Deep Learning based Driving Posture Stability Analysis for People with Mobility Challenges

Ruoqi Huang
Computer Science Department
Cal Poly Pomona
Pomona, CA, USA
ruoqihuang@cpp.edu

Mai Narasaki-Jara
Kinesiology & Health Promotion Department
Cal Poly Pomona
Pomona, CA, USA
mnjara@cpp.edu

Tingting Chen
Computer Science Department
Cal Poly Pomona
Pomona, CA, USA
tingtingchen@cpp.edu

2023 IEEE International Conference on Big Data (BigData) | 979-8-3503-2445-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigData59044.2023.10386922

***Abstract*—Mobility scooters are critical in facilitating social participation of people with mobility challenges and thus improving their life quality. However, the safety issues of driving mobility scooter are real concerns and may not be assessed in a timely fashion for many patients. In this paper we enable driving posture stability analysis at home or in the community setting for people with mobility challenges by using only video recordings of their driving. In particular, we design a system that extracts upper body keypoints' 2D coordinates from video frames and builds an autoencoder model to perform stability analysis. We explore two architectures of the autoencoder, with Long Short Term Memory (LSTM) and Convolutional Neural Networks (CNN), respectively emphasizing the temporal and spatial relationships of upper body keypoints in driving movements. Evaluations using patients driving posture data collected have shown that both architectures have achieved over 0.99 Precision-Recall Area Under the Curve (AUC), and 0.8 ROC AUC, indicating excellent model accuracy levels.**

***Index Terms*—Driving Stability Analysis, Pose Estimation, Computer Vision, AutoEncoders, Deep Learning, Patient Driving Data**

I. Introduction

Mobility disability is the most common kind of disability in the USA. According to Centers for Disease Control and Prevention (CDC) statistics in 2018, 13.7 percent of U.S. adults have a mobility disability [1]. Mobility scooters, a type of electrically powered scooters, are an affordable and popular type of assistive mobility technology. Mobility scooters are critical in facilitating social participation of people with mobility disability and thus improving their life quality. However, A high number of mobility scooter accidents such as falling and colliding with pedestrians or obstacles have been reported [8]. According to data from the National Electronic Injury Surveillance System [11], the number of accidents involving mobility scooters were treated in American Emergency Departments has been increasing every year [4].

People with mobility challenges have various medical conditions that affect their physical and cognition abilities to drive (e.g., limited vision, lack of motion stability or slow reaction time). These conditions are often progressive (such as Parkinson's disease) and thus drivers' abilities change faster than the healthy population. Therefore, it is critical to perform more frequent and convenient driving safety assessment of people with mobility challenges when using the mobility scooters.

In this paper, we focus on driving motion stability analysis for patients based on their posture data. Motion stability is an important metric for safety assessment, which indicates muscle ability and associates with the risks of falling. To assess patients' motion stability, most studies have mainly focused on the clinical setting, where patients are required to perform certain tasks [7]. However, for patients who cannot or are not willing to go to the clinics, these measurements cannot be applied. Our work enables stability analysis for patients' driving motions when they are at home or in the local community, which will increase the accessibility and timeliness of driving safety assessment in the tele-health setting.

Unlike existing works that rely on special sensors or devices (e.g., [5]), our work is leveraging posture data of patients which can be collected from the camera on a smart phone when driving, to perform stability analysis. To the best of our knowledge, there is no existing dataset of driving postures for people with mobility disabilities/challenges riding mobility scooters, for research or other purposes. Our work is the first in collecting such patients' data and using it to perform data-driven intelligent stability analysis.

In order to generate reliable stability analysis results, we build deep learning models which take sequences of video frames as input, and produce a preliminary analysis result, i.e., a loss value that reflects how unstable the postures are, which can also be combined with a threshold to generate the binary classification result of stable or unstable. The reason we choose to use this coarse-grained result is that the goal of the system is to monitor the driving safety of patients outside of hospital and detect any conditions that need further interventions or closer examinations by rehabilitation experts, neurologist, or other medical experts. More fine-grained stability scales can be measured at a later stage.

