# Cyclist Maneuvers Prediction with Bicycle-mounted IMUs

ILJA KANISCEV, University of Twente, The Netherlands

Fig. 1. Safely driving cyclists on an intersection.

Cycling is one of the most common types of transportation in the Netherlands, accounting for about a quarter of all types of transportation used by the population in 2023 [1]. This type of transport continues to lead to deaths due to underdeveloped bicycle safety systems. Hence, the paper introduces new research to create an improved deep learning model regarding F1-score and advance notice time, predict cyclist maneuvers, and allow for a safe environment. Experiments were conducted using bicycle/cyclist-mounted IMUs and a smartphone device, which captured GPS data on 20 participants. This data is analyzed to identify potential indicative signs of pre-maneuvers. These signs are then utilized as contributing features in two prediction models: CNN-LSTM and CNN, expectantly predicting the maneuvers with substantial prior time and F1-score. Feature importance analysis is conducted to determine what features contribute most to prediction. Moreover, different combinations of prediction gaps and window sizes are tested on models to determine the optimal configurations. The final configured CNN model could predict the cycling maneuvers 1.2 seconds in advance and the window size of 1.6 seconds with an F1-score of 0.84 (see Fig. 9). The final configured CNN-LSTM model achieved an F1-score of 0.83, predicting 1.2 seconds in advance with the window size of 2.6 seconds (see Fig. 8).

Additional Key Words and Phrases: IMU, GPS, Maneuver Prediction, Bicycle, Deep learning, CNN-LSTM, CNN.

*TScIT 41, July 5, 2024, Enschede, The Netherlands*


## 1 INTRODUCTION

The Netherlands is famous for its well-developed cycling infrastructure, coupled with cultural preferences for sustainable transport solutions, which has resulted in a substantial portion of the population adopting bicycles. In 2023, bicycles accounted for more than one-quarter (28%) of all trips in the Netherlands. This indicates that there is a significant reliance on cycling as a mode of transportation within the country [1]. However, the recent emergence of speed pedelecs introduced new challenges to cyclist safety, especially in populated cities.

Speed pedelecs, electric bicycles capable of reaching higher speeds, have become increasingly popular among commuters, who seek faster travel times. However, their integration into urban travelling dynamics raises concerns regarding safety, particularly in interactions with traditional bicycles and other road users.

Hence, the research aims to address these challenges by developing a predictive deep learning model for cyclist maneuvers utilising Inertial Measurement Units (IMUs) [2] mounted on a bicycle and speed data collected via smartphones' GPS services. With a specific emphasis on identifying pre-turning features, the ultimate focus is to make bicycles smarter and safer for everyone.

In addition, to addressing these challenges, current study conducted enhanced experiments, in comparison to analysed previous researches. The experiments are conducted with mounted Inertial Measurement Units (IMUs) on bicycles which are capable of capturing crucial data such as acceleration, angular velocity, because of accelerometer and gyroscope integrated inside them. That is why the information gathered by such units can provide us with valuable insights into cyclist behavior. Building upon existing knowledge, such as handlebar counter-steering and the tilting of a frame before the actual maneuver [3], the research aims to explore new probable signs indicative of upcoming maneuvers.

These new signs may include the slowing down of speed, cessation of pedaling, and turning of the head before the maneuver, all of which will be measured using IMUs and speed data, and incorporated as additional indicators, which will probably assist towards understanding of the cyclist behavior preceding maneuvers.

The results of this research have the potential to not only improve cyclist safety but also contribute to advancements in V2X communication systems and improve current solutions.

## 2 STATE OF THE ART

There are multiple researches that have explored various solutions. One of them is a proposed method that utilizes the communication behavior features of bicyclists to predict their maneuver intentions at intersection approaches. This method involves the extraction of explicit and implicit communication cues from bicyclists using a bicycle simulator, focusing on their behavior at non-signalized intersections. An Intel RealSense D435 depth camera was used to collect the implicit and explicit communication features upon which AI model has been trained [4]. The result has the potential of predicting the trajectory of a bicyclist in traffic, but not explicitly predicting a maneuver in advance. Moreover, the experiments were conducted on a simulator, which does not represent real-world scenarios.

Another research was conducted in a natural setting using a wide-angle stereo camera system mounted at the research intersection of the University of Applied Sciences Aschaffenburg. The camera was mounted approximately 5 meters above the ground and used for movement detection and a probabilistic trajectory forecast [5]. Even though, the research achieved promising results in terms of high F1-score for the trajectory prediction, the solution is not predicting a maneuver in advance explicitly. Furthermore, suppose the current solution is widely adopted at intersections. In that case, each of them requires such a camera, which limits the safety regions drastically, as the ultimate goal is to make cycling safe at all times and places.

