Driver behavior detection using ST-Gait++

MAXIM ROSCA, University of Twente, The Netherlands

Detecting driver behavior is crucial for enhancing road safety and developing intelligent transportation systems. Over 70% of accidents are attributed to human behavior, underscoring its significance in this field. This results in the need to understand human driving behavior to reduce this percentage. Computer vision is used to understand This research assesses the performance of ST-Gait++ model, originally designed to predict human emotion based on body position, for detecting driver behavior. The AIDE dataset is used and multiple models are trained, with various training inputs and configurations. Most of the trained models always predicted the same label, except for one that had greater accuracy and predicted two labels out of 3, instead of one.

1 INTRODUCTION

1.1 Motivation

Over 70% of accidents are attributed to human behavior, underscoring its significance in this field [1, 6]. Around 20,400 people were killed in road crashes in the EU last year. Although there are improvements in regulations, the numbers are still high and alarming.

A vehicle equipped with a driver behavior detection system can notify the driver to concentrate on driving based on the system's detection of distracted behavior. It can be also used to assist with semi-autonomous driving systems [23].

1.2 Aim

This research aims to develop a driver behavior recognition system using a human body graph representation and a model based on graph convolutional neural networks called ST-Gait++ [11]. This system aims to accurately analyze and classify driver behaviors such as normal driving, looking around, using the phone, talking, or dozing off based on the patterns within a graph representation of the human body. By using graph-based modeling such as ST-GCN [21], the research aims to make the effectiveness and robustness of driver behavior detection systems better, contributing to improved road safety and accident prevention.

1.3 State of the art

Several works were done previously regarding driver behavior detection using different techniques [2, 3, 15, 17, 20]. Although some of them have great results [2], the driver behaviors, such as Normal Driving or Using the phone, are somehow limited, detecting fewer activities than a driver would normally do. Other papers have better results compared to the state-of-the-art methods and use graph neural networks but the datasets that were used have flaws, for example, all the driving conditions are induced [17].

TScIT 37, July 5, 2024, Enschede, The Netherlands

1.4 Research question

How to train and adapt the ST-Gait++ for driver behavior detection? This question can be divided into two further sub-questions.

1.4.1 Training. For the ST-Gait++ to predict driver behavior such as Normal Driving, Looking Around, or using the phone training using different inputs is required. Finding the right input is no easy task and the first part of the research question focuses on: What data should be provided to ST-Gait++ to perform well on driver behavior recognition such as Normal Driving, Looking around, or using the phone?

1.4.2 Adapting. Once the data that works well is found, further improvement can be done to the model, by changing some of its parameters to get the best accuracy, this leads to the next sub-research question: What parameters and to which values should be changed to improve the accuracy of the ST-Gait++ for the task of driver behavior recognition such as Normal Driving, Looking around, or using the phone?

In the next sections, we will talk more in detail about the methodology used 3, the setup used for the experiments 4, and the results that we got after the model training 5. At the end, there will be a discussion section 5 and conclusions regarding the research 6.

2 SCIENTIFIC BACKGROUND

Computer vision. As a result of a large number of accidents, 1.35 million a year [14] Driver Monitoring Systems (DMS) were created. Currently, vision is the most effective and best source for gathering information [16], making possible the rapid development of DMS [8]. There are several commercial DMS that are based on measures from the car, such as steering [8], as well as scientific papers that use computer vision DMS [2, 13, 15].

Body language. Some researches successfully used deep learning and computer vision to encode body language for recognizing human emotions, suggesting that body expressions are a vital component in creating affect-aware technology [10, 19]. A common limitation in practical applications of body language and other affective features, such as facial expressions, gaze, gestures, and physiological indicators like heart rate and respiratory rate is that they require the user to be facing a camera, ensuring that these affective sources (e.g., face, eyes, arms) are continuously visible. A solution to this problem is to use the human pose as input for behavior detection, as it can be visible at all times.

ST-Gait++. Advances on Graph Convolutional Networks, especially the proposal of the ST-GCNs by Yan et al. [21] allowed for a very robust way to learn the spatiotemporal relationship. Based on that, ST-Gait++ was developed. ST-Gait++ is a skeletal trajectory classification model with graph convolutional networks originally used to predict human emotions, achieving great accuracy.