There are many existing works of gait analysis using deep learning models [31]. Body deep representations based on silhouettes and skeleton have been explored to perform tasks such as action recognition and identity detection (e.g., [14], [23]). Like gait analysis, we focus on the temporal features

This work is supported in part by grants NSF CISE-MSI #2318671 and Google exploreCSR Summer Research 2023.

of body movements by time series analysis, and also the spatial relationships among multiple body key points when moving. However, our work is different than most gait analysis works in that we focus on upper body movements instead of the full body, which will result in different choices in deep model architectures. Moreover, driving postures possess distinct characteristics such as the unique angle and distance that wrists and other key joints move during driving, yielding unique spatial and temporal features.

In this paper, to perform preliminary stability analysis for mobility scooter drivers with mobility challenges, we propose a deep autoencoder based system which is driven by the pose estimation result of driver's upperbody video recordings. In particular, we utilize MoveNet [18] to produce upper body keypoints 2D coordinates for each video frame. The resulting sequence of keypoints coordinates from the driving video frames labeled as stable will be used to train the recurrent neural network and convolutional neural network based autoencoder. Here the stability labels for driving video frames are generated by Kinesiologists who possess the domain expertise to distinguish between stable and unstable upper extremity movements in driving. After the autoencoder is trained, sequences of new video frames can be processed and analyzed, based on the loss values of the test sequences. In general, unstable driving video frames will yield in much higher loss because the input's features are distant from the embedding vector trained using the stable samples.

The contributions of this paper include the following:

- We design and develop a driving posture stability analysis system for people with mobility challenges, which can provide accessible safety assessment of the drivers at home or in the communities.
- Our deep autoencoder based approach extracts the spatial and temporal features of the movement of drivers' upper body key points which are generated by a pose estimation module.
- We explore two different architectures of the deep autoencoder, one having an LSTM based encoder and the other having a CNN based encoder, and implement the two models separately.
- We collect six patients' driving posture data on mobility scooters to form a pilot dataset, and run the system pipeline to test the system performance. The evaluation results of ROC and precision-recall show that our system has high levels of analysis accuracy.

The rest of the paper is organized as follows. In Section II, we review the existing works related to this paper, including car driving behavior analysis, time series analysis and pose estimation and deep autoencoders. Then in Section III, we describe the details of our method. Section IV presents our data collection procedure with patients' demographic information, data collection and preprocessing methods. After describing the evaluation and results in Section V, we conclude the paper in Section VII.

II. Related Work

A. Driving Behavior Analysis and Classification

Safety-related mobility scooter driving behavior analysis has mostly been conducted qualitatively in the Health research communities of Geriatrics, Rehabilitation and Disability, in the forms of participant reports and case studies [19], [29], [30]. Research shows that compared to driving experience, the physical cognitive and psychological factors of the drivers are affecting the driving behaviors of mobility scooter users in the longer term. Other contributing factors include distractions from the environment, road hazards and interference from pedestrians. There are very few works on quantitatively analyzing the driving skills of mobility scooter users. In [24], a driving operation logging system has been developed to collect driving data of steering and speed adjustment in an indoor environment, to assist in driving skill assessment. Our project will fill up the gap of data-driven mobility scooter driving behavior analysis based on recent advances in deep learning.

B. Time Series Analysis and Pose Estimation

To characterize driving behavior, time-series data analysis is often used. Conventional recurrent neural networks such as Long Short Term Memory (LSTM) [16] or Gated Recurrent Unit (GRU) [36] have been widely applied to many domains [15], [20], [33]. Pose estimation [13], [35], [38] has also been successfully applied for driver behavior detection. In [6], a Convolutional Neural Network (CNN) based framework was proposed for car driver's head localization and pose estimation on depth images. Pose estimation based on 2D images, depth data and 3D models [22], [25] have been applied to, for example, pediatric population, athletes, to track development, injury prevention and optimize performance [12], [17], [27]. However, validations of pose-estimation applied to people with motor challenges are under-explored [32]. The learning tasks in this paper leverage both pose estimation and time-series data analysis. Our work explores the specific applications of these models on real-world patient posture data, and validates the appropriate parameter settings.

C. Deep AutoEncoders

Deep autoencoder is a type of deep neural networks that generate an embedding vector that represents latent features of the input data through encoder layers, and the decoder layers transform the encoded features to the original input format, trying to minimize the loss between the decoded data and the input data [21], [34]. In autoencoders, layers are usually convolutional or recurrent neural network layers, to extract the latent spatial or temporal features. Most autoencoders' architectures are symmetrical (e.g., [39]), i.e., the decoder layers are the inverse/transpose of the encoder layers, even with the weights shared . There are also some works on autoencoders that yield good results with asymmetrical architectures [9], [37]. In our paper, we are going to study and compare the performance of these two types of architectures in mobility scooter driving behavior analysis, where input data is time series of upper body skeleton keypoints coordinates.

