Motion Prediction of Vulnerable Road Users at Signalized Intersections Using Lightweight Feed-Forward Neural Networks

Bence Jekl
HUN-REN Institute for Computer Science and Control
Kende u. 13-17., H-1111 Budapest, Hungary
Department of Control for Transportation and Vehicle Systems
Faculty of Transportation Engineering and Vehicle Engineering
Budapest University of Technology and Economics
Műegyetem rkp. 3., H-1111 Budapest, Hungary
bence.jekl@sztaki.hu

Dániel Fényes
HUN-REN Institute for Computer Science and Control
Kende u. 13-17., H-1111 Budapest, Hungary
daniel.fenyes@sztaki.hu

Balázs Németh
HUN-REN Institute for Computer Science and Control
Kende u. 13-17., H-1111 Budapest, Hungary
Department of Control for Transportation and Vehicle Systems
Faculty of Transportation Engineering and Vehicle Engineering
Budapest University of Technology and Economics
Műegyetem rkp. 3., H-1111 Budapest, Hungary
balazs.nemeth@sztaki.hu

***Abstract*—The accurate prediction of Vulnerable Road Users (VRUs) is essential for the safe and efficient operation of autonomous vehicles in urban environments. This paper proposes a lightweight Feed-Forward Neural Network (FFNN) for multi-step acceleration prediction of bicycles and e-scooters at signalized intersections using the IMPTC dataset. A frequency-domain analysis of VRU velocity profiles is performed to identify the dominant motion time scale, which guides the selection of the history window and prediction horizon. The model uses motion history and traffic light state information without requiring complex scene representations.**

Quantitative evaluation demonstrates that the proposed FFNN outperforms zero-acceleration and constant-acceleration baselines in both road and sidewalk scenarios. Cross-scenario evaluation confirms that scenario-specific training improves performance. The model contains only 1803 trainable parameters and achieves an inference time of approximately 0.03 ms per sample on a standard CPU, demonstrating very high computational efficiency and suitability for real-time automotive applications.

***Index Terms*—Acceleration, Prediction, VRU, Learning**

I. Introduction

Prediction of the future motion of Vulnerable Road Users (VRUs), including pedestrians, cyclists, and e-scooters, is an important task in the field of intelligent transportation systems and autonomous vehicles. Effective short-term acceleration prediction enhances trajectory planning, collision avoidance, and interaction handling in urban environments. Nevertheless, VRU motion is nonlinear, context-dependent, and affected by the behavior of humans.

Early motion prediction methods used physics-based or rule-based models; however, these techniques struggle to capture the complex and nonlinear behavior of real-life VRUs [1], [2]. With the increasing availability of data, deep learning became the dominant methodology for trajectory prediction tasks. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) were some of the early architectures that used sequence-to-sequence learning, to capture temporal dependencies [3], [4]. There are several surveys that discuss deep learning-based VRU trajectory prediction and highlight challenges i.e., uncertainty, interaction modeling, and scene understanding [5], [6].

Newer methods include environmental context along with historical motion data for better prediction accuracy. Models that include context make use of scene layout and scene information in the form of lane graphs and occupancy grid maps to ensure physically feasible trajectories [7], [8]. The environment plays a key role in the behavior of VRUs, such as the location of crosswalks, signalized intersections, and the phases of traffic signals [9], [10]. While such models achieve strong performance in complex urban scenarios, they often rely on computationally intensive rasterized or graph-based scene representations.

In contrast, feature-based approaches include chosen contextual elements directly as part of the input. With traffic light status and distance to stop line included as features, the impact of context can be captured efficiently without full-scene rasterization or graph construction. Although recurrent

The paper was funded by the National Research, Development and Innovation Office (NKFIH) under NKKP Grant Agreement No. STARTING 153675. The work of Dániel Fényes was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

and attention-based architectures dominate recent trajectory prediction research, Feed-Forward Neural Networks (FFNNs) remain computationally efficient and competitive for short- and mid-term prediction tasks [11]. When temporal information is encoded within a fixed-length input vector, multilayer perceptrons can approximate nonlinear motion dynamics effectively [12]. However, most existing approaches focus on trajectory prediction rather than direct multi-step acceleration forecasting.

