Multimodal Sensing for Predicting Real-time Biking Behavior based on Contextual Information

Prasenjit Karmakar, Ajay Kumar Meena, Kushal Natani, Sandip Chakraborty
Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India.
{prasenjitkarmakar52282, ajaymeena3210, kushalnatani123}@gmail.com, sandipc@cse.iitkgp.ac.in

***Abstract*—Overspeeding is a significant cause of road accidents, especially when the target vehicle is a two-wheeler. Coupled with infrastructural limitations and the general reckless driving behavior, it becomes challenging to reduce the problem of overspending, mainly because the optimal speed depends not only on road types but also on several spatiotemporal contexts. To mitigate this, in this paper, we propose *Pathik*, which uses multimodal contextual information to accurately predict the speeding behavior of a bike driver for the next road segment. *Pathik* then aggregates this information with the demographic and map-based information for the next road segment and recommends decelerating if the bike speed exceeds. Principled evaluation on an in-house dataset with different bike types (both geared and gearless) shows that *Pathik* can accurately predict the speed for the next patch with a mean R_2-score of 0.92 (±0.015).**

***Index Terms*—Biking, Contextual sensing, Multimodal.**

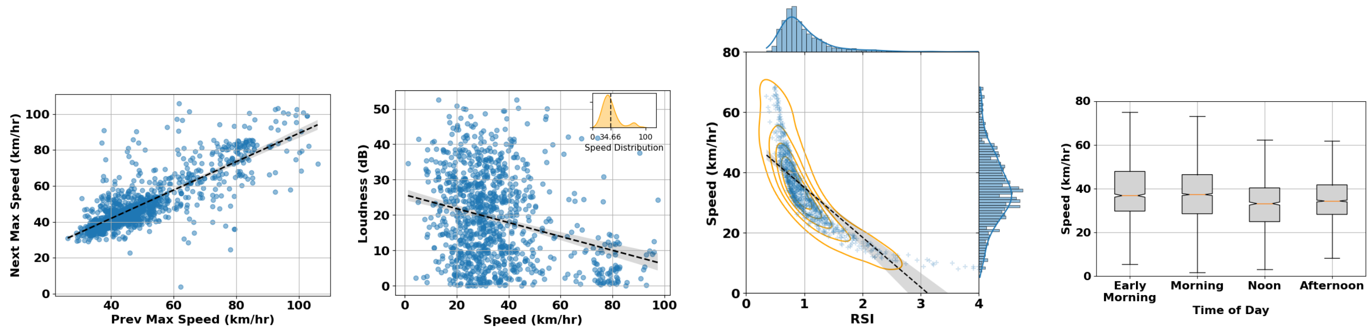
I. Introduction

In recent years, several countries in the world, particularly the low and middle-economy ones, have witnessed a tremendous increase in the number of two-wheelers on the road. This can be attributed to the growing population, urbanization, and increased disposable income. However, this has also led to an increase in the number of accidents, especially those involving two-wheeler bikes. According to the Ministry of Road Transport and Highways (MoRTH) data for 2021, India witnessed a staggering number of road accidents, with a total of 412432 incidents recorded. These accidents resulted in the loss of 153972 lives, while 384448 individuals were injured. Notably, two-wheelers have claimed the highest number of lives in road accidents, 69544, which is around 45.2% of the total road accident deaths [1]. The government has implemented several measures to address this issue, such as imposing fines for traffic violations, improving road infrastructure, and promoting road safety awareness campaigns; however, more than these measures are needed to curb the rising number of accidents by monitoring and regulating the biking behavior.

Notably, the expected bike speed in a region depends not only on the road type but also on several other spatiotemporal contextual information as well as the personal traits of the biker. For example, the demography of the region, such as schools or hospitals nearby, or the level of congestion or crowdedness significantly impacts the biking behavior and hence the bike speed. Several studies [2], [3], [4], [5] have focused on GPS trajectories to analyze inter-city traffic behavior and vehicle speeds over road segments. Further, GPS along with population demographic data [5], [6], [7] have been used to understand how neighborhood characteristics, such as income level and race, affect vehicle speeds. Researchers [8] have also explored modalities like video, honking sound, GSM radio signals, meteorological data, and social media feeds to complement GPS. The recent literature [9], [10] has also developed traffic risk detection algorithms by combining GPS data with camera images to recognize hazardous driving scenarios automatically. However, these works have analyzed traffic behavior offline for four-wheelers and thus do not apply to personalized real-time biking scenarios. The challenge here is to predict the biking speed in real time while considering the driving behavior of the biker influenced by the spatiotemporal environmental context.