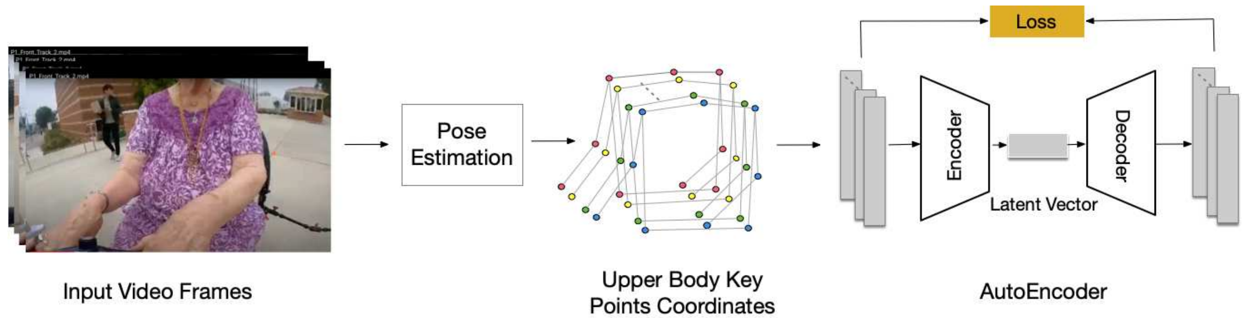


Fig. 1: Posture-based Mobility Scooter Driving Stability Analysis System Overview

III. Mobility Scooter Driving Stability Analysis

In this section, we first discuss our system pipeline, which takes video frames of patients' driving behavior as input and outputs a loss value for each frame. Then we will present in details two components of the system, upper body keypoints extraction and autoencoder-based stability analysis model.

A. System Pipeline

In this system, our goal is to analyze the mobility scooter drivers' upper body stability when driving. It is an important safety metric for people with mobility challenges and/or different medical diagnosis. We focus on upper body movement because unlike car driving, mobility scooters do not have breaks or gas pedals. Change of speed is controlled by speed potentio-meters which can be adjusted by hands.

To analyze drivers' upper body movements while driving, we rely on vision data, which is captured by cameras facing drivers' upper body. Videos frames in a sequence contain rich information of drivers' stability. We choose to apply pose estimation to extract the sequential skeletons from the video frames as the features we use for further analysis. The skeleton includes upper body keypoints coordinates, which is less sensitive to noisy variations such as the clothes that drivers are wearing compared to other segmentation results like object masks. Once a sequence of keypoint is extracted from a sequence of video frames, we use the stable driving keypoint coordinates as the input to the autoencoder structure for training. The encoder will generate a compressed vector which is a deep representation of the keypoint movement pattern within a time frame. The decoder will reconstruct features trying to minimize the loss between the output of decoder and the original input of encoder in training. In testing, the loss value of one given sequence's reconstruction will be used to determine the binary classification result, i.e., stable sequence or non-stable sequence. In Section III-C, we are going to introduce and compare two different architectures for this autoencoder, one symmetric and the other asymmetric, which include detailed descriptions of layer types and dimensions. Figure 1 illustrates the complete pipeline of our posture-based mobility scooter driving stability analysis system.

B. Upper Body Keypoints Extraction

To extract drivers' upper body keypoint information, we apply one open-source pose estimation model, MoveNet (lightning version) offered on Tensorflow Hub [18]. MoveNet detects 17 keypoints of a body, but we only use 9 of them, i.e., neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hip, and right hip. 2D coordinates of the 9 keypoints are generated for each input video frame with 30 frames per second. In MoveNet, first a person body heatmap is used to calculate the center. The initial prediction of keypoints coordinates is produced by slicing the keypoint regression output from the pixel corresponding to the center. Then weights are used for each keypoint which are inversely proportional to the distance to the center, so that the effect of keypoints from distant other objects can be reduced. In the last step of calculating the location of the maximum heatmap value, the local 2D offset predictions are added to refine the final 2D coordinate output for each keypoint.

