angle formed by the line segment joining the ankle and knee with P exceeds the 10 degrees limit by *at least one of the legs.*



(b) Balancing

Figure 3: Range based classification

In Figure 3a, the angles formed by the leg segments of both the legs with P fall within the 10 degrees range. Here $\theta 1$ represents the angle formed by the line segment joining the right ankle and the right knee with P whereas $\theta 2$ represents the angle formed with that of the left ankle and left knee. Since both $\theta 1$ and $\theta 2$ ranges from 0 to 10 degrees, the subject under consideration has been identified to be in a standing position. On the contrary, Figure 3b depicts the scenario where the subject is in a balancing position. Although $\theta 1$ is within 10 degrees, $\theta 2$ exceeds the 10 degrees limit. So, it has been identified to be in a balancing position.

5.1.2 Deep Neural Network: An improvement is to use a deep neural network instead of the previously mentioned naive approach as the first classifier to identify the balancing state of children. In this approach, the feature obtained from the feature extraction module is used to train a neural network. For each image in the training dataset, the extracted features and associated first label are used. The first label describes the balancing state of the children which is either "Standing" or "Balancing". For each image, the extracted feature matrix has a shape of $17X3$ which is the x, y coordinates, and the confidence score of the 17 key body points. The extracted features are first normalized and then flattened into a vector of shape $51X1$ which is used to train a deep neural network model. Once the model is trained, the testing dataset is used to validate the model. If the output of the first model is "Balancing", then the input key body points of the system is passed to the second classifier to calculate the ergonomics score. If the output of the first model is "Standing", an ergonomic score of 0 is assumed for that frame.

5.2 Ergonomics Scoring

Researchers have widely used ergonomic postural assessment methods to evaluate the risks of musculoskeletal disorders[19]. In this paper, we have analyzed the ergonomic posture of children to assess their executive functions.

Figure 4: Ergonomic scores are assigned between 1 to 3. Subject in (a) is assigned an ergonomic score of 1, (b) is assigned an ergonomic score of 2, and (c) is assigned an ergonomic score of 3.

We have proposed three different approaches to score and evaluate body posture on a scale of 1 to 3 depending on how well the children balance. A score of 1 indicates that the child's posture is **poor**, a score of 2 indicates that the child's posture is **average**, and a score of 3 indicates that the child's posture is **excellent**.

5.2.1 Weighted Matching: We calculated the weighted score for each test image by comparing its key points extracted using MoveNet to those derived from the training images for all three classes. We used the weighted distance, which takes into consideration the confidence score as well as the x and y coordinates of the retrieved keypoints [2].

$$D(F, G) = \frac{1}{\sum_{k=1}^{17} F_{Ck}} X \sum_{k=1}^{17} F_{Ck} \, \|Fxy_k - Gxy_k\|$$

Here, F and G are the two pose vectors that could only be compared after $L2$ normalization, F_{xy} and G_{xy} are the x and y coordinates of the k^{th} keypoint for each vector, F_{ck} represents the confidence score of the k^{th} keypoint of F. After computing the weighted

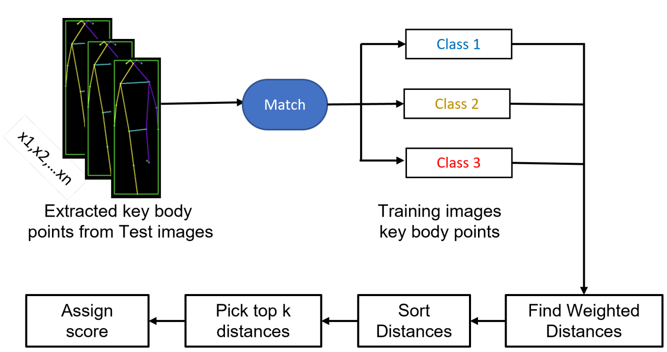


Figure 5: Weighted Matching Ergonomic Scoring

distance of each test image with that of the training images, we sort the instances in increasing order of their distance along with the corresponding class labels. Finally, we pick the top k distances and score the test images based on the maximum occurrence of the class labels in the top k values.

5.2.2 Range Based Scoring: Similar to the previous range-based classification, we adapted the range-based scoring and assigned a score of 1 to 3 to each of the three selected ranges. The angle ranges are selected based on expert observations. Table 1 depicts the different ranges and its corresponding scores.