The proposed FFNN-based framework is characterized by computational efficiency, stable training behavior, and ease of integration into real-time automotive applications. Unlike graph-based, attention-based, or recurrent architectures, it does not rely on rasterized scene representations or implicit sequence memory. Rather, it exploits motion and traffic-signal features together with a frequency-dependent temporal structure [13]. Although it does not achieve the same level of detail as large-scale transformer-based models, it provides a good balance between model size and predictive performance for context-aware acceleration prediction in intersection scenarios [14].

The remainder of this paper is organized as follows. Section II describes the dataset with spectral analysis and details the preprocessing steps. Section III presents the model architecture and training procedure. Section IV explains the results, and Section V concludes the paper.

II. Data Collection and Pre-processing

As an initial step, the dataset used in this study is briefly introduced. There are limited publicly available datasets that capture observations of vulnerable road users (VRUs), particularly pedestrians, cyclists, and e-scooter riders. In this research, the Interaction Multi-Person Trajectory and Context (IMPTC) dataset [14] is utilized, as it includes a representative number of e-scooter user cases.

The IMPTC dataset: The dataset was recorded in an intelligent public inner-city intersection in Aschaffenburg, Germany, with visual sensor technology. The full traffic and the road user behavior were perceived by a multi-view camera and LiDAR system operating at 25 Hz. The data acquisition system is focused on Vulnerable Road Users (VRUs) and multi-agent interactions. The sequences are recorded at various times of the day, across different seasons and weather conditions. The dataset provides high-quality trajectories of VRUs and vehicles, along with contextual information, e.g. the traffic light signal status. More than $2,500$ VRU trajectories are included, containing pedestrian, cyclist, and e-scooter rider data. The sequences have an average duration of $32.2s$, with the shortest and longest lasting 1.6 and $137.8s$, respectively.

Within the synchronized reference timestamp data, each dataset includes the VRU class with associated probability, three-dimensional coordinates expressed in the intersection coordinate system in meters, object velocity, ground classification of the track, and source sensor information. Nine traffic-light signal groups are tracked, each represented by one of six light states: red, red-yellow, yellow, green, yellow-blinking, and off.

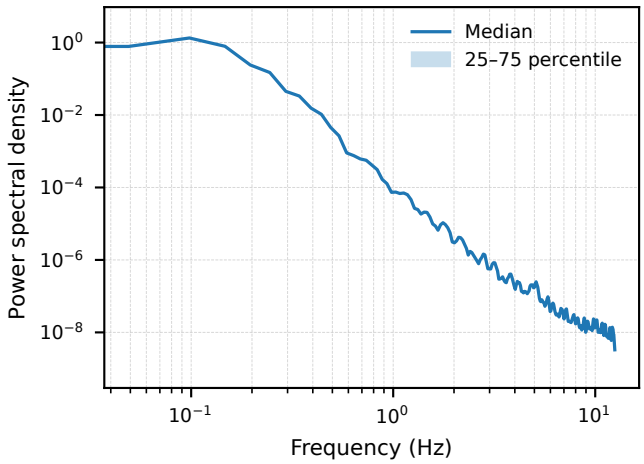


Fig. 1. Power spectral analysis

A. Spectral Analysis of VRU Velocity Profiles

To analyze the temporal dynamics of the VRU motion, the power spectral density (PSD) of each velocity signal was estimated using Welch's averaged periodogram method [13].

In Fig. 1, the median PSD and interquartile range are shown for all VRUs. The spectrum is displayed on logarithmic frequency and power scales. Most spectral energy is concentrated at low frequencies, and the PSD decreases rapidly with increasing frequency, indicating smooth and slowly varying motion patterns. The narrow percentile band suggests consistent spectral characteristics across the dataset.

The dominant frequency was defined as the frequency with the maximum spectral power for each VRU. The 95^{th} percentile of the dominant frequency distribution is 0.164 Hz, corresponding to a dominant period of 6.10 s. This shows that the velocity dynamics are dominated by low-frequency components, i.e., the motion dynamics of the VRUs are slow.

B. Preprocessing

After filtering the dataset, additional motion and contextual features were generated for each VRU trajectory. Acceleration was obtained using a finite difference approximation of the velocity signal. Orientation was determined from consecutive position differences to represent instantaneous heading. For each trajectory, the traffic light relevant to the VRU was identified by determining whether the VRU path intersected a manually defined stop line associated with a specific traffic signal. The signed distance to the relevant stop line was computed by projecting the VRU position orthogonally onto the stop-line segment. Positive values indicate positions before the stop line, while negative values indicate positions after crossing. Finally, each VRU was classified as moving on the road or on the sidewalk based on the type of traffic light it crossed. This classification enables scenario-specific modeling of motion behavior.