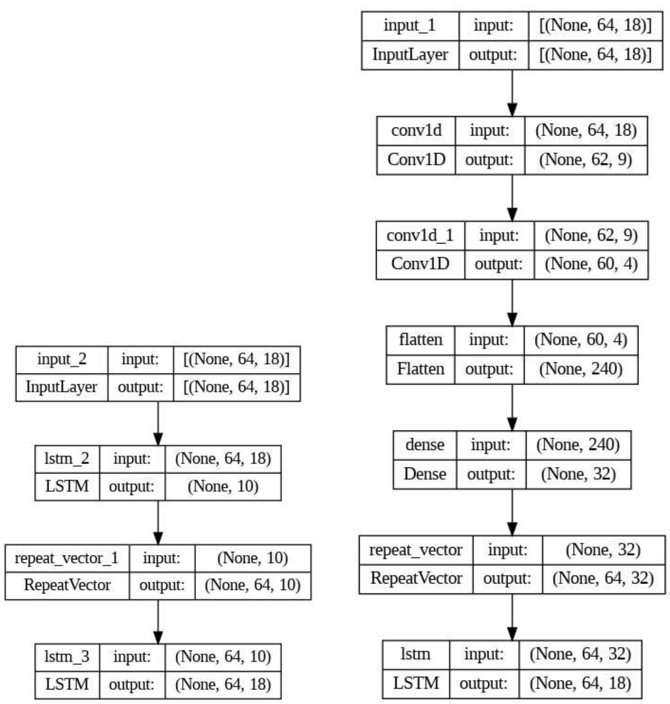
C. Driving Behavior Stability Analysis Deep Models based on AutoEncoders

To analyze the driving behavior stability, we need sequential data which can capture the characteristics of the upper body movement in a time frame when driving. To this end, we take a sequence of keypoints' coordinates at a time and make one stable-or-not decision for that sequence. The decision is enabled by a deep autoencoder model.

In this work, we study two different autoencoder architectures, where both architectures have an LSTM based decoder but the encoders incorporate different types of layers. In one architecture, we utilize LSTM to extract the latent temporal features of upper body driving among the sequence of keypoints coordinates, while in the other, the encoder is composed of a CNN network with one-dimensional convolutional layers. The second architecture leverages the spatial patterns in the 9 keypoints at different locations of the upper body when generating the embeddings. For both models, the goal is to minimize the distance between the reconstructed samples by the decoder and the original input to the encoder. We use the mean squared error (MSE) as the loss metrics for both models.

LSTM based symmetric AutoEncoder

In this model, the autoencoder has a symmetric architecture where an LSTM layer is deployed in the encoder part to extract an embedding vector that contains the features of upper body movements within the sequence of a certain time-step, from the input with the dimension of $time\_step * number\_of\_features$. Then the embedding vector is repeated for $time\_step$ number of times, before passing into the decoder LSTM layer. The output of the decoder is the reconstructed sample from the embedding vector, having the same dimension as the original input sample. Figure 2a illustrates a sample architecture of the LSTM based autoencoder, where $time\_step = 64$ considering 64 frames at a time in a sequence, and $number\_of\_features = 18$ with 9 joints' x, y coordinates. The embedding vector size is 10 in this example. In Section V, we will vary the size of embedding vectors and test their impact on the model accuracy.



(a) Sample LSTM based AutoEncoder Model Architecture

(b) Sample CNN-encoder - LSTM-decoder Architecture

Fig. 2: Two different AutoEncoder Model Architectures

Asymmetrical Architecture -

To capture the spatial features in the 9 keypoints, an asymmetrical model architecture is explored. In the encoder part, the input sample with the dimension of $time\_step * number\_of\_features$ will be processed by two consecutive 1D convolutional layers with ReLU as the activation function. Then a flatten layer is applied before a fully connected layer, which produces the embedding vector. After the embedding vector is repeated for $time\_step$ number of times, for the decoder, an LSTM layer is used to reconstruct the sample to the same size as input. In the example configuration of this architecture as shown in Figure 2b, $time\_step = 64$, $number\_of\_features = 18$. The first Conv1D layer has 9 filters, and the second Conv1D layer has 4 filters, both with filter size 3 and valid padding. The embedding vector size is 32.