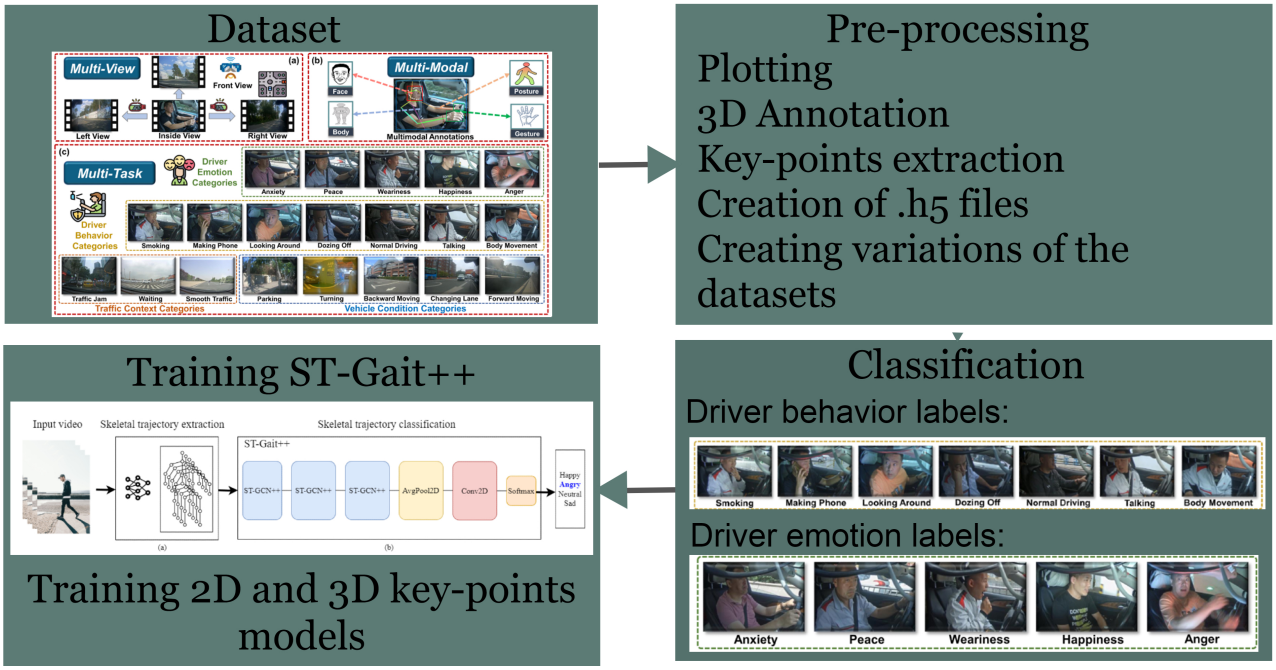

Fig. 1. Experimental setup pipeline

AIDE. AIDE captures rich information inside and outside the vehicle from several drivers in realistic driving conditions. It has three significant characteristics: multi-view (four distinct cameras used to capture inside and outside information), multi-task (different annotations such as emotion or driver behavior), and multi-modal (contains face, body, posture, and gesture information) as well as all the non-induced driving environment [22].

2.1 Related work

Several works have been done previously to detect driver behavior detection. This research [7] uses guidance sensors and gyroscopes to train models using the SVM machine learning approach. Using good sensors is expensive and are used and the attachment of sensors may cause discomfort [4], as well as might be inaccurate in certain conditions, such as line detection sensors on badly marked roads. This paper [24] proposed a Temporal Convolutional Network and four datasets were used to evaluate the proposed model, having a better accuracy than the state of the art by 2.24%. Another work [5] has achieved an accuracy of 94.7% for two-way recognition and 96% for three-way recognition using a deep learning architecture "DriverRep". One of the latest researches [2] focuses on multiple labels and achieves a 95% accuracy with 5 labels using a CNN-based model. This work [18] focuses on detecting fatigue based on vehicle speed, driver position, and lane position. All these works are either based on sensors, that might be expensive and hard to mount, use induced and limited datasets, or use a maximum of 5 labels.

3 METHODOLOGY

3.1 Methodology Idea

Based on the proven success of ST-Gait++ in human emotion recognition, we propose to adapt and train several models to evaluate its effectiveness in driver behavior recognition. The primary dataset utilized will be AIDE [22], chosen for its multimodal, non-induced nature and presence of emotion and driver behavior labels. We will experiment with models using both 2D and 3D coordinates. To annotate the AIDE videos, which initially contain only 2D key points, we will use AlphaPose [9] to generate 3D key points. Multiple models will be trained using these 3D coordinates. Additionally, to enhance the dataset's volume and diversity, we will incorporate video samples from the Drive and Act dataset [12]. The main evaluation metrics will include accuracy and confusion matrices.