Moving on to more mobile solutions, [6] presented the solution of HeadMon, the maneuver prediction system, utilizing the head dynamics of a rider by installing an inertial measurement unit on the helmet. After experiments conducted with 20 participants, the results demonstrated an overall precision of at least 85% for a maneuver prediction under a 4 seconds prediction time gap.

In [3], the authors confronted the limitations and potential inaccuracies associated with predicting data solely from a head turns. Consequently, they explored a more reliable approach and enhanced it by utilizing IMU sensors mounted directly on the bicycle, namely its handlebar and frame. Through this method, they successfully predicted bike maneuvers with a substantial prediction gap, including left or right turns and cruising.

Their strategy involved analyzing both the bike's side tilt and the countersteering of the handlebars preceding a turn. By training a CNN-LSTM model on the data they collected, the researchers achieved a significant predictive accuracy, as evidenced by improvements in F1-score and advance notice time. This innovative approach not only addressed the shortcomings of previous methods but also offered a more coherent and effective means of anticipating bike maneuvers.

## 3 PROBLEM STATEMENT

Therefore, drawing inspiration from the findings of [3] and [6] researches, by utilizing additional IMU sensors, GPS, experimenting with different deep learning model layers, and conducting data collection experiment close to real life setting, this research is aimed to investigate other indicative features that can be measured, preceding a cyclist's maneuvers and if the prediction results can be improved in terms of F1-score and the advance notice time window. Hence, based on the above, the research question follows:

***To what extent can cycling maneuvers be predicted from bicycle/cyclist-mounted IMUs and speed data using deep learning?***

Sub research questions:

(1) *What are pre-maneuver indicators in a real-world setting and how to detect them using multiple bicycle/cyclist mounted IMUs?*
(2) *How to predict turn maneuvers with the identified indicative signs?*
(3) *To what extent and how early can the deep learning model predict cycling maneuvers based on F1-score?*

## 4 DATA COLLECTION

This section discusses the specifics and the tools used in the experiment setup.

### 4.1 Experiment setup

Specifically for this research a set of 4 IMU sensors was employed [7], along with a smartphone utilizing application for GPS data capturing, and an e-bike with a helmet. The IMU sensors were strategically positioned: one on the handlebar of the bike next to the smartphone to measure turning dynamics, another on the frame - to monitor bicycle's tilting movements, helmet - for capturing pre-maneuver head behaviors, and pedals - to detect any deviations from typical pedaling patterns, such as cessation before turning.

In addition to the above mentioned setup, a Full HD camera has been used in order to document each experiment and be further used for the manual labeling of the classified events (left, right turns and cruising) as the measure for the ground truth.

First, the camera was turned on and until the end of an experiment. Right after, GPS recording application was started on the smartphone sampling data at the rate of 1 Hz. Later, all 4 IMU sensors were synchronously being started, sampling the data with 200 Hz frequency.

Before, a participant of the experiment could start off, the handlebar was moved left-right-left in order to set an indicative starting point in the data records after which the data can be analyzed.

In total, there were 20 people taking part in the experiment. The bike seat and the helmet were adjusted to the personal preference


Fig. 2. Experiment setup

of a participant, ensuring a comfortable setup akin to their own bike in a real-world scenario. The speed constraint has not been set. It was allowed to cycle with an approximate maximum speed of 25 km/h due to e-bike software speed limitation. Participants were encouraged before the ride to maintain all the safety and road rules while conducting the experiment.

### 4.2 Experiment route

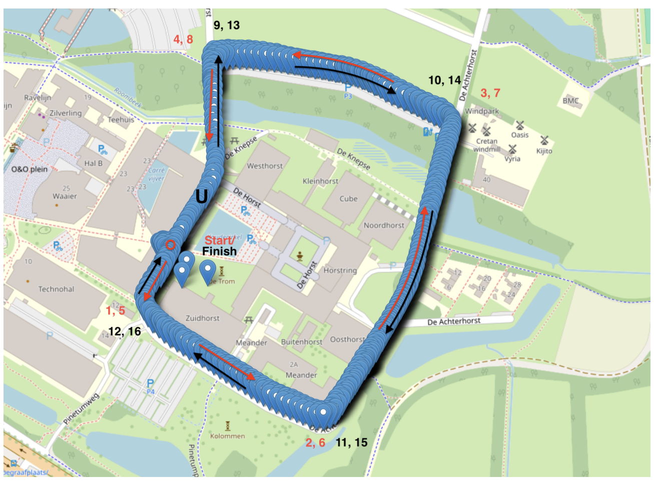



Fig. 3. Experiment route

The experiment was conducted on a circular route within the campus grounds of the University of Twente. The route included 16 turns (8 left, 8 right). Participants had to make 4 consecutive laps in total. First, 2 laps had to be cycled in one direction with 8 consecutive left turns. Later, they had to make a U-turn and repeat the same route in reverse direction with 8 consecutive right turns. The start and finish points were at the exact same location. This setup ensured the simplicity of the route, eliminating the need to have a virtual map attached to the handlebar, hence maximizing the participant's focus on the road and making the helmet sensor's result the most accurate in capturing natural head movements and pre-maneuver behaviors.