IV. Patient Data Collection and Prepossessing

Our patient mobility scooter driving data collection is in collaboration with California State University Northridge (CSUN) Center of Achievement (CPP IRB protocol #IRB-22-88). Six Participants with mobility challenges have been recruited.

A. Patient Demographic Data

The demographic data details of the patients with mobility challenges are shown in Table I.

TABLE I: Patient Demographic Information

ID	*Age*	*Sex*	*ht*	*wt(lb)*	*Medical Condition*	*Impairment*
1	84	M	5'7"	145	Stroke	Right upper extr.
2	87	F	5'8"	200	Neuropathy	neck, trunk
3	53	M	6'1"	182	Brain Injury	Lower extr.
4	85	F	5'6"	230	Sciatica	Left lower extr.
5	90	F	5'7"	156	Arthritis	Both hips
6	62	M	5'7"	180	back&shoulder	Left upper extr.

B. Data Collection Methods

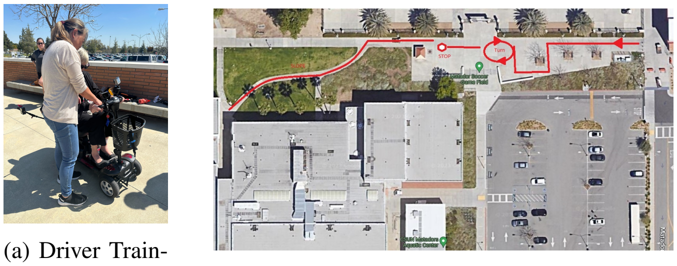
To collect the video recordings of patients' upper-body movements when driving, we mount an action camera on the handle, facing the driver. All video frames only contains upper-body parts below the neck, without face information being recorded by the camera for privacy protection. Participants are instructed to complete various driving tasks on a Drive Medical Phoenix LT 4 Wheel Mobility Scooter [3] on the CSUN campus after brief training. The list of driving tasks for participants is included in Figure 3. A sample picture of preparing the participants for driving on the mobility scooter and the participants driving route and are shown in Figure 4a and 4b respectively.

Rolls forwards (10 m)
Rolls backwards (2 m)
Turns while moving forwards (90 degree)
Turns while moving backwards (90 degree)
Turns in place (180 degree)
Gets through hinged door
Rolls 100 m
Avoids moving obstacles
Ascends 5degree incline
Descends 5-degree incline
Ascends 10-degree incline
Descends 10-degree incline
Rolls across side-slope (5 degree)
Rolls on soft surface (2 m)
Gets over gap (15 cm)
Gets over threshold (2 cm)
Ascends low curb (5 cm)
Descends low curb (5 cm)

Fig. 3: Tasks that Participants Perform for Data Collection

C. Data Prepossessing

Each frame of the videos (30 fps) is labeled by Kinesiologists into as Stable or Unstable based on the movement patterns of drivers' upperbody while driving. We have developed a video data annotation tool to accelerate the labeling process [2], which allows to generate the stability label for each video frame while playing the video only once.



(a) Driver Training (b) Participants Driving Routes

Fig. 4: Mobility Scooter Driving Behavior Data Collection

V. Implementation and Evaluations

In this section, after explaining the details of system implementation, we focus on illustrating the evaluation results of the system on analysis accuracy.

A. Implementation

The driving posture stability analysis system is developed in Python. It contains the following components: a per-frame stability labeling tool, MoveNet based pose estimation, deep autoencoder training, and the test module. In implementation, we have used the libraries: Tensorflow Keras [10], scikit-learn [28], opencv [26], and MoveNet. All experiments are performed on a high-performance computing cluster. The cluster has 20 DL160 compute nodes, and four GPU nodes with a total of 8 Tesla P100 GPUs. The cluster contains 3.3TB of RAM in total.

In total, in addition to the patients' driving data collected, we also collected stable driving video recordings from a group of healthy people (36 volunteers in the age range 18-36, with 30 males and 6 females), to enlarge the dataset of training the autoencoders. there are about 10 hours and 9 minutes of video footage in total and 128 minutes video footage of patients on the mobility scooter have been collected. After data cleaning by removing the frames without driving actions, in total we have labeled 847036 stable frames and 8084 frames as unstable, among which 606238 stable frames have been used for training and 240798 stable frames and 8084 unstable frames have been used for testing.