III. Methodology

A fully connected Feed-Forward Neural Network (FFNN) was implemented to predict the motion of the VRUs using historical motion data and contextual traffic-light information. The FFNN was selected for its effectiveness with fixed-dimensional feature vectors and its ability to model nonlinear relationships between motion dynamics and traffic signal states. Given that the input is a fixed-length window of features rather than raw variable-length sequential data, a fully connected architecture provides a deterministic structure with moderate complexity and stable training behavior.

A. Network architecture

The proposed FFNN model is designed to map a fixed-length historical feature vector to a multi-step acceleration prediction. The network processes a flattened representation of sequential multi-feature inputs. The historical time window is transformed into a fixed-dimensional vector suitable for fully connected layers.

The input layer size d_{in} is determined by the number of selected features per timestep n_{feat} multiplied by the number of historical time steps n_{history}. The input vector is processed through two fully connected hidden layers with d_{hidden} neurons and ReLU activations. Dropout with probability p_{dr} is applied after the second hidden layer to improve generalization. The output layer maps the final hidden representation to a vector of size $n_{\text{future}} + 1$, where n_{future} denotes the number of future time steps for which acceleration is predicted. No activation function is applied in the output layer, allowing the network to predict both positive and negative acceleration values.

B. Feature representation

The features at each historical time step include the VRU speed, the distance to the relevant traffic light stop line, and the past and future traffic light states encoded as discrete values. The VRU speed and stop-line distance values are normalized by $v_{\max}$ and $s_{\max}$, respectively.

For each VRU, the historical window contains the last n_{history} samples taken with a fixed sampling interval $\Delta t_{\text{history}}$. The resulting sequence is flattened into a single vector, forming the input of the FFNN at time step t:

$$\mathbf{x}_t = \left[f_i^{(t-k\Delta t_{\text{history}})} \right]^{\top}_{\substack{i=[1,\dots,n_{\text{feat}}] \\ k=[0,\dots,n_{\text{history}}-1]}} \tag{1}$$

where $f_i^{(t)}$ denotes the i-th feature at time step t, and n_{feat} is the total number of features per time step.

C. Output definition

The FFNN model is designed to predict the future acceleration of the VRUs. The network output at time step t is a multi-step prediction vector containing normalized acceleration values at uniformly spaced future time instants. Specifically, the model predicts accelerations for the next n_{future} time steps with a fixed time interval Δt_{future}:

$$\hat{\mathbf{a}}_t = \left[a^{(t)}, a^{(t+\Delta t_{\text{future}})}, \dots, a^{(t+(n_{\text{future}}-1)\Delta t_{\text{future}})} \right]^{\top}, \tag{2}$$

where $a^{(t)}$ denotes the normalized acceleration at time step t.

D. Training setup

The model was trained to minimize a weighted regression loss between the predicted and ground-truth acceleration sequences over the defined prediction horizon.

A linearly decaying weight was applied across the prediction horizon in order to assign higher importance to acceleration estimates closer to the current time step. The weight decreases linearly from 1 for the first predicted step to w_{last} for the final step of the horizon.

The loss function is formulated as a weighted Mean Squared Error, where each predicted acceleration value is multiplied by its corresponding horizon weight.

The dataset was split into 70% training, 20% validation, and 10% test subsets. Model parameters were optimized using the Adam optimizer with a learning rate of η. The training was performed using mini-batches of size N_{batch} to balance computational efficiency and gradient-estimation stability. This enables faster convergence while maintaining reliable updates to the network weights. Early stopping was used to prevent overfitting and to avoid unnecessary computations. Training is terminated when the validation loss does not improve for 30 consecutive epochs; the model parameters corresponding to the lowest validation loss are retained.

E. Optimal hyperparameter selection

The FFNN was configured using a structured hyperparameter selection procedure that balances predictive performance and computational efficiency. The optimal parameters were determined using the validation-set multi-step prediction loss and the test-set average root-mean-square error (RMSE) of the predicted acceleration. This subsection details the selected hyperparameters.