Table 1. Features (F)

| Helmet | Pedal | Frame | Handlebar | Smartphone |
|---|---|---|---|---|
| ax-515 | ax-559 | ax-560 | ax-558 | speed |
| ay-515 | ay-559 | ay-560 | ay-558 | |
| az-515 | az-559 | az-560 | az-558 | |
| gx-515 | gx-559 | gx-560 | gx-558 | |
| gy-515 | gy-559 | gy-560 | gy-558 | |
| gz-515 | gz-559 | gz-560 | gz-558 | |

Table 2. Data set: total samples obtained for each maneuver

| | All data | Training data | Testing data |
|---|---|---|---|
| **Left turn** | 160 | 112 | 38 |
| **Right turn** | 160 | 112 | 38 |
| **Cruise** | 160 | 112 | 38 |
| **Total** | 480 | 336 | 144 |

## 5 DATA ANALYSIS

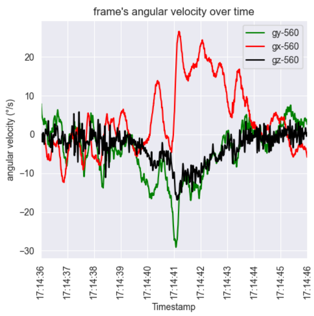
Once all the data has been collected, namely from 4 IMU sensors and the GPS application, all the files have been combined and synchronized using the timestamps for every participant separately. Due to frequency incompatibility, extracted speed data from GPS application for each participant have been linearly interpolated. The dataset resulted in having 25 features (F) (see Fig. 1).
6 features from each IMU sensor and additionally speed feature, extracted from GPS file.
'ax, y, z' columns represent acceleration around a corresponding axis in m/s$^2$. 'gx, y, z' columns represent rotational speed per second along a corresponding axis in °/s. 'speed' represents speed in m/s.

## 6 MANEUVER PREDICTION MODELS

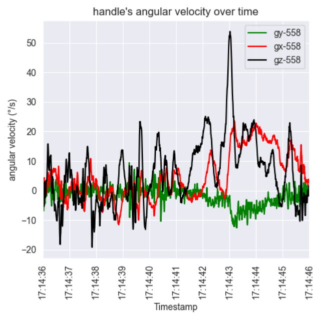
In this section the preparation of the data set is discussed as well as model architecture and their tuning parameters.

### 6.1 Events labeling

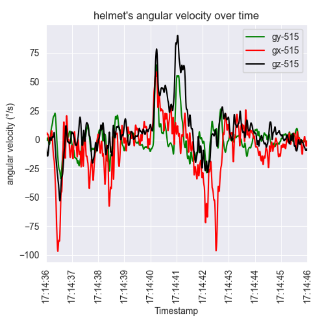
In preparation for training the dataset, the events were labeled manually by reviewing the video footage of the turns and cruises. The mid-maneuver reference point of a turn has been selected based on the peak of 'gz-560' (see Fig. (a) Frame under Fig. 4.) at the turning area. This selection is based on the assumption that the middle of a turn is defined as the point where the frame is tilted the most from its normal position [8]. Hence, the selected time point is considered to be a middle of a turn prior to which the prediction is expected to be done. Worth mentioning that no data calibration has been done from the start of a participant's trip.



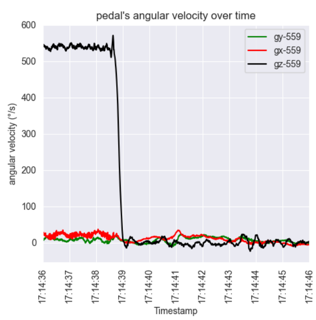

(a) Frame




(b) Handle




(c) Helmet




(d) Pedal

Fig. 4. Right turn raw data

Eventually, each participant's data file has been split to 3 individual ones, containing labeled only left turns, only rights turns and only cruises. Consequently, corresponding events were combined, thus the labeling resulted into 480 events total, with an equal distribution of classified events, namely 160 each in total. Training and testing sets were split with 7:3 ratio. (see Table 2)

## 6.2 Parameters

For each labeled event (right, left turn, cruise) a 10 seconds data flow was selected upon the determined mid point. The variations of the window sizes and prediction gaps resulting in significant F1-scores were interchanged within these 10 seconds time frames (see Fig. 5).