The pose estimation module produces a 2D-coordinate for each of the 9 upper body keypoints, so the number of features for the autoencoder input is $2*9 = 18$. We take a sequence of 64 frames at a time into the system. With fps being 30, each sequence covers about 2 seconds information of upper body movements in driving. For the CNN based encoder, the first 1D convolutional layer input size is $(64, 18)$, and with 9 filters of size 3 and valid padding, the ouput size is $(62, 9)$. The second 1D convultional layer has output size $(60, 4)$ with 4 filters of size 3. The dense layer following the flatten layer ouputs the embedding vector. After the repeating vector layer, the LSTM based decoder reconstructs the output of size $(64, 16)$. The loss function is MSE.

B. Experiment Metrics

Our driving posture stability analysis system produces the loss value of each test video frame by calculating the distance between reconstruction result based on the stable posture feature vector and the input frame. The larger the loss value, more unstable the driving posture is. To yield a binary classification result as either stable or unstable, a threshold is needed to define the cutoff. If the loss value is greater than the threshold, the input frame is tested as unstable; otherwise stable. To determine the threshold, it highly depends on the specification of system requirements. For different patients, medical experts may suggest different threshold with more background information of the patient being considered.

To evaluate the accuracy of our system, we apply the receiver operating characteristic (ROC) and precision-recall (PR) curve which show the performance of classification at all thresholds. We will also use the area under the curve scores (AUC) for the two autoencoder architectures we implement.

C. Evaluation Results

In our experiments, we would like to study the impact of different embedding vector sizes generated by the encoders on the system performance. We vary the embedding vector size from 16 to 512 and test the AUC values correspondingly for each of the two autoencoder architectures. The results are shown in Table II. For the LSTM encoder architecture, we can see that both the ROC AUC value and the PR AUC value reach the highest (0.852 and 0.994 respectively) when the embedding vector size is 128. For the CNN encoder, the highest ROC AUC score and PR AUC score (0.807 and 0.992) are both observed when the vector size is 256. This indicates while both architectures achieve excellent accuracy levels especially in terms of Precision-Recall, LSTM based encoder outperforms CNN based encoder in the stability analysis task, i.e., capturing the temporal features of keypoint coordinates is slightly more effective than the spatial features. Moreover, for both architectures, when the embedding vector size increases it improves the model accuracy level, until the vector size reaches 128 or 256. It means for this application and input size, vector of size 128/256 is the best in representing the features of upper body movements when driving mobility scooters.

TABLE II: ROC and PR Area Under the Curve Results

Vector Size	**16**	**32**	**64**	**128**	**256**	**512**
LSTM Encoder						
ROC AUC	0.811	0.8361	0.837	0.852	0.850	0.854
PR AUC	0.991	0.992	0.993	0.994	0.994	0.993
CNN Encoder						
ROC AUC	0.750	0.780	0.761	0.805	0.807	0.771
PR AUC	0.988	0.990	0.988	0.992	0.992	0.990

We also plot one ROC curve and one PR curve for each of two models when vector size is 256. In Figure 5, we can see that overall the LSTM encoder yields better results than the CNN encoder when plotting the (false positive rate, true positive rate) points for all possible thresholds. Figure 6 shows the Precision-Recall curves for both models. The precision values remain very close to 1 in both curves. The difference between the two models in this figure is minimum.