Table 1: Range Based Scoring.

Angle Ranges	Score
0 to 20 degrees	1
20 to 40 degrees	2
Otherwise	3

To begin, we must determine which leg is elevated above the ground when balancing. Then we should measure the angle it is creating with the line perpendicular to the ground. If the angle is between 0 to 20 degrees, we can say that the subject is not balancing well and we assign an ergonomic score of 1. If the angle is between 20 and 40 degrees, we consider it average and give it a score of 2. If this angle goes beyond 40 degrees, we can say that it the subject is in an excellent balancing state and hence assign a maximum score of 3.

5.2.3 Neural Network Based Scoring: We can use a deep neural network based classifier to assign an ergonomic score to the balancing state of children in each frame. Same as the neural network based classifier used to identify the static balancing state of the children, the second deep neural network based classifier is trained with the key body points. But instead, the first label, the second label of the image frames is used from the annotation, which is the ergonomic score of the children for that particular frame. The input shape of the second classifier is the same as the first deep neural network based classifier with the difference in the output. The output of the second classifier is either the label "1", "2", or "3" which corresponds to the ergonomic score of the children for that particular image frame.

6 EXPERIMENTS AND RESULTS

The experiments are performed in a system with an intel core i7-8750 quad-core CPU, 16GB of RAM, and NVIDIA GTX 1060 GPU with 120 Cuda cores and 14GB of graphics memory.

For training and testing the models, an 80-20 split is done on the dataset where 80% data are used for training and 20% of the data are used for testing and evaluation. The models are trained for 5 such different 80-20 splits of the dataset and the accuracy of each of the 5 iterations is averaged and reported in this section. We use SVM in place of both the first and second classifier in the system as a baseline for comparison.

For the Range Based Classification, we have selected a range of 10 degrees on either side of the line drawn perpendicular to the horizon based on expert opinion. Using this method, we have achieved an accuracy of 87.15%.

For the first classifier when a deep neural network is used, we get an accuracy of 97% which is a substantial improvement over the naive approach. This deep neural network classifier is trained in Tensorflow [3] framework for 100 epochs with ADAM optimizer[20]. The learning rate and batch sizes are 0.001 and 48 respectively. These values are obtained empirically. The models are trained using the feature extracted using MoveNet [1] from the segmented images. The following table shows the accuracy of the various approach used as the first classifier to identify the static balancing state.

Table 2: Classifiers accuracy to identify static balancing state.

Method	Overall Accuracy
Range Based Classification	87.15%
SVM	95.48%
Deep Neural Network	97%

For the weighted matching, we first analyzed the number of samples from each of the individual classes in the training dataset. Computing the weighted matching using an imbalanced dataset may lead to a biased result. Hence, we needed to down-sample the number of instances of the classes to make it balanced. Then, we computed the weighted distance of the key points of each test image with that of all the available key points for the training images. The lower the distance, the closer are the points. We sort the distances, pick the top 3 and assign a score of 1 to 3 based on the maximum occurrence of the class labels. We achieved an accuracy of 77.24% using the weighted matching technique. Using the range-based scoring method, where we have selected three different angle ranges for the three different classes, we have achieved an accuracy of 61.53%.

Table 3: Classifiers accuracy to assign ergonomics score.

Method	Overall Accuracy
Range Based Scoring	61.53%
Weighted matching	77.24%
SVM	80.1%
Deep Neural Network	86.5%

For the second classifier, we can also use a deep neural network to obtain the ergonomics score. The deep neural network classifier performs better than the rest of the approaches. This deep neural network classifier is trained in Tensorflow [3] framework for 200 epochs with ADAM optimizer[20]. The learning rate and batch sizes are 0.001 and 48 respectively. These values are also obtained empirically.

7 RELATING RAW SCORES TO COGNITIVE MEASURES

The proposed system automatically calculates the raw score of the balancing task and the ergonomics score of the task. As each collected video consists of 30 frames per second, we can easily calculate the raw score for the standing on one-foot task, which is the time the children are maintaining balance on one foot. If the children are balancing on one foot in F_B frames in an F_N frame video, and the video is captured in 30FPS, then the time children are balancing on one foot is,

$$T_B = \frac{1}{30} * F_B$$

Where T_B is the total time in seconds the children are balancing on one foot, F_B is the number of frames in which children are performing the static balancing task, and F_N is the number of frames in the video.