To personalize vehicle speed estimates, the authors in [11] have attempted to model traffic conditions using smartphone sensor data and propose the permissible speed limit for road segments. In contrast, [12] effectively used the smartphone's microphone data to identify traffic conditions. Furthermore, authors in [13], [14] have shown that the state of the road surface is also a significant factor affecting how fast vehicles travel. The smartphone's IMU sensor [15], [16] is used to estimate road quality by detecting potholes, roughness, etc. However, these studies only involve public transport services like taxis, buses, personal four-wheeler vehicles, etc., not exploring the challenges due to ad-hoc and more chaotic nature of the substantial two-wheeler traffic. Consequently, we need to develop a robust context-sensitive model for real-time biking behavior prediction.

In this paper, we propose *Pathik*, an assistive framework for two-wheelers, which utilizes sensor data from COTS smartphone and spatial housing demographics, points of interest (i.e., schools, hospitals), etc., from the map API to predict the driving speed in the upcoming road segment and detect reckless driving behavior to warn the driver. At its core, our method uses a hybrid neural network that processes the previous mobility patterns influenced by the biking behavior along with the acoustic signals from the environment and observes the spatial demographics of the approaching area to estimate the biking behavior for the upcoming road segment. To assess *Pathik*, we have collected an in-house dataset of 1448 km of biking (both geared and gearless) trails for three different routes. The experimental results demonstrate that *Pathik* can predict the biking speed for the next road segment with a R_2-score of 0.92 (±0.015) and mean absolute error of 3.14 (±0.41) kmph.



(a) Previous vs Next speed (b) Speed vs Loudness (c) Speed vs RSI (d) Speed vs Time of Day

Fig. 1: Correlation of biking speed with different contextual information.

II. Pilot Study

Two-wheeler trails are recorded using smartphones over four weeks, considering three routes (approx. 9 km each) at Durgapur city in India (Details in Sec. IV-B). Our analysis of the initial week's data reveals correlations between the biking speed in the next road segment and several factors such as historical driving behavior, traffic noise levels, road surface quality, etc. *We observe that reckless driving behavior tends to persist over time rather than being a one-time occurrence*. Fig. 1a demonstrate a positive correlation between the speed of the previous and next 100 m road segment. *The traffic noise level also influences driving behavior.* Fig. 1b shows a negative correlation between traffic loudness and two-wheeler speed; thus, people tend to drive slowly in noisy traffic conditions. Needless to say, *road quality impacts the speed of the two-wheeler*. We have computed RSI (Road Surface Index [14], lower value represents better road quality) for each road segment and observe that speed negatively correlates with RSI as shown in Fig. 1c.

Moreover, temporal factors such as weekends *time of day also impact driving patterns due to city-specific traffic patterns*. Fig. 1d depicts that the median speed of the two-wheeler is slightly higher during the busy hours in the early morning and evening, compared to the late morning and noon, respectively. Most importantly, *spatial demographics of the road segment plays a vital role in driving behavior.* For example, people drive bikes slowly in residential localities where educational institutions, hospitals, and parks are located compared to highways. Underscoring the above observations, we discuss the design of *Pathik* framework to exploit these correlations and accurately estimate the speed of the immediate road segment in the following section.

III. Methodology

Building upon the findings of Section II, we compute the following features from the smartphone sensors and Google Maps API for each 100 m road segment.

- **Mobility:** We utilize GPS data to compute the *Previous Speed* of the two-wheeler in before entering new road segment. *Road Surface Index* is computed from the smartphone's IMU.

TABLE I: Availability of different point of interests.

	Park	Medical		Education	Temple	Malls
		Morning	Evening			
Opened	00:00	10:00	13:00	07:00	07:00	09:00
Closed	23:00	15:00	19:00	17:00	20:00	21:30

TABLE II: Hyper-parameters of the model.

Parameter	Value	Parameter	Value
Optimizer (∇)	Adam	Learning rate (α)	0.001
Activation (σ)	ReLU	Epoch (E)	100
Loss ($\mathcal{L}$)	MSE	Batch size (B)	64

- **Traffic Noise:** The smartphone's microphone records the traffic sounds, which is then divided according to the road segments to compute average *Loudness* and *MFCC* (Mel-Frequency Cepstral Coefficients) values that extract the complete auditory spectrum of traffic noises for a road segment.
- **Temporal:** In our pilot experiments, we observe that two-wheeler speed patterns vary based on the *Time of Day*, along with which *Day of the Week* the trail is collected. For instance, weekends and holidays imply lesser driving speed. Moreover, different *Points of Interest* open at different times of the day (see TABLE I), impacting its surrounding traffic behavior.
- **Spatial Demographics:** We query the demographic image for the upcoming road segment from the static Maps API. The image represents the structure and placement of different objects within the approaching area, such as the percentage of *Housing*, *Roads*, presence of *School*, *Hospital* or *Amusement* regions, etc. The raw image is also fed as a feature to preserve the structural information of the buildings and road network. Moreover, the number of *Unique WiFi* devices is used to estimate the population density of the locality.