The feature space and output targets were normalized based on the largest observed values for each parameter in the dataset: $a_{\max} = 6$ m/s^2, $v_{\max} = 40$ km/h, $s_{\max} = 60$ m. Here, $a_{\max}$ refers to the maximum acceleration, $v_{\max}$ to the maximum velocity, and $s_{\max}$ to the maximum distance to the traffic light stop line.

The input vector contains $n_{\text{history}} = 11$ historical time steps, including the current step, each represented by $n_{\text{feat}} = 4$ features, as discussed previously. The historical observations are sampled at intervals of $\Delta t_{\text{history}} = 0.64s$. The selected history length $(n_{\text{history}} - 1)\Delta t_{\text{history}} = 6.4s$ slightly exceeds the dominant period of $6.1s$, ensuring that one full characteristic motion cycle is observed by the model.

The output vector contains predictions for the current step and the next $n_{\text{future}} = 10$ time steps. Future samples are generated at intervals of $\Delta t_{\text{future}} = 0.60s$. The prediction horizon of $6.0s$ is approximately equal to the dominant period.

The hidden layer size $d_{\text{hidden}} = 32$ was selected to provide sufficient nonlinear modeling capacity while maintaining real-time feasibility. With an input dimension of $d_{\text{in}} = 44$, the chosen hidden size corresponds to a moderate compression ratio ($\frac{d_{\text{hidden}}}{d_{\text{in}}} \approx 0.73$), enabling feature abstraction without excessive parameter growth. The resulting network has 1803 trainable

TABLE I
LOCAL SENSITIVITY ANALYSIS AROUND THE SELECTED CONFIGURATION

Modified hyperparameter	Modified Value	Δ Validation Loss (%)	Δ Test RMSE (%)
$(n_{\text{history}}, \Delta t_{\text{history}})$	$(6, 1.28)$	132.04	9.86
$(n_{\text{history}}, \Delta t_{\text{history}})$	$(21, 0.32)$	245.25	1.76
d_{hidden} (low)	16	55.42	1.06
d_{hidden} (high)	64	-3.90	0.35
η (low)	10^{-6}	137.63	1.06
η (high)	10^{-4}	64.92	3.52
N_{batch} (low)	32	103.39	-0.35
N_{batch} (high)	128	-5.42	4.93
p_{dr} (low)	0.1	-7.80	8.45
p_{dr} (high)	0.4	59.83	5.99

parameters, which provides sufficient representational capacity while maintaining computational efficiency.

The model was trained for $n_{\text{epoch}} = 120$ epochs with a learning rate of $\eta = 10^{-5}$, batch size $N_{\text{batch}} = 64$, dropout probability $p_{\text{dr}} = 0.2$. By tuning the final prediction step weight $w_{\text{last}} = 0.1$, the relative importance of short- and long-horizon predictions can be modified according to the specific requirements of the use case.

Table I summarizes the local sensitivity analysis around the selected configuration. Each hyperparameter is individually increased and decreased while keeping all remaining parameters fixed. The table shows the relative percentage deviation of validation loss and test RMSE from the optimal configuration. The reference values are a validation loss of 0.0059 and a test RMSE of $0.284 m/s^2$. All reported percentages are calculated with respect to these baseline metrics.

Reducing the history window to $(6, 1.28)$ increases both validation loss and test RMSE, indicating insufficient motion information. Selecting it to $(21, 0.32)$ further increases the validation loss, indicating that finer discretization adds redundant information without improving predictive accuracy.

Variations of the hidden dimension show moderate sensitivity. A reduced hidden size increases validation loss due to limited representational capacity, while a larger hidden size does not improve test performance, indicating limited benefit from increased model complexity.

The learning rate strongly affects performance. A lower value leads to higher validation loss due to slower convergence, whereas a higher value slightly increases validation loss relative to the selected configuration, indicating reduced training stability.

The batch size mainly influences optimization behavior. A smaller batch size increases validation loss due to higher gradient variance, while a larger batch size improves validation loss but slightly increases test RMSE, suggesting reduced generalization.

Finally, dropout probability controls regularization strength. Lower dropout increases validation loss due to overfitting, while higher dropout decreases performance due to excessive regularization.

TABLE II
QUANTITATIVE PERFORMANCE COMPARISON ON THE TEST SET.