Given a video $V \in \mathbb{R}^{N*H*W*3}$ with N frames, height H and width W, and a set of driver behaviors K or driver emotions E, our task is to classify the perceived driver behavior or emotion of a person present in such video by extracting features related to body language and joint position. We first extract a set of 2D or 3D key points $P \in \mathbb{R}^{16*3}$ or $P \in \mathbb{R}^{16*2}$, in which $k_1, k_2, ..., k_16$ and each k_i represents the location of a body joint in space related to the person in the video.

Skeletal trajectory extraction. The human body can be represented as a graph. Each body joint, such as the left elbow or nose can be considered nodes and the bones that connect these body parts can be considered edges. The ST-Gait++ receives as an input these graphs,

which are further processed using GCNs. At a given timestamp t, we extract the skeleton of the person visible in the scene and represent it as a graph: $G = (V, E)$, where V are the joints and E the edges and $N = |V|$.

Body Graph representation classification. This graph is further used as an input for the ST-Gait++ model. The input has the shape of $|V| * N * (K * 2)$ for 2D inputs and $|V| * N * (K * 3)$ for 3D inputs. ST-Gait++ consists of 3 ST-GCN++ blocks with 32, 64, and 64 kernels each, followed by an average pooling, a 2D 1x1 convolution layer, and a softmax layer for the 7 driver behaviors recognition or 5 driver emotions.

Model input. As previously mentioned, having C as the number of coordinates, the model receives as input the videos in the form of a matrix with the following size $|V| * N * (K * C)$. This input is fed to the model as a .h5 file containing multiple h5 datasets with this input, each having a specific key value. Each video V is a separate h5 dataset of the form $N * (K * C)$. The labels are stored in a separate h5 file that contains a number representing the label and has the same key as the features h5 file. To assess the performance of the model and find an input that would have a good accuracy multiple input files were created, with different combinations of X, Y, and Z coordinates, different numbers of labels, and different sizes of input, which are represented in the table below.

4 EXPERIMENTAL SETUP

4.1 Dataset

We used the AIDE [22] dataset to run the experiments. The dataset contains 2898 data samples with 521.64K frames. Each sample consists of 3-second video clips from four views, where the durations share a specific label from each perception task. There are 45 annotated frames in each video, all the frames having the same label. This dataset has a similar distribution of labels 4 as the previously used dataset to train the ST-Gait++. An experiment also includes data from the Drive and Act dataset [12]. The labels from the dataset are transformed, such that they are the same AIDE ones. Sitting still is transformed into normal driving, looking or moving around is transformed into looking around and talking on phone and interacting with phone is transformed into making phone. The train/validation/test split is of 7:2:1.

The AIDE dataset contains already extracted 2D skeleton key points and has multiple labels, for emotion, driver behavior, vehicle condition, and traffic context labels. The ones used by us are mainly the driving behavior labels: Normal Driving, Looking Around, Making phone, Body Movement, Talking, Smoking, Dozing off, and the emotion labels: peace, anxiety, weariness, happiness, and anger. Each sample is shaped $T * V$ with T being the number of times stamps and V being the number of coordinates. In our case, it's 48 or 32, 16 joints, each having 2 or 3 coordinates. The joints that are used are the ones from the upper body, which can be seen by the camera inside the car: nose, left eye, right eye, left ear, right ear, left shoulder, left elbow, right elbow, left wrist, left hip, right hip, head, neck and hip. A visualization of the joints in 3D can be seen in 2 as well as a visualization of the 2D key points in 3. To obtain the 3D key points from the AIDE images AlphaPose was used [9].