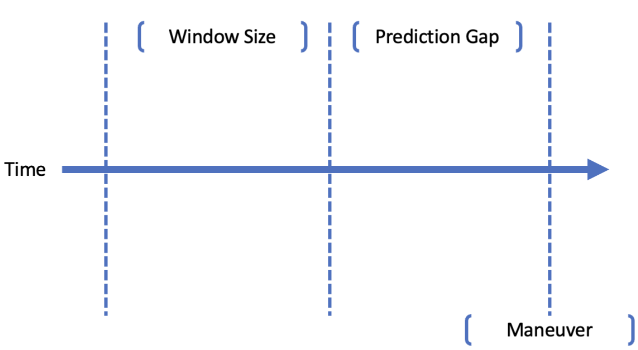



Fig. 5. Parameters

In order to comprehensively evaluate the performance of models and assess how it is affected by the trade-off between window size (W) and prediction gap (G), we considered 341 combinations. This involved training 341 models for each sensor combination with prediction gaps ranging from 0 to 2 seconds and window sizes ranging from 1 to 7 seconds, with a step value of 0.2 seconds. In order to facilitate the scale of the computational work, the cloud computing of the University of Twente has been exploited.

## 6.3 Models

For this research, two models have been used: CNN-LSTM with a similar structure (see Fig. 6) as proposed by Han et al. [6] and by Smit et al. [3], and CNN (see Fig. 7). The CNN-LSTM models were trained and evaluated on a dataset of cycling maneuvers captured with bicycle-mounted IMUs, focusing on predicting turn dynamics. Meanwhile, CNN was employed to explore alternative time series modeling approaches [9], comparing its performance against the CNN-LSTM models in terms of F1-score and computational efficiency.

Table 3. Hyper parameter Values

| **Parameter** | **Value** |
|---|---|
| Batch Size | 60 |
| Optimizer | Adam |
| (Initial) Learning Rate | 0.001 |
| Numpy Randomizer Seed | 42 |
| Loss | Categorical crossentropy |

Through out all the computational and training experiments, the same hyper parameters were used for both models(see Table 3).

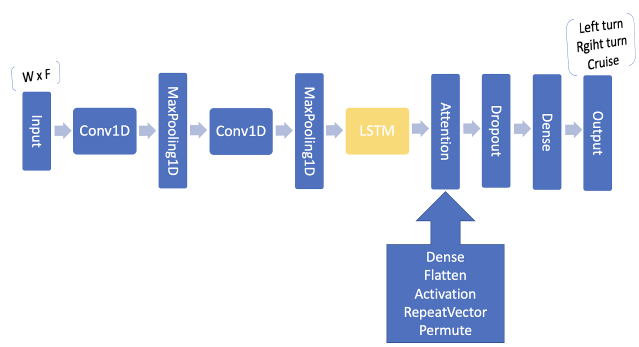



Fig. 6. CNN-LSTM model architecture

CNN-LSTM features with two 1 dimensional convolutional layers, each of them followed by max pooling layer with a pool size 2. First convolutional layer has 16 filters (output channels) and the second one has 2 filters. Both of them are with a kernel size of 7, activation 'ReLU' to introduce non-linearity, and padding 'same'. The purpose of those layers is to find spatial dependencies, whereas LSTM layer's purpose is to find temporal dependencies. LSTM layer is configured with dimensionality of 64, returning sequences over previously analyzed times steps, and activation 'tanh'. Importantly, the attention layer is used to prioritize and weigh different parts of the input sequence [10]. It uses Dense layer to compute attention scores with 'tanh' activation and 'softmax' normalization. Lastly, it uses drop out layer to prevent over fitting with an argument of 0.5.

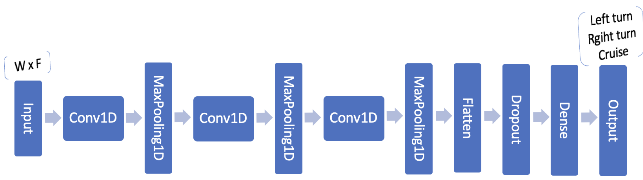



Fig. 7. CNN model architecture

CNN model architecture consists out of 3 pair of 1 dimensional convolutional layer followed by max pooling layer. Each convolutional layer has 64, 128 and 256 filters respectively with kernel size of 2, padding 'causal' and activation 'ReLu'. Later, flatten layers follows in order to output 1 dimensional vector, which is followed by dropout layer with an argument of 0.5 to prevent over fitting. Lastly, dense layer is present performing classification based on previously extracted features [11].