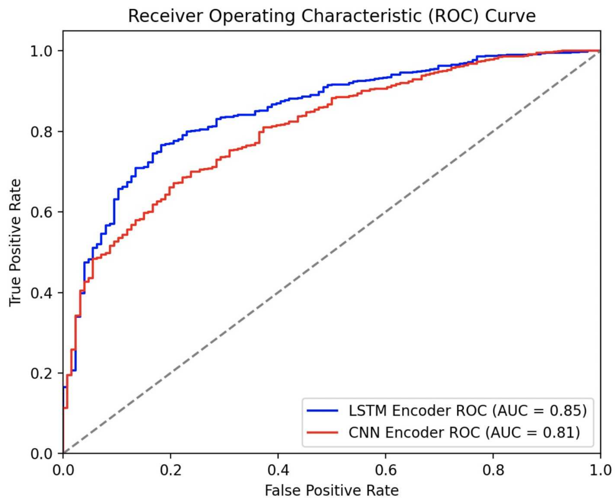


Fig. 5: ROC Curves of the Two Models when $Vector\_Size = 256$

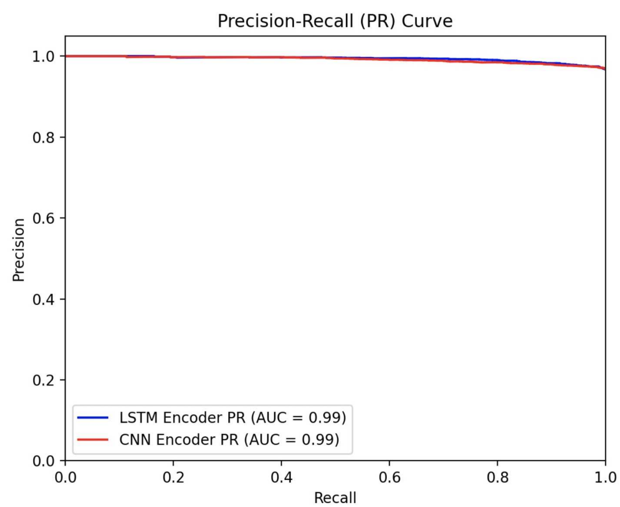


Fig. 6: PR Curves of the Two Models when $Vector\_Size = 256$

VI. Discussions

In this paper, the stability analysis test results of the input driving videos are produced on a high performance computing cluster. No system delay or time efficiency data has been collected yet. In practical scenarios, real-time stability analysis can be performed on user-end smart phones as an application or on Raspberry Pi based embedded systems attached to the vehicle, after the pre-trained autoencoder model is loaded. With more driving video samples being collected, the stability model can be continuously updated to yield more representative feature vectors for the mobility scooter drivers. In future works of system deployment, more experiments and evaluations will be conducted on system delay and the tradeoff between efficiency and accuracy will be studied in this application.

VII. Conclusion

In conclusion, our stability analysis system for mobility scooter drivers is the first effort in focusing on the upper-body posture movement for people with mobility challenges. Our system is easy to deploy as it only relies on videos collected from cameras. Deep autoencoder models have been built to extract the embedding features of upper-body movement represented by the sequence of keypoints' 2D coordinates in stable driving. We have collected patients' driving posture data and evaluated the system's accuracy for two different model architectures, i.e., the LSTM encoder and the CNN encoder. Test results are excellent in terms of ROC AUC score and Precision-Recall AUC score.

References

[1] Disability Impacts All of us. https://www.cdc.gov/ncbddd /disabilityand-health /infographic-disability-impacts-all.html.

[2] Driving stability labeling tool for upper extremely. https://github.com/Mobility-Scooter-Project/labeler/releases/tag/v0.2.4.

[3] Drive Medical Phoenix LT 4 Wheel Mobility Scooter. https://www.drivemedical.com/us/en/, 2022.

[4] Emilie S. Bakgaard. Mobility scooter accidents - need for preventative action? *Clinical Medical Reviews and Case Reports*, 4, 02 2017.

[5] Adrian Bingham, Xavier Hadoux, and Dinesh Kant Kumar. Implementation of a safety system using ir and ultrasonic devices for mobility scooter obstacle collision avoidance. In *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*, pages 1–5, 2014.

[6] Guido Borghi, Marco Venturelli, Roberto Vezzani, and Rita Cucchiara. Poseidon: Face-from-depth for driver pose estimation. pages 5494–5503, 07 2017.

[7] Sjoerd M Bruijn, OG Meijer, PJ Beek, and Jaap H van Dieen. Assessing the stability of human locomotion: a review of current measures. *Journal of the Royal Society Interface*, 10(83):20120999, 2013.

[8] Anna Carlsson and Jörgen Lundälv. Acute injuries resulting from accidents involving powered mobility devices (pmds)—development and outcomes of pmd-related accidents in sweden. *Traffic injury prevention*, 20(5):484–491, 2019.

[9] Xiangru Chen, Yanan Sun, Mengjie Zhang, and Dezhong Peng. Evolving deep convolutional variational autoencoders for image classification. *IEEE Transactions on Evolutionary Computation*, 25(5):815–829, 2020.

[10] Francois Chollet et al. Keras, 2015.

[11] US Consumer Product Safety Commission et al. National electronic injury surveillance system coding manual. 2019. 2020.