This system-provided raw score has to be meaningful to specialists, which is why this raw score is converted to a meaningful cognitive measure using the ATEC rubric. The following table contains the ATEC rubric for converting the raw score to the ATEC score.

Table 4: Raw score to ATEC score conversion rubric.

Balancing Time T_B	ATEC Score
$T_B < 5$	0
$5 <= T_B <= 8$	1
$9 <= T_B <= 10$	2

For calculating the ergonomics score, only the frames where the children are balancing on one foot is considered. If the children are balancing on one foot in F_B frames and summation of ergonomics score for the session is S_{total}, then ergonomics score for the session is,

$$S_{Eg} = \frac{S_{total}}{F_B}$$

8 CONCLUSION AND FUTURE WORK

In this paper, we presented a novel dataset of 27 children performing the task of balancing on one foot. We also proposed a computer vision-based system to automatically score the task and assign an ergonomics score for the task with acceptable accuracy. Our proposed approach is portable and cheap which can be deployed in the classroom or home environment using a smartphone or tablet. Our goal is to design a fully automated and low-cost system for the other ATEC tasks and collect more data to improve the accuracy of the tasks. Apart from that, we want to develop an automatic ergonomic scoring system for the other ATEC tasks which can be extended to other domains.

REFERENCES

[1] 2021. *Next-Generation Pose Detection with MoveNet and TensorFlow.js*. https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

[2] 2022. *Move Mirror: An AI Experiment with Pose Estimation in the Browser using TensorFlow.js*. https://blog.tensorflow.org/2018/07/move-mirror-ai-experiment-with-pose-estimation-tensorflow-js.html

[3] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ Software available from tensorflow.org.

[4] Federica Aprigliano, Michelangelo Guaitolini, Angelo Maria Sabatini, Silvestro Micera, and Vito Monaco. 2019. Pre-impact detection algorithm to identify lack of balance due to tripping-like perturbations. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2430–2433.

[5] Ashwin Ramesh Babu, Akilesh Rajavenkatanarayanan, James Robert Brady, and Fillia Makedon. 2018. Multimodal approach for cognitive task performance prediction from body postures, facial expressions and EEG signal. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*. 1–7.

[6] Ashwin Ramesh Babu, Mohammad Zakizadeh, James Robert Brady, Diane Calderon, and Fillia Makedon. 2019. An Intelligent Action Recognition System to assess Cognitive Behavior for Executive Function Disorder. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 164–169.

[7] Hend Basly, Wael Ouarda, Fatma Ezahra Sayadi, Bouraoui Ouni, and Adel M Alimi. 2020. CNN-SVM learning approach based human activity recognition. In *International Conference on Image and Signal Processing*. Springer, 271–281.

[8] Morris D Bell, Andrea J Weinstein, Brian Pittman, Richard M Gorman, and Maher Abujelala. 2021. The Activate Test of Embodied Cognition (ATEC): Reliability, concurrent validity and discriminant validity in a community sample of children using cognitively demanding physical tasks related to executive functioning. *Child Neuropsychology* 27, 7 (2021), 973–983.

[9] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. arXiv:1812.08008 [cs.CV]

[10] Catherine L Davis and Stephanie Cooper. 2011. Fitness, fatness, cognition, behavior, and academic achievement among overweight children: do cross-sectional associations correspond to exercise trial outcomes? *Preventive medicine* 52 (2011), S65–S69.

[11] Emma E Davis, Nicola J Pitchford, and Ellie Limback. 2011. The interrelation between cognitive and motor development in typically developing children aged 4–11 years is underpinned by visual processing and fine manual control. *British Journal of Psychology* 102, 3 (2011), 569–584.

[12] Daniela De Venuto and Giovanni Mezzina. 2020. High-specificity digital architecture for real-time recognition of loss of balance inducing fall. *Sensors* 20, 3 (2020), 769.

[13] Adele Diamond. 2013. Executive functions. *Annual review of psychology* 64 (2013), 135–168.