## 7 RESULTS

### 7.1 Different sensor combinations

In order to evaluate effectiveness in terms of F1-score performance and to assess the predictive lead time prior to an event, comparisons were made with previous research results, namely [6] and [3]. For this evaluation, only helmet features were present in the CNN-LSTM and CNN model training set, following the approach of Han et al. [6], while only handlebar features were considered in the the CNN-LSTM and CNN model training set, following the methodology of Smit et al. [3].

For further evaluation, strong performance results are characterized by attaining high F1-scores (close to 0.8 is considered to be sufficient), while also maximizing prediction gap and minimizing window size. Smaller window sizes facilitate faster model evaluation calculations. The higher prediction gaps, the higher chance of communicating the maneuver in advance.

### 7.2 Only helmet

By exploring different combinations of window size and prediction gap, and utilizing only helmet features (ax-515, ay-515, az-515, gx-515, gy-515, gz-515) and speed, the best results achieved for CNN-LSTM model were F1-score of 0.78 and predicted 1 second before the maneuver's mid-point with a window size of 2 seconds (see Fig. 13).

Moreover, as for CNN model, the best performance results were F1-score of 0.79, prediction gap of 1 second and window size of 1.2 seconds (see Fig 14).

Evaluating the performance, the results were not that promising as reported by Han et al. [6]. Possible that if higher prediction gaps evaluated, than CNN-LSTM model could have shown better results in terms of higher prediction gap with a decent F1-score metric.

### 7.3 Only handle

By exploring different combinations of window size and prediction gap, and utilizing only handlebar features (ax-558, ay-558, az-558, gx-558, gy-558, gz-558) and speed, the best results achieved for CNN-LSTM model were F1-score of 0.79 and predicted 0.4 second before the maneuver's mid-point with a window size of 2 seconds (see Fig. 19).

CNN model, however, achieved the best performance results with F1-score of 0.79, prediction gap of 0.4 second and window size of 1.2 seconds (see Fig 20). Important to notice that with higher prediction gap, CNN performed way worse with handlebar features than CNN-LSTM.

### 7.4 Other combinations

Other than combinations for evaluation for the previous researches comparison sake, combinations of only handlebar and pedal were explored, as well as handlebar and helmet.

Table 4. Other combinations: CNN-LSTM

| | **F1-score** | **Prediction gap (G)** | **Window size (W)** |
|---|---|---|---|
| **Handlebar-helmet** | 0.8 | 1.4 s | 3.8 s |
| **Handlebar-pedal** | 0.81 | 0.8 s | 3.4 s |

Table 5. Other combinations: CNN

| | **F1-score** | **Prediction gap (G)** | **Window size (W)** |
|---|---|---|---|
| **Handlebar-helmet** | 0.81 | 0.8 s | 1.4 s |
| **Handlebar-pedal** | 0.83 | 0.4 s | 1.4 s |

From the reported results (see Table. 4) and 5, it is evident that CNN-LSTM performed significantly better in terms of almost doubling prediction gap. However, this improvement came with a trade-off of computation time, as the window size was doubled too, while maintaining the same prediction F1-score.

### 7.5 Full sensors setup

Evaluating CNN-LSTM and CNN performance with all 4 sensors' and speed data, the following results are reported:

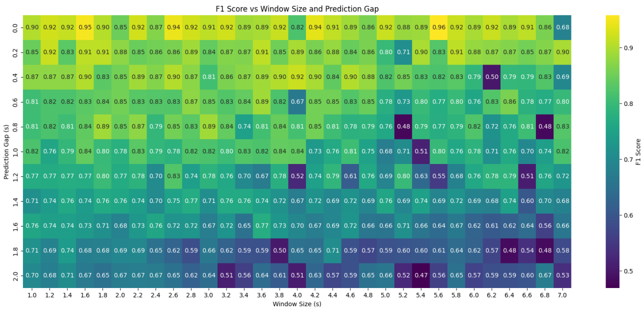



Fig. 8. CNN-LSTM with all sensors

CNN-LSTM obtained F1-score of 0.83 with a prediction gap of 1.2 seconds and window size of 2.6 seconds (see Fig. 8).

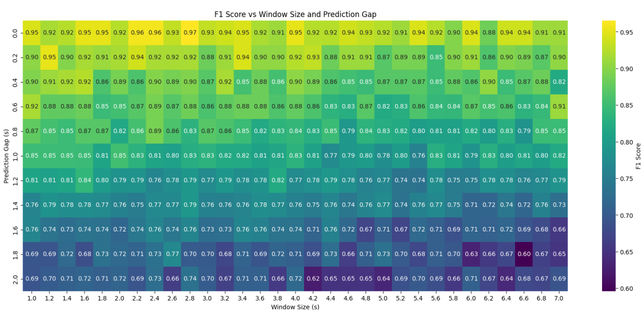
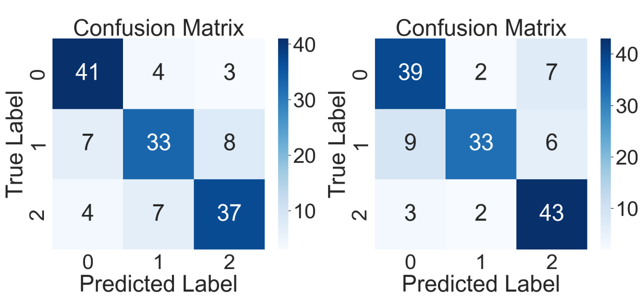