[12] Alessandro Crippa, Christian Salvatore, Paolo Perego, Sara Forti, Maria Nobile, Molteni Massimo, and Isabella Castiglioni. Use of machine learning to identify children with autism and their motor abnormalities. *Journal of autism and developmental disorders*, 45, 02 2015.

[13] Qi Dang, Jianqin Yin, Bin Wang, and Wenqing Zheng. Deep learning based 2d human pose estimation: A survey. *Tsinghua Science and Technology*, 24(6):663–676, 2019.

[14] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7297–7306, 2018.

[15] Zhongyang Han, Jun Zhao, Henry Leung, and King Ma. A review of deep learning models for time series prediction. *IEEE Sensors Journal*, PP:1–1, 06 2019.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[17] Zhanyuan Huang, Yang Liu, Yajun Fang, and Berthold K. P. Horn. Video-based fall detection for seniors with human pose estimation. In *2018 4th International Conference on Universal Village (UV)*, pages 1–4, 2018.

[18] Tensorflow Hub. Movenet singlepose lightning. https://tfhub.dev/google/movenet/singlepose/lightning/4.

[19] Michal Isaacson and Dov Barkay. Mobility scooters in urban environments: A research agenda. *Journal of Transport Health*, 18:100917, 2020.

[20] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33, 07 2019.

[21] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

[22] Felix Kuhnke and Jorn Ostermann. Deep head pose estimation using synthetic images and partial adversarial domain adaption for continuous label spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[23] Beibei Lin, Shunli Zhang, and Feng Bao. Gait recognition with multiple-temporal-scale 3d convolutional neural network. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3054–3062, 2020.

[24] Hiroshi Momma, Jun Suzurikawa, Rie Takeshima, Yuki Sawada, Yusuke Harada, and Tomoko Kondo. Assessment of driving skills of a mobility scooter using driving operation logs. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 3212–3215, 2020.

[25] Sankha S. Mukherjee and Neil Martin Robertson. Deep head pose: Gaze-direction estimation in multimodal video. *IEEE Transactions on Multimedia*, 17(11):2094–2107, 2015.

[26] OpenCV. *The OpenCV Reference Manual*, 2.4.13.7 edition, April 2014.

[27] Megumi Ota, Hiroshige Tateuchi, Takaya Hashiguchi, and Noriaki Ichihashi. Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm. *Gait Posture*, 85, 02 2021.

[28] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[29] Mike Prescott, William C. Miller, Jaimie Borisoff, Polly Tan, Nova Garside, Robert Feick, and W. Ben Mortenson. An exploration of the navigational behaviours of people who use wheeled mobility devices in unfamiliar pedestrian environments. *Journal of Transport Health*, 20:100975, 2021.

[30] Laura A. Rice, Alexander Fliflet, Mikaela Frechette, Rachel Brokenshire, Libak Abou, Peter Presti, Harshal Mahajan, Jacob Sosnoff, and Wendy A. Rogers. Insights on an automated fall detection device designed for older adult wheelchair and scooter users: A qualitative study. *Disability and Health Journal*, 15(1, Supplement):101207, 2022. Improving Community Participation for People with Mobility Disabilities.

[31] Alireza Sepas-Moghaddam and Ali Etemad. Deep gait recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):264–284, 2022.

[32] Jan Stenum, Kendra M. Cherry-Allen, Connor O Pyles, Rachel Reetzke, Michael F. Vignos, and Ryan T. Roemmich. Applications of pose estimation in human health and performance across the lifespan. *Sensors (Basel, Switzerland)*, 21, 2021.

[33] Odongo Steven Eyobu and Dong Han. Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. *Sensors*, 18, 08 2018.

[34] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.

[35] Wei Yang, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3073–3082, 2016.

[36] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, 31(7):1235–1270, 07 2019.

[37] Junbo Zhao, Michael Mathieu, Ross Goroshin, and Yann Lecun. Stacked what-where auto-encoders. *arXiv preprint arXiv:1506.02351*, 2015.

[38] Ce Zheng, Wenhan Wu, Taojiannan Yang, Sijie Zhu, Chen Chen, Ruixu Liu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. Deep learning-based human pose estimation: A survey. *Journal of ACM*, 37(4), 12 2018.

[39] Chong Zhou and Randy C Paffenroth. Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 665–674, 2017.