Model	Scenario	RMSE (average)	RMSE $(t+0)$	RMSE $(t+10)$
Zero acc.	Road	0.376	0.382	0.385
Const. acc.	Road	0.318	0.286	0.402
FFNN-sidewalk	Road	0.312	0.286	0.334
FFNN-road	Road	0.284	0.276	0.314
Zero acc.	Sidewalk	0.876	0.850	0.962
Const. acc.	Sidewalk	0.365	0.308	0.449
FFNN-road	Sidewalk	0.315	0.296	0.334
FFNN-sidewalk	Sidewalk	0.301	0.295	0.250

F. Training

Due to differences in motion behavior, the dataset was divided into two subsets based on whether the VRU is moving on the road or on the sidewalk. Instead of training a single unified model, two separate FFNN models were trained (FFNN-road and FFNN-sidewalk), each was trained and validated on its corresponding subset.

Both models were trained using the same network architecture and hyperparameter configuration described previously. This way, any performance differences can be attributed to the different motion characteristics of road- and sidewalk-riding VRUs rather than to changes in model design or optimization settings.

The proposed FFNN models contain only 1803 trainable parameters each. On a standard CPU (Intel i7), with a batch size of 1, the average inference time per sample is approximately $0.033 ms$, corresponding to roughly 30,300 predictions per second. This performance exceeds typical real-time requirements and demonstrates the high computational efficiency of the proposed architecture.

To evaluate robustness to random initialization, the model was trained 10 times using different random seeds. The average test RMSE varied only by ± 0.002. This shows that the model performance is stable not significantly affected by random training variations.

IV. MODEL EVALUATION

A. Training Convergence and Optimization Stability

The training and validation losses decreased steadily during optimization and converged after approximately 100 epochs. An initial rapid decrease was followed by gradual stabilization, indicating effective learning of the dominant motion patterns. The validation loss remained slightly lower than the training loss due to dropout being active during training but disabled during evaluation. No oscillatory behavior, divergence, or instability was observed throughout the training process. The smooth convergence behavior confirms appropriate selection of the learning rate and well-conditioned optimization dynamics.

B. Model comparison with kinematic baselines

Table II presents the quantitative comparison between the proposed FFNN models and two baseline approaches: zero-acceleration and constant-acceleration prediction. Performance

is evaluated using the average RMSE over the prediction horizon as well as the RMSE at the first prediction step $(t+0)$ and the final prediction step $(t+10)$.

For the road scenario, the FFNN-road reduces the average RMSE from 0.376 (zero baseline) and 0.318 (constant baseline) to 0.284. Improvements are most evident in the final prediction step, where the FFNN-road achieves lower long-horizon error than the baselines. When the FFNN-sidewalk is evaluated on road-riding VRUs, the average RMSE increases to 0.312, indicating reduced cross-domain generalization compared to the scenario-specific model.

For the sidewalk scenario, the performance gap is even more significant. The zero-acceleration baseline produces considerably higher error (0.876 average RMSE), while the constant-acceleration baseline achieves 0.365. The FFNN-sidewalk further reduces the average RMSE to 0.301 and maintains lower error across both short- and long-horizon predictions. When the FFNN-road is applied to sidewalk-riding VRUs, the average RMSE increases to 0.315.

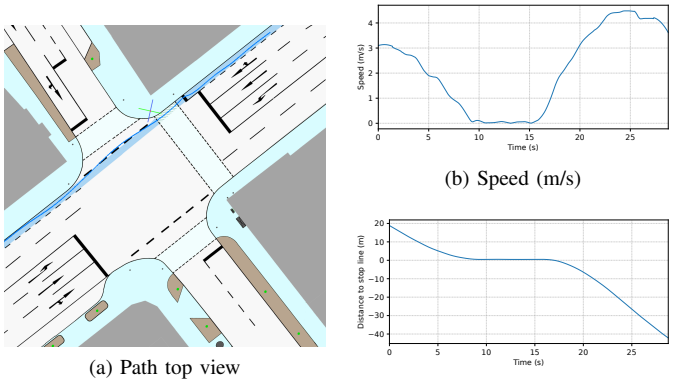
Overall, the results demonstrate that the proposed FFNN model consistently improves predictive accuracy over simple kinematic baselines, particularly for longer prediction horizons.

C. Results

The predictive performance was also evaluated by comparing the ground-truth acceleration with the predicted acceleration over time. The first analyzed example corresponds to a bicycle driving on the road using the bike lane, see the blue path in Fig. 2a. The motion sequence illustrates a typical stop-and-go behavior at the intersection influenced by traffic light changes. The FFNN model used for this evaluation was trained exclusively on road-riding VRUs, ensuring that the prediction reflects scenario-specific motion characteristics.