[14] Alex Dillhoff, Konstantinos Tsiakas, Ashwin Ramesh Babu, Mohammad Zakizadehghariehali, Benjamin Buchanan, Morris Bell, Vassilis Athitsos, and Fillia Makedon. 2019. An automated assessment system for embodied cognition in children: from motion data to executive functioning. In *Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction*. 1–6.

[15] Jeffrey W Elias and Zoe Clelland. 2000. Aging, Executive Function, and Human Factors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 4–80.

[16] Srujana Gattupalli, Ashwin Ramesh Babu, James Robert Brady, Fillia Makedon, and Vassilis Athitsos. 2018. Towards deep learning based hand keypoints detection for rapid sequential movements from rgb images. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 31–37.

[17] Pushpajit Khaire, Praveen Kumar, and Javed Imran. 2018. Combining CNN streams of RGB-D and skeletal data for human activity recognition. *Pattern Recognition Letters* 115 (2018), 107–116.

[18] Ue-Hwan Kim, Dongho Ka, Hwasoo Yeo, and Jong-Hwan Kim. 2020. A Real-time Vision Framework for Pedestrian Behavior Recognition and Intention Prediction at Intersections Using 3D Pose Estimation. arXiv:2009.10868 [cs.CV]

[19] Woojoo Kim, Jaeho Sung, Daniel Saakes, Chunxi Huang, and Shuping Xiong. 2021. Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose). *International Journal of Industrial Ergonomics* 84 (2021), 103164.

[20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[21] Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. 2020. VIBE: Video Inference for Human Body Pose and Shape Estimation. arXiv:1912.05656 [cs.CV]

[22] Tomonori Koyama, Hirokazu Osada, Hiromi Tsujii, and Hiroshi Kurita. 2009. Utility of the Kyoto Scale of Psychological Development in cognitive assessment of children with pervasive developmental disorders. *Psychiatry and clinical neurosciences* 63, 2 (2009), 241–243.

[23] AJ Piergiovanni and Michael S. Ryoo. 2018. Fine-Grained Activity Recognition in Baseball Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[24] Thomas Raberger and Heinz Wimmer. 2003. On the automaticity/cerebellar deficit hypothesis of dyslexia: balancing and continuous rapid naming in dyslexic and ADHD children. *Neuropsychologia* 41, 11 (2003), 1493–1497.

[25] Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Ashish Jaiswal, Alexis Lueckenhoff, Maria Kyrarini, and Fillia Makedon. 2020. A multi-modal system to assess cognition in children from their physical movements. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 6–14.

[26] Hannah R Snyder. 2013. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological bulletin* 139, 1 (2013), 81.

[27] Catherine J Stoodley, Angela J Fawcett, Roderick I Nicolson, and John F Stein. 2005. Impaired balancing ability in dyslexic children. *Experimental Brain Research* 167, 3 (2005), 370–380.

[28] Marcela Tenorio Delgado, Paulina Arango Uribe, Andres Aparicio Alonso, and Ricardo Rosas Diaz. 2016. TENI: A comprehensive battery for cognitive assessment based on games and technology. *Child Neuropsychology* 22, 3 (2016), 276–291.

[29] Amin Ullah, Jamil Ahmad, Khan Muhammad, Muhammad Sajjad, and Sung Wook Baik. 2017. Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE access* 6 (2017), 1155–1166.

[30] Erik G Willcutt, Alysa E Doyle, Joel T Nigg, Stephen V Faraone, and Bruce F Pennington. 2005. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological psychiatry* 57, 11 (2005), 1336–1346.

[31] Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey, and Shaik Ali Akbar. 2021. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems* 223 (2021), 106970.

[32] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. 2021. Self-supervised human activity recognition by augmenting generative adversarial networks. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. 171–176.

[33] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, Maria Kyrarini, Morris Bell, and Fillia Makedon. 2021. Automated system to measure Tandem Gait to assess executive functions in children. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. 167–170.

[34] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, and Fillia Makedon. 2022. Self-Supervised Human Activity Representation for Embodied Cognition Assessment. *Technologies* 10, 1 (2022), 33.

[35] Philip David Zelazo, Jacob E Anderson, Jennifer Richler, Kathleen Wallner-Allen, Jennifer L Beaumont, and Sandra Weintraub. 2013. II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development* 78, 4 (2013), 16–33.