Fig. 9. CNN with all sensors

As for CNN, one of the best performance results was of 0.84 of F1-score with a prediction gap of 1.2 seconds and window size of 1.6 seconds (see Fig. 9). Comparing both best performance model results, CNN has an advantage of the computation time since the window size is less by 1 second .




(a) CNN-LSTM (b) CNN

Fig. 10. Confusion matrices of CNN-LSTM and CNN with all sensors (best performance)

### 7.6 Feature importance evaluation

In order to understand what feature contribute towards prediction the most and what is influencing it negatively, the feature permutation technique has been used. The CNN-LSTM model with all sensors set up (having 25 features) has been used with 2.6 second window size and 1.2 seconds prediction gap.

The baseline for the evaluation is the F1-score obtained during training the model with above mentioned parameters - 0.83 (see Fig. 8).

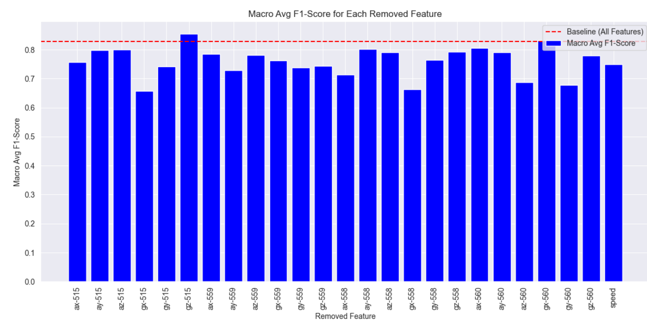



Fig. 11. CNN-LSTM with all sensors

Figure 11 demonstrates that the removal of gx-515, gx-558, az-560, and gy-560 results in a significant drop in the F1-score, indicating the importance of these features in the model's learning and prediction processes.
On the other hand, removal of gz-515 improves the F1-score, meaning that including this feature in the training set worsens the prediction performance.
Apart from negatively or positively influencing features, gx-560 turned out to have no influence at all, since by removing it, F1-score has not been changed at all.

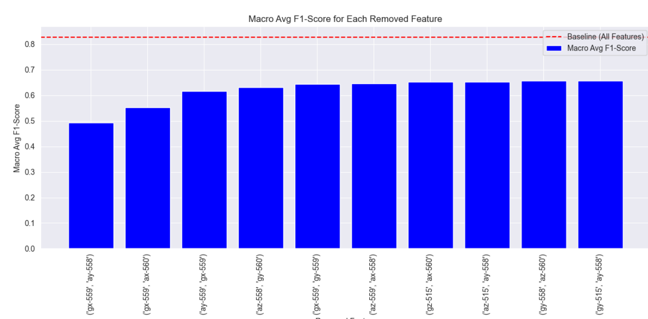



Fig. 12. CNN with all sensors

In addition to analyzing the model performance dependency on the removal of individual features, the top 10 most significant relationships are plotted (see Fig. 12). The graph should be interpreted as showcasing the most critical relationships contributing to a high F1-score. It is evident that the rotational speed of the pedals (gx-559) and the handlebar rotation acceleration (ay-558) combined are the most significant indicators. This suggests that a pedaling behavior and a handlebar turning prior to a maneuver are crucial for an accurate advance prediction.

The second crucial relationship is between pedals and a frame over x axis, proving that acceleration of a frame tilting combined with rotational speed of pedals are also significant behavior patters before a maneuver.

It is worth highlighting that the rotational speed over the z-axis of the helmet also plays a significant role in prediction when combined with a frame acceleration along the x-axis. This finding suggests that human side gazing prior to a maneuver helps to predict it more accurately.

Therefore, all sensor positions were found to provide useful indicators of behavioral patterns preceding a maneuver.

## 8 CONCLUSION

In conclusion, CNN-LSTM and CNN models were able to predict maneuvers of left and right turns and cruises, where input is bicycle-mounted IMU sensors and speed data in time windows. CNN-LSTM was able to predict maneuvers 1.2 seconds in advance using a window size of 2.6 seconds with an F1-score of 0.83. Another model CNN was able to predict maneuvers with 1.2 seconds in advance, with a window size of 1.6 seconds, achieving an F1-score of 0.84.