The traffic signal is initially red and at $t = 13s$, it transitions to red-yellow and subsequently to green at $t = 14s$. The bicycle decelerates smoothly from approximately $v = 3.1m/s$ to standstill between $t \in [0, 9]$ s. This standstill phase is observed between $t \in [9, 15]$ s, corresponding to the red signal. Following the transition to green, the speed increases steadily, reaching approximately $v = 4.3m/s$ before gradually stabilizing, see Fig. 2b. The distance to the stop line decreases continuously during the initial approach and becomes $s = 0m$ during the red phase. Following the green transition, the distance becomes negative, indicating that the bicycle has crossed the stop line and entered the intersection, see Fig. 2c.

The acceleration prediction results are shown in Fig. 3. The predicted acceleration closely follows the ground-truth signal during both deceleration and acceleration phases. During the braking phase, the model captures the negative acceleration trend with minor smoothing. After the light signal turns green, the model correctly predicts the beginning of positive acceleration and the following increase in magnitude. For the first-step prediction, the maximum absolute deviation between



(a) Path top view

(b) Speed (m/s)

(c) Distance to stop line (m)

Fig. 2. Scenario of VRU riding on the road

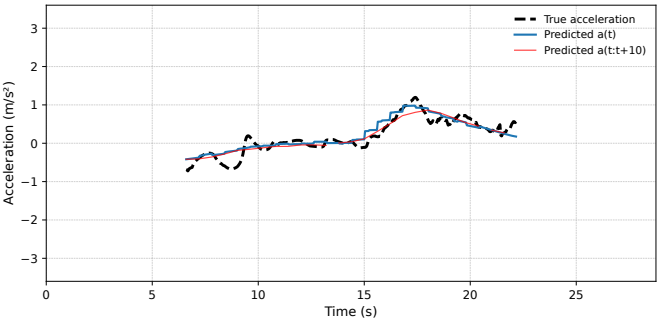


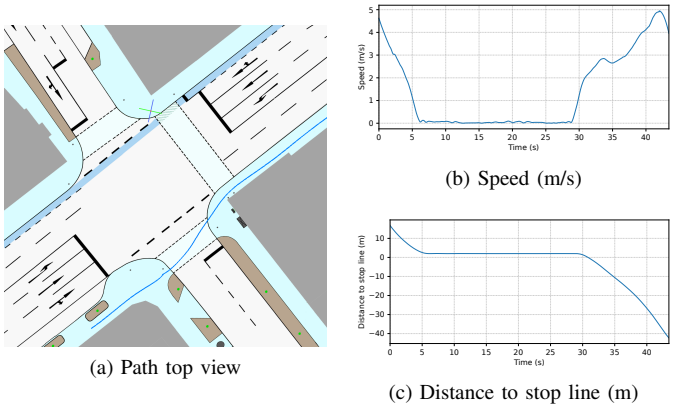
Fig. 3. Predicted and true acceleration for VRU riding on the road

predicted and true acceleration is $a = 0.46m/s^2$. The multi-step prediction trajectories also remain stable and do not exhibit oscillatory or divergent behavior. At $t = 14s$, the multi-step predictions correctly predict the initiation of acceleration at $t = 15s$ and accurately reproduce the subsequent fluctuation patterns. A slight underestimation of peak acceleration values is observed. This behavior is typical of regression-based models and does not significantly affect the overall quality of trajectory prediction.

The second analyzed example corresponds to a bicycle approaching a pedestrian crossing on the sidewalk, see the blue path in Fig. 4a. The FFNN model used for this evaluation was trained exclusively on sidewalk-riding VRUs.

The traffic light signal is initially red and changes to green at $t = 26s$. The bicycle decelerates smoothly from approximately $v = 4.5m/s$ to standstill between $t \in [0, 6]$ s. The standstill phase is observed between $t \in [6, 28]$ s, corresponding to the red signal. Following the transition to green, the speed increases with minor fluctuations, reaching approximately $v = 5.0m/s$, see Fig. 4b. The distance to the stop line decreases continuously during the initial approach and becomes approximately $s = 2m$ during the red phase. This is the usual behavior of VRUs driving on the sidewalk to maintain a safety space from the road. Following the green transition, the distance becomes negative, indicating that the bicycle has crossed the sidewalk edge and is riding on the pedestrian crossing.