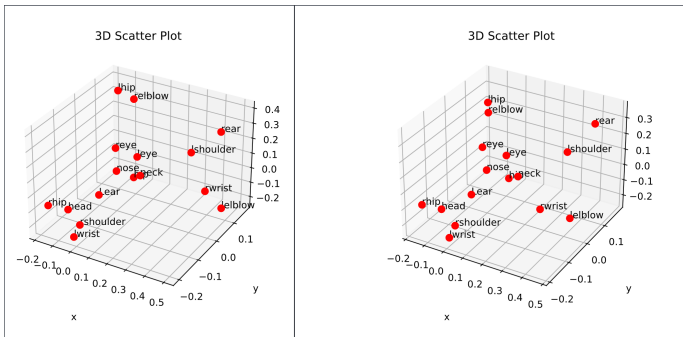

Fig. 2. Example 3D Representation of key-points from the same video, with very similar image

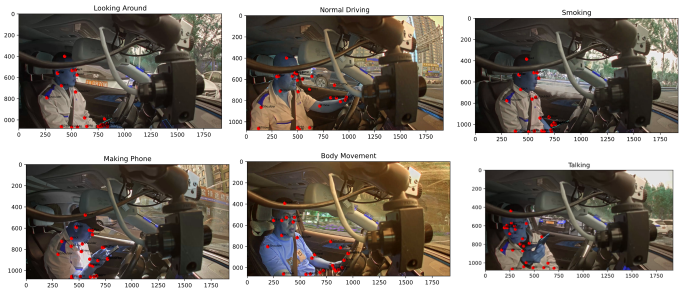

Fig. 3. Example 2D Representation of key-points

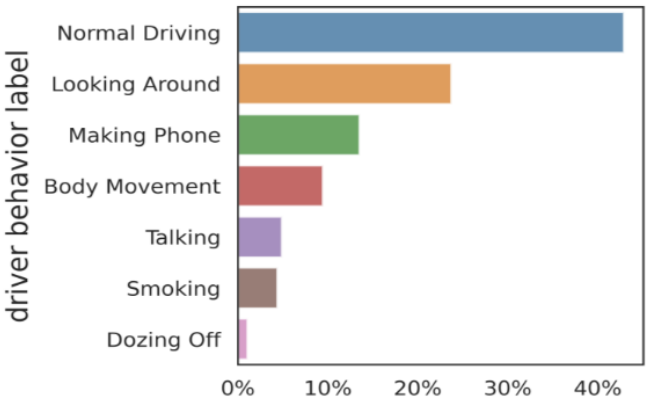

Fig. 4. Distribution of labels in dataset

4.2 Validation metrics

Accuracy metric is used, being used in previous ST-Gait++ research and being one of the most reliable and popular choices. To be sure that not the same labels are always given confusion matrix is used as well.

4.3 Implementation details

Environment. Python 10 is used in combination with multiple Ubuntu WSLs and Conda environments. This is necessary because

each of the used Python libraries has different dependencies, not only Python dependencies but also operating-system-specific dependencies. Everything was run on a machine with an AMD Ryzen 7 5800H processor and Nvidia GeForce RTX 3070 graphics card.

Data processing. AIDE data is structured in several JSON files, each file representing a video, and containing all the frame information, such as coordinates and labels. Each video has a single label for each category (behavior or emotion, for example). These coordinates and labels are transformed into h5 files. The AIDE dataset contains only 2D coordinates, while, initially the ST-Gait++ is set up for 3D keypoints. Because of that, the AIDE dataset needed to be re-annotated with 3D key points. This was done using AlphaPose [9]. Multiple models are trained, using the inputs mentioned in the previous paragraph. Before training the data was visualized as can be seen in 3 and 2. Only the required joints, the ones from the upper body are selected and, for each frame for a video a $N*(K*C)$ matrix is created, and for each frame an array of size $K*C$ is created. The format of the input is then compared to the original input used for ST-Gait++, in other successful research, using an h5 file visualizer and everything was similar. Furthermore, the data was re-plotted again from the h5 file and compared with the coordinates before writing them to the h5 file and it was the same. The same procedure was done with the labels.

In order to further enhance the data, another dataset, called Drive and Act [12] was used in combination with the AIDE dataset. The labels from the dataset are transformed, such that they are the same AIDE ones. Sitting still is transformed into normal driving, looking or moving around is transformed into looking around and talking on phone and interacting with phone is transformed into making phone. From these features and labels, new h5 files are created that are used as input for the model.

Model setup. The parameters for the model are present in the table 2. The setup was tested with the original input of ST-Gait++ and had nearly the same accuracy as claimed by the authors of approximately 87%. The original setup requires only 3D key points, that's why the X, Y, and confidence inputs were used initially. The model was afterward adapted for 2D key points, by changing some hardcorded values. For each of the inputs from 1, a model was trained and it's performance was assessed.