It is discovered that the rotational speed of the pedals and the handlebar rotation acceleration combined are the most significant indicators in order to make a prediction, followed by the pedal rotational speed and the frame acceleration over the axis.

Moreover, the rotational speed over the z-axis of the helmet, representing the person's gazes to the sides, showcased to be important too (see Fig. 12).

Compared to the research conducted by Smit et al. [3], the experiments were conducted in a natural traffic environment with people, bicycles, and cars being obstacles on the road and intersections. This implies that the prediction results achieved in this research are more accurate and significant, as the data was collected in a real-case scenario environment.

## 9 ACKNOWLEDGEMENTS

During the preparation of this work, the author used ChatGPT-4o in order to debug, familiarize with library documentation, and automate quick changes to exploited scripts for running the combinations of models. Grammarly software has been used to correct grammar and spelling mistakes. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

## REFERENCES

[1] Mathijs de Haas and Lukas Kolkowski, "Cycling facts 2023 Netherlands Institute for Transport Policy Analysis | KiM," Tech. Rep., 2023. [Online]. Available: https://english.kimnet.nl/publications/publications/2024/01/10/cycling-facts-2023

[2] "ProMove V - wireless inertial sensor from Inertia Technology." [Online]. Available: https://inertia-technology.com/product/inertial-sensor-promove-v/

[3] G. d. Smit, D. Yeleshetty, P. J. Havinga, and Y. Huang, "Predicting Turn Maneuvers of Cyclists Using Bicycle-Mounted IMU with CNN-LSTM," *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 587–592, 3 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10502681/

[4] Georgios Grigoropoulos, Patrick Malcolm, Fritz Busch, and Andreas Keler, "Predicting Bicyclist Maneuvers using Explicit and Implicit Communication," 10 2022. [Online]. Available: https://www.researchsquare.comhttps://www.researchsquare.com/article/rs-2108504/v1

[5] S. Zernetsch, H. Reichert, V. Kress, K. Doll, and B. Sick, "A Holistic View on Probabilistic Trajectory Forecasting - Case Study. Cyclist Intention Detection," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2022-June, pp. 265–272, 2022.

[6] Z. Han, L. Xu, X. Dong, Y. Nishiyama, and K. Sezaki, "HeadMon: Head Dynamics Enabled Riding Maneuver Prediction," *2023 IEEE International Conference on Pervasive Computing and Communications, PerCom 2023*, pp. 22–31, 2023.

[7] "phyphox – Physical Phone Experiments." [Online]. Available: https://phyphox.org/

[8] A. L. Schwab and J. P. Meijaard, "A review on bicycle dynamics and rider control," *Vehicle System Dynamics*, vol. 51, no. 7, 2013.

[9] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu, "Convolutional neural networks for time series classification | BIAI Journals & Magazine | IEEE Xplore," 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7870510

[10] Y. Qiao, K. Xu, and K. Zhou, "Research on time series based on improved LSTM," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications, ICPECA 2023*, 2023.

[11] Granat Margarita, "How to Use Convolutional Neural Networks for Time Series Classification | by Margarita Granat | Towards Data Science." [Online]. Available: https://towardsdatascience.com/how-to-use-convolutional-neural-networks-for-time-series-classification-56b1b0a07a57