The acceleration prediction results are shown in Fig. 5.



(a) Path top view

(b) Speed (m/s)

(c) Distance to stop line (m)

Fig. 4. Scenario of VRU riding on sidewalk

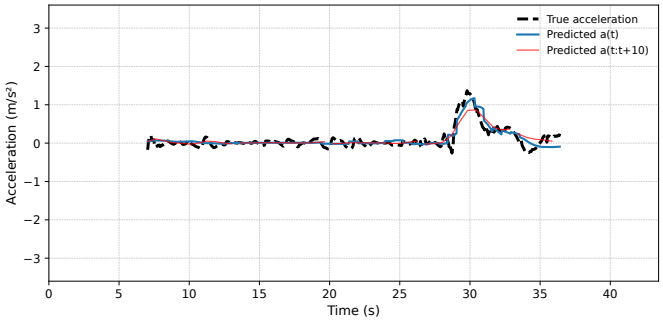


Fig. 5. Predicted and true acceleration for VRU riding on sidewalk

The predicted acceleration closely follows the ground-truth signal. The model correctly predicts the beginning of positive acceleration after the light turns green. For the first-step prediction, the maximum absolute deviation between predicted and true acceleration is $a = 0.32m/s^2$. The multi-step prediction trajectories remain stable and do not exhibit oscillatory or divergent behavior. At $t = 26.5s$, the multi-step predictions correctly predict the initiation of acceleration at $t = 28s$ and accurately reproduce the subsequent fluctuation patterns. A slight underestimation of peak acceleration values is also observed in this scenario. Overall, the prediction quality remains consistent despite the increased velocity fluctuations in the sidewalk scenario.

V. Conclusion

This paper presents a feed-forward neural network approach for multi-step acceleration prediction for vulnerable road users using the IMPTC dataset. A frequency-domain analysis was performed on the dataset to identify the dominant motion time scale, which guided the selection of the history window and prediction horizon. The proposed lightweight architecture achieves improved predictive performance compared to zero- and constant-acceleration baselines while maintaining computational efficiency suitable for real-time operation. The structured hyperparameter selection strategy and the local sensitivity analysis of the model show that the definition of the temporal structure has a greater impact on predictive performance than moderate changes in optimization settings. The results demonstrate that a compact feed-forward architecture provides a robust and efficient solution for short-term VRU motion prediction in urban traffic environments.

Future work will focus on extending the input features with additional environmental information, e.g., the behavior of surrounding traffic participants and weather conditions (precipitation, visibility). This may further improve prediction robustness in complex urban scenarios. In addition, the limited availability of high-quality micromobility datasets, particularly for e-scooter motion, motivates the collection of dedicated measurement data. The development of a dedicated dataset with sensor-equipped e-scooters and synchronized environmental sensing is planned. This will enable a more accurate modeling of e-scooter motion behavior.

References

[1] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 33–47, 2020.

[2] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH journal*, vol. 1, no. 1, p. 1, 2014.

[3] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2891–2900.

[4] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[5] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[6] M. Golchoubian, M. Ghafurian, K. Dautenhahn, and N. L. Azad, "Pedestrian trajectory prediction in pedestrian-vehicle mixed environments: A systematic review," *IEEE transactions on intelligent transportation systems*, vol. 24, no. 11, pp. 11 544–11 567, 2023.

[7] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, "Learning lane graph representations for motion forecasting," in *European Conference on Computer Vision*. Springer, 2020, pp. 541–556.

[8] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 1179–1184.

[9] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 3, pp. 900–918, 2019.

[10] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Conference on robot learning*. PMLR, 2018, pp. 947–956.

[11] A. Laudani, G. M. Lozito, F. Riganti Fulginei, and A. Salvini, "On training efficiency and computational costs of a feed forward neural network: A review," *Computational intelligence and neuroscience*, vol. 2015, no. 1, p. 818243, 2015.

[12] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," *arXiv preprint arXiv:1910.05449*, 2019.

[13] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[14] M. Hetzel, H. Reichert, G. Reitberger, E. Fuchs, K. Doll, and B. Sick, "The imptc dataset: an infrastructural multi-person trajectory and context dataset," in *2023 IEEE Intelligent vehicles symposium (IV)*. IEEE, 2023, pp. 1–7.