We took the feature sequence of the past 1 km trail. Note the features are computed at each 100 m, resulting in a sequence length of 10, along with a demographic image (64px$\times$64px representing 100m$\times$100m area) of the upcoming road segment. Due to the multi-modal nature of the data, we have used two separate encoding heads to process and compute embeddings. The feature sequence is fed to a temporal attention layer to efficiently compress and capture

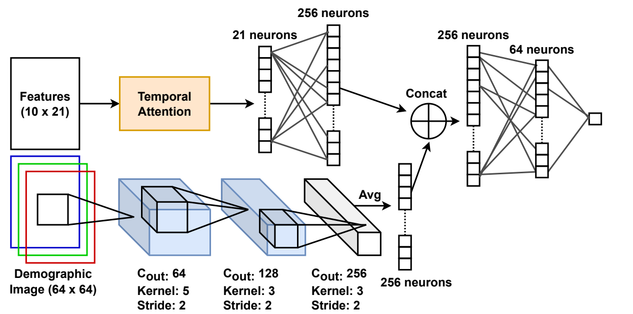


Fig. 2: The neural network architecture of *Pathik*. Output channels of a convolution layer is represented as C_{out}.

TABLE III: The data collection routes in Durgapur.

Route	Length (km)	Trails		Total (km)
		Bike	Scooty	
R1	9.1	63	12	682.72
R2	9.11	17	0	154.89
R3	8.26	74	0	611.61

the dependencies and further provided to a fully connected layer having 256 neurons. At the same time, the demographic image is processed through a convolutional neural network to compress to 256-dimensional embedding. Finally, both the embeddings are concatenated and fed to an Artificial Neural Network regressor to estimate the speed value in the next 100m road segment. The regressor has two fully connected layers of 256 and 64 neurons, respectively, before the final output neuron. The complete neural network architecture is depicted in Fig. 2. All the layers have ReLU (Rectified Linear Unit) activation for non-linearity in the model. The hyper-parameters are listed in Table II.

IV. Evaluation

In this section, we evaluate the *Pathik* framework on the collected two-wheeler dataset in various scenarios as follows:

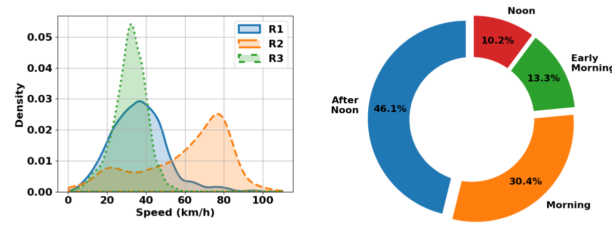
A. Implementation Detals

Hardware setup: The deep learning model is trained on an iMac M1 (with 16GB primary memory running MacOS v12.6 with base-kernel version: 21.6.0). We utilized a Workstation (48 × vCPU, GPU Nvidia TitanX 12GB, & RAM 128GB) for hyper-parameter tuning. We have used the software package based on `Python3.10.6`, `Tensorflow v2.12.0`, and `Scikit-learn v1.2.1` for the implementation. The source code and the collected two-wheeler dataset are publicly available at https://github.com/prasenjit52282/Pathik.

Evaluation Metric: To evaluate the model, we have considered three metrics: MAE (Mean Absolute Error), MSE (Mean Square Error), and R_2-score to be the performance measures as it is considered the standard for regression model evaluation.

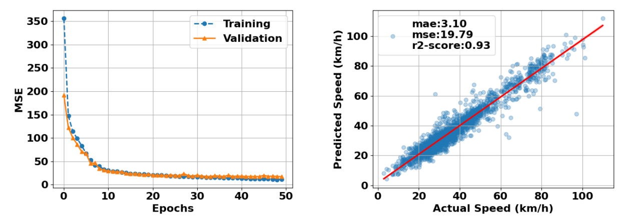
B. Data Collection

We have collected the two-wheeler trails at Durgapur city in India for three routes: R1, R2, and R3. Fig. 3a shows the speed distribution in each route. We observe that the average speed



(a) Speed distribution (b) Data in times of the day

Fig. 3: Collected two-wheeler dataset in Durgapur.



(a) MSE during training (b) R_2-score during testing

Fig. 4: Performance of the model on 80-20 train test split.