## A RESULTS VISUALS

### A.1 Only helmet

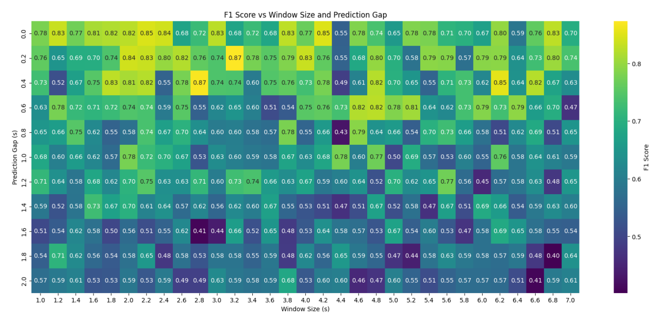



Fig. 13. CNN-LSTM helmet

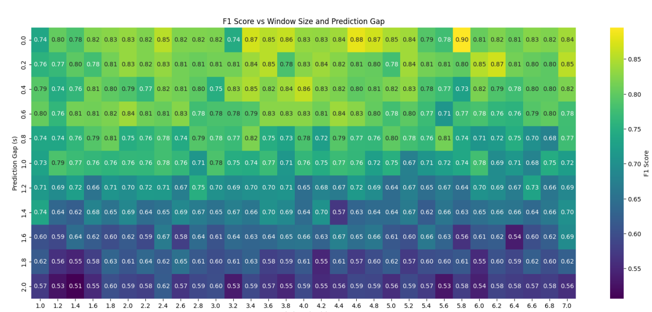



Fig. 14. CNN helmet

## A.2 Only handle

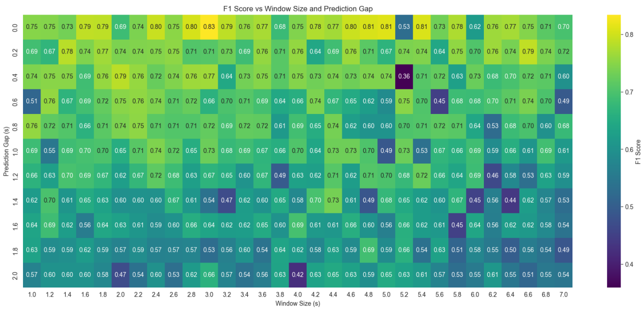



Fig. 15. CNN-LSTM handle

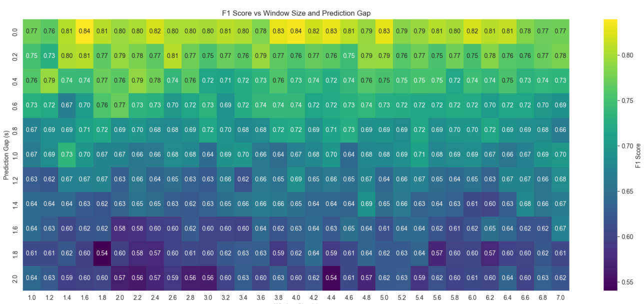



Fig. 16. CNN handle

## A.3 Handlebar-helmet

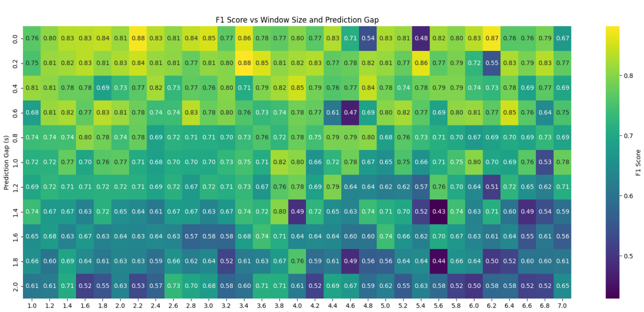



Fig. 17. CNN-LSTM handlebar-helmet

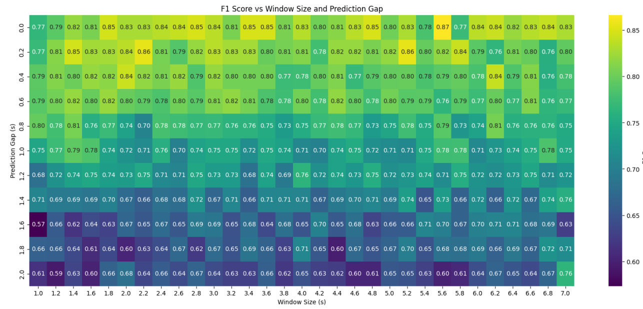



Fig. 18. CNN handlebar-helmet

## A.4 Handlebar-pedal

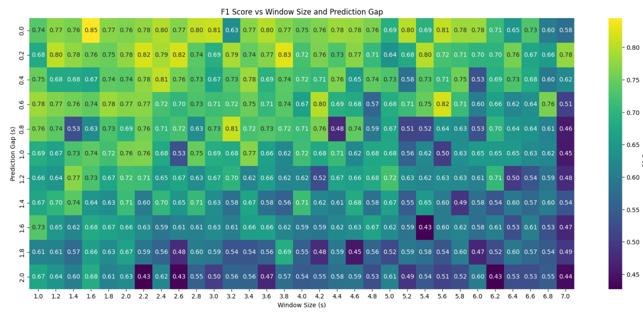



Fig. 19. CNN-LSTM handlebar-pedal

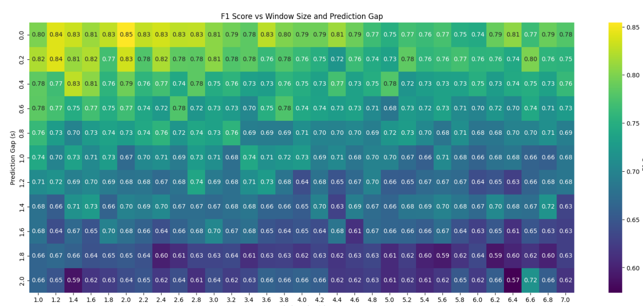



Fig. 20. CNN handlebar-pedal