5 RESULTS AND DISCUSSION

The most important inputs and trained models can be seen in the table 1. Besides these multiple other models with slightly different inputs were tested, but were not included in the results due to loss best epoch files. The results are compared with the random probability of selecting the correct label out of 7 labels, which is 14.28%. Most of the trained models perform better than 40%. The high accuracy is due to the model always predicting the same label, this is confirmed by the confusion matrix 6. On the last day, a model was trained using only 3 labels for driver behavior: Normal Driving, Looking Around, and Making phone it resulted in an accuracy of 76.89% with two labels out of 3 being predicted. This model predicted correctly only two labels: Normal Driving and Using Phone, Looking around was not predicted correctly even once 5. The data was taken from both AIDE and Drive And Act. Another model with only 2 labels was planned to be tested, but there was not enough time. The possible reasons for these results are discussed below:

- Error in data processing. Although it was thoroughly checked, there could still be an error in the process of transforming the dataset into the model input
- Model configuration. The model parameters were left untouched, meaning that changing the parameters might have a good result.
- Imbalance in the dataset. Even though the AIDE dataset has a similar label distribution as the originally used dataset for training ST-Gait++, it might still be too imbalanced.
- Wrong annotation of the 3D key points. Visualization of the 3D annotations shows differences in key point's coordinates for very similar images, meaning that the coordinates might not be accurate and thus confusing the model.
- Not enough data was given to the model. The original ST-Gait++ training had a total of 466.000 frames and an input size of 2177 samples ∗240 frames ∗48 coordinates = 25.079.040 parameters. The AIDE dataset has a total of 130.000 frames and 2892 samples ∗45 frames ∗48 coordinates = 6.246.720 parameters.

5.1 Veryfing the pipeline

In order to test out what exactly was wrong a test was done to see if the pipeline, or at least some parts of it are correct. Previous research using ST-Gait++ [11] achieved high accuracy for human emotion recognition. Having a correct input for the model, we can transform it to the format in which the AIDE dataset [22] is given and run it through the pipeline to get the h5 file. If the files are identical, the model would give a similar accuracy as the initial model of 82.41%. This was done and the results were that the model had the same accuracy as previously. The only thing that was changed is the lack of points extraction, as the original dataset input already had 16 joints, and selecting the right joints from them is redundant. Also, visualization using an h5 file visualizer and heatmap for each dataset was used to compare them and the randomly selected samples were all the same.

5.2 Veryfing imbalance in the dataset

Even though the predominant label of "Normal Driving" has a high percentage of approximately 40%, in the original dataset the percentage of the predominant label was even higher, at 54%. To further test this assumption, datasets, using equalized percentages were used, meaning that "Normal Driving" for example was reduced to 30%, the same as "Looking around". This still gave the same result.

5.3 Discussion about wrong 3D annotation

During visualization, it can be seen that a very similar image, from the same video, has some of the key points, specifically the right wrist and left elbow 2. This could be a reason for the wrong output of the model. At the same time, the visualization of the 2D key points shows that they are correct, but the output of the model in 2D mode with 2D input gives the same output as the 3D mode. This could also be because of a mistake in adapting the ST-Gait++ for 2D

X	Y	Z	Input size	Sample size	Dataset	Output label	Accuracy	Otputs the same label
X	Y	Zero	45 * 16 * 3	2898	AIDE	Behavior	43%	Yes
X	Y	Confidence	45 * 16 * 3	2898	AIDE	Behavior	43%	Yes
X	Y	Z	45 * 16 * 3	2898	AIDE	Behavior	58.68%	Yes
X	Y	Z	45 * 16 * 3	2898	AIDE	Emotion	42.19%	Yes
X	Y	Z	45 * 16 * 3	6627	AIDE and DriveAndAct	Behavior	60.62%	Yes
X	Y	Z	45 * 16 * 3	6627	AIDE and DriveAndAct	Behavior (3 labels)	76.89%	No
X	Y	Z	45 * 16 * 3	2893	AIDE	Behavior (2 labels)	60%	Yes
X	Y	Z	45 * 16 * 3	2898	AIDE	Emotion	60.24%	Yes
X	Y	Z	45 * 16 * 3	3729	DriveAndAct	Behavior	75.67%	Yes
X	Y	Z	27 * 16 * 3	2898	AIDE	Behavior	42.19%	Yes
X	Y	Z	27 * 16 * 3	2898	AIDE	Emotion	42.19%	Yes
X	Y	Z	225 * 16 * 3	2898	AIDE (padding)	Behavior	42.19%	Yes
X	Y	-	45 * 16 * 2	2898	AIDE	Emotion	63.02%	Not sure

Table 1. Trained models and their inputs

Parameter	ST-Gait++
Train/Val/Test	7:2:1
Epochs	500
Batch Size	8
Optimizer	Adam
Basic LR	0.1
Momentum	0.9
Weight Decay	5e-4
GCN Initializer	Offeset

Table 2. Model parameters

annotation and the wrong generation of confusion matrices, which is discussed in the next sections.

5.4 Discussion about not enough data

As previously said, the data from only AIDE [22] has 4 times fewer parameters than the one originally used for ST-Gait++ [11]. To fix this the Drive And ACT [12] dataset was used to add more samples, around 3.000 for the three most predominant labels, resulting in a bigger amount of parameters than the original ST-Gait++ input. This still did not solve the problem, the output being the same label every time. Padding was also used to achieve a size of 240 frames for each video from the AIDE dataset, the 45 frames were repeated until 240 frames were achieved, and this still resulted in the same label prediction. On the other hand, the only model that did not predict the same label using more data and fewer labels, this combination resulted in the model starting to predict different labels and having higher accuracy. This might mean that more data with fewer labels might result in the model performing better and having a high percentage of being true.

5.5 Error in the generation of the confusion matrix

Using the confusion matrix generation algorithm from the ST-Gait++ always resulted in the prediction of the same label, even though the model did not predict the same label. This was discovered later in the research and was rectified. The 'Not sure' label from the models' table 1 is because the confusion matrix was generated before the error was discovered and the model is different from the others because it is adapted to use only 2D inputs. There was not enough time for rectifying the confusion matrix generation algorithm as well.

5.6 Limitations of the research

The main limitation of the research is the lack of time. The last-minute discovery of a working input would've been studied more and maybe a better result would be obtained. The previous discussions would be checked more thoroughly as well.
The body of the driver not being fully seen is a general limitation of all driver behavior research. The research only uses human skeleton representation, without including other factors, such as face, outside factors, and other aspects.

5.7 Answering the research question

In order to answer the question of "How to train and adapt the ST-Gait++ for driver behavior detection?" we will answer the two sub-research questions.

What data should be provided to ST-Gait++ to perform well on driver behavior recognition such as Normal Driving, Looking around, or using the phone?
Even though multiple inputs were tried to obtain good predictions from the ST-Gait++ we were not able to find an input that would have good results. For the ST-Gait++ probably more data is required or a reduction in the behavior labels.

What parameters and to which values should be changed to improve the accuracy of the ST-Gait++ for the task of driver behavior recognition such as Normal Driving, Looking around, or using the phone?
This question remained unanswered, as we could not find the right input for the model that would not predict the same labels, except at the end of the research. Without a proper input, the parameters could not be adjusted.

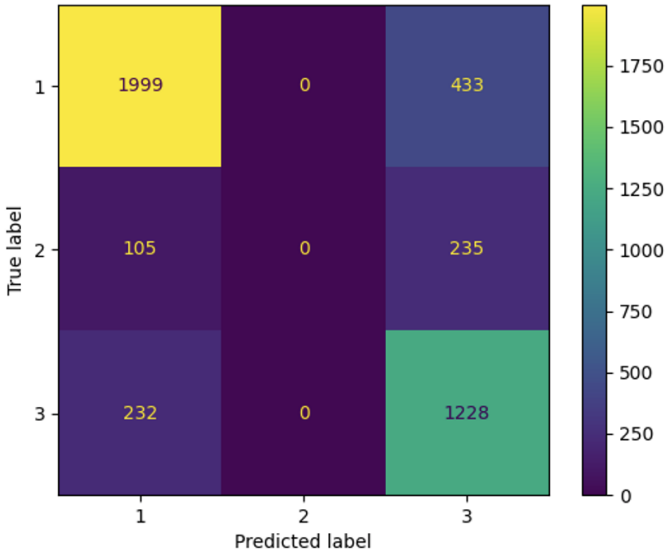

Fig. 5. 3 labels model confusion matrix (1- Normal Driving, 2 - Looking Around 3 - Making phone)

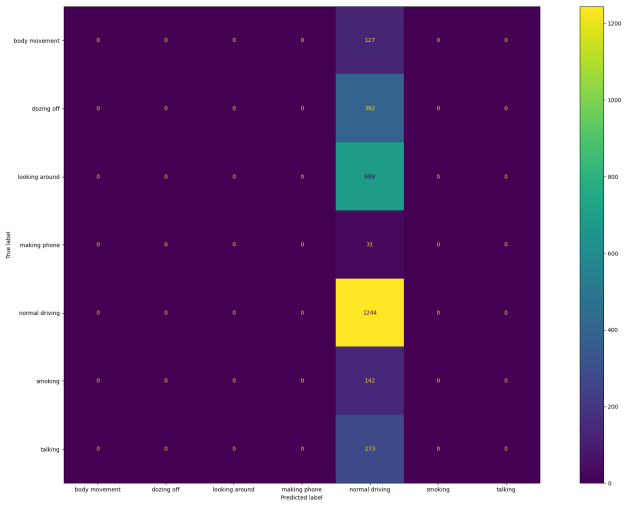

Fig. 6. 7 labels confusion matrix for Driver Behavior

6 CONCLUSION AND FUTURE WORK

In this research, it was tried to train the ST-Gait++ for driver behavior recognition such as Normal Driving, Looking around, or using the phone. It resulted in finding a single input file that would result in predicting different labels and a higher accuracy than the other inputs.

Future work needs to be done in order to find the correct input. Following the last-minute finding of a 3-label input, combining the AIDE and Drive and Act that did not predict always the same label other inputs might be discovered using different combinations of data and labels. More labels from the Drive And Act, besides the Normal Driving, Looking Around, and Using phone should be used as well as other datasets. A model using only two labels "Normal Driving" and "Abnormal Driving" can be tested and would probably have great accuracy. After the input is discovered the parameters of the ST-Gait++ can be adapted, by using for example hyperparameter search.

ACKNOWLEDGEMENTS

I would like to thank the following persons for their support and help in the research.

Estefania Talavera Martínez (supervisor) - for guidance, help, and advice during the research.

Willams de Lima Costa - for helping with advice regarding the process, next steps and ST-Gait++

Maria Luísa Lima - for helping with advice related to ST-Gait++

REFERENCES

[1] Pires Abdullah and Tibor Sipos. 2022. Drivers' behavior and traffic accident analysis using decision tree method. *Sustainability* 14, 18 (2022), 11339.

[2] Hamad Ali Abosaq, Muhammad Ramzan, Faisal Althobiani, Adnan Abid, Khalid Mahmood Aamir, Hesham Abdushkour, Muhammad Irfan, Mohammad E. Gommosani, Saleh Mohammed Ghonaim, V. R. Shamji, and Saifur Rahman. 2023. Unusual Driver Behavior Detection in Videos Using Deep Learning Models. *Sensors* 23, 1 (2023). https://doi.org/10.3390/s23010311

[3] Saif Al-Sultan, Ali H. Al-Bayatti, and Hussein Zedan. 2013. Context-Aware Driver Behavior Detection System in Intelligent Transportation Systems. *IEEE Transactions on Vehicular Technology* 62, 9 (2013), 4264–4275. https://doi.org/10.1109/TVT.2013.2263400

[4] Monagi H. Alkinani, Wazir Zada Khan, and Quratulain Arshad. 2020. Detecting Human Driver Inattentive and Aggressive Driving Behavior Using Deep Learning: Recent Advances, Requirements and Open Challenges. *IEEE Access* 8 (2020), 105008–105030. https://doi.org/10.1109/ACCESS.2020.2999829

[5] Mozhgan Nasr Azadani and Azzedine Boukerche. 2022. Driverrep: Driver identification through driving behavior embeddings. *J. Parallel and Distrib. Comput.* 162 (2022), 105–117.

[6] Arun Chand, S Jayesh, and AB Bhasi. 2021. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. *Materials Today: Proceedings* 47 (2021), 5135–5141.

[7] Zhongyang Chen, Jiadi Yu, Yanmin Zhu, Yingying Chen, and Minglu Li. 2015. D 3: Abnormal driving behaviors detection and identification using smartphone sensors. In *2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 524–532.

[8] Alaa El Khatib, Chaojie Ou, and Fakhri Karray. 2020. Driver Inattention Detection in the Context of Next-Generation Autonomous Vehicles Design: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 21, 11 (2020), 4483–4496. https://doi.org/10.1109/TITS.2019.2940874

[9] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. 2022. AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[10] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective Body Expression Perception and Recognition: A Survey. *Affective Computing, IEEE Transactions on* 4 (01 2013), 15–33. https://doi.org/10.1109/T-AFFC.2012.16

[11] Maria Luísa Lima, Willams De Lima Costa, Estefania Talavera Martínez, and Veronica Teichrieb. 2024. ST-Gait++: Leveraging spatio-temporal convolutions for gait-based emotion recognition on videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 302–310.

[12] Manuel Martin, Alina Roitberg, Monica Haurilet, Matthias Horne, Simon Reiß, Michael Voit, and Rainer Stiefelhagen. 2019. DriveAct: A Multi-modal Dataset for Fine-grained Driver Behavior Recognition in Autonomous Vehicles. In *The IEEE International Conference on Computer Vision (ICCV)*.

[13] Anthony Mcdonald, Thomas Ferris, and Tyler Wiener. 2019. Classification of Driver Distraction: A Comprehensive Analysis of Feature Generation, Machine Learning, and Input Measures. *Human Factors The Journal of the Human Factors and Ergonomics Society* (05 2019). https://doi.org/10.1177/0018720819856454

[14] World Health Organization. 2015. Global status report on road safety 2015. (2015).

[15] Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, and Mohammad Sabokrou. 2020. Driver behavior detection and classification using deep convolutional

neural networks. *Expert Systems with Applications* 149 (2020), 113240. https://doi.org/10.1016/j.eswa.2020.113240

[16] Michael Sivak. 1996. The Information That Drivers Use: Is it Indeed 90% Visual? *Perception* 25, 9 (1996), 1081–1089. https://doi.org/10.1068/p251081 arXiv:https://doi.org/10.1068/p251081 PMID: 8983048.

[17] Mingkui Tan, Gengqin Ni, Xu Liu, Shiliang Zhang, Xiangmiao Wu, Yaowei Wang, and Runhao Zeng. 2022. Bidirectional Posture-Appearance Interaction Network for Driver Behavior Recognition. *IEEE Transactions on Intelligent Transportation Systems* 23, 8 (2022), 13242–13254. https://doi.org/10.1109/TITS.2021.3123127

[18] Shigeyuki Tateno, Xia Guan, Rui Cao, and Zhaoxian Qu. 2018. Development of drowsiness detection system based on respiration changes using heart rate monitoring. In *2018 57th annual conference of the society of instrument and control engineers of Japan (SICE)*. IEEE, 1664–1669.

[19] Selvarajah Thuseethan, Sutharshan Rajasegarar, and John Yearwood. 2022. EmoSeC: Emotion recognition from scene context. *Neurocomputing* 492 (1 July 2022), 174–187. https://doi.org/10.1016/j.neucom.2022.04.019 Publisher Copyright: © 2022 Elsevier B.V..

[20] Xing Wei, Shang Yao, Chong Zhao, Di Hu, Hui Luo, and Yang Lu. 2022. Graph Convolutional Networks (GCN)-Based Lightweight Detection Model for Dangerous Driving Behavior. In *Wireless Algorithms, Systems, and Applications*, Lei Wang, Michael Segal, Jenhui Chen, and Tie Qiu (Eds.). Springer Nature Switzerland, Cham, 27–39.

[21] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018). https://doi.org/10.1609/aaai.v32i1.12328

[22] Dingkang Yang, Shuai Huang, Zhi Xu, Zhenpeng Li, Shunli Wang, Mingcheng Li, Yuzheng Wang, Yang Liu, Kun Yang, Zhaoyu Chen, et al. 2023. Aide: A vision-driven multi-view, multi-modal, multi-tasking dataset for assistive driving perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 20459–20470.

[23] Li Zhang, Guizhen Yu, Bin Zhou, Zhangyu Wang, and Guoyan Xu. 2019. Detection Algorithm of Takeover Behavior of Automatic Vehicles' Drivers Based on Deep Learning. In *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*. 126–130. https://doi.org/10.1109/ICITE.2019.8880230

[24] Yunyun Zhao, Hongwei Jia, Haiyong Luo, Fang Zhao, Yanjun Qin, and Yueyue Wang. 2022. An abnormal driving behavior recognition algorithm based on the temporal convolutional network and soft thresholding. *International Journal of Intelligent Systems* 37, 9 (2022), 6244–6261.