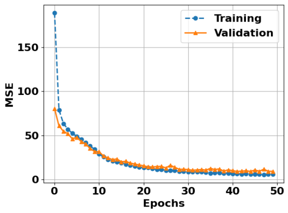
of R2 is higher as it goes through the highway. In total, six participants voluntarily took part in the data collection process, recording a total of 1448 km trail from both bike and scooty as shown in TABLE III for four weeks. Fig. 3b shows the data quantity for different times of day. The smartphone's sensor data is recorded with an open-source Android application available at https://github.com/prasenjit52282/LoggerApp.

C. Overall Performance

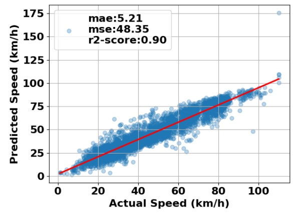
To evaluate the performance of our model, we divide the total data into an 80% training set and a 20% testing set. In Fig. 4a, we observe that the MSE drops exponentially with the number of training epochs, and the drop is similar for both the train and validation sets (20% of the training set). This indicates that the model is not overfitting to the training data and can generalize well to unseen data. Further, we calculate the R_2-score for the testing set and get approximately 0.93 as shown in Fig. 4b, which suggests that the model is highly accurate in predicting the two-wheeler speed for the upcoming road segments. The MAE on the testing set is 3.10 kmph. Therefore, the model is slightly deviated from the actual speed of the two-wheeler.

D. Year-wise Performance

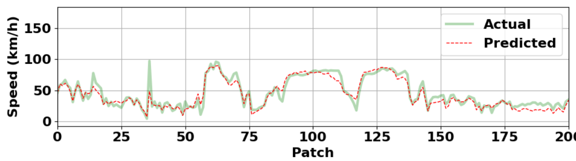
We evaluate the model for biases by selecting the data collected in 2019 as the training data, while the data collected in 2018 is used as a testing set. Fig. 5a shows the MSE drops for the training and validation set with training epochs. Moreover, we observe the R_2-score to be 0.90 as shown in Fig. 5b, suggesting the adaptability of the model across years. The framework has an admissible MAE of 5.21 kmph on the



(a) MSE during training



(b) R_2-score during testing



(c) Variations between actual and predicted speed

Fig. 5: Performance of the model trained on 2019 data and tested on 2018 data.

TABLE IV: Cross-validation performance for 10-folds.

Split no	Train MAE	Train MSE	Train R_2-score	Test MAE	Test MSE	Test R_2-score
0	2.46	11.44	0.96	3.19	22.54	0.92
1	2.41	11.29	0.96	2.79	13.31	0.94
2	2.39	11.03	0.96	3.19	19.29	0.93
3	2.45	11.51	0.95	3.88	42.75	0.90
4	2.72	15.03	0.94	2.78	14.12	0.92
5	2.44	11.39	0.96	3.33	20.16	0.92
6	2.52	12.12	0.95	3.74	27.04	0.89
7	2.65	13.38	0.95	2.77	13.79	0.94
8	2.36	10.58	0.96	3.01	17.61	0.93
9	2.58	12.77	0.95	2.71	13.86	0.94

testing set. Compared to overall testing, the slight decrease in the performance in across-year testing is attributed to the change in demographic data within the year. Moreover, other factors could change over the year and impact the speed, such as the weather conditions, which are not considered in our model. The model is also tested against several consecutive road patches or segments in Fig. 5c. We observe that the framework closely predicts the next segment's speed to the collected ground truth speed.

E. 10-fold Cross-Validation

We further evaluate our model using the 10-fold cross-validation method, dividing our data into ten splits. We select one such split in each experiment as the testing set while the model is trained over the remaining nine splits. Table IV shows the performance of the model, where the i^{th} row corresponds to i^{th} split being used as the testing set, and the rest are used as the training set. The model shows an average $0.92(\pm0.015)$ R_2-score and 3.14 (±0.41) kmph MAE for the testing set. In summary, we consistently observe a commendable R2-score, indicating a strong correlation between the predicted and actual two-wheeler speed, thus ensuring the model's reliability across different two-wheeler trails.

V. Conclusion

This paper aims to develop the *Pathik* framework that estimates the future speed of two-wheelers and identifies instances of reckless over-speeding utilizing spatial demographics and temporal sensor data as features. We have collected an extensive dataset from smartphone sensors covering more than 1448 km of trails to extract the training features and validate the learning model. To the best of our knowledge, this is the first system to focus on two-wheeler speed prediction using contextual information from a smartphone. However, further work is required before the system becomes deployable in the real world. Although the proposed deep learning model has demonstrated an excellent R_2-score of 0.92 in predicting future speed based on processed data, we must evaluate its adaptability and usability in different real-world scenarios, such as various routes, cities, and bike types. As a future direction, we plan to develop prototypes for two-wheelers of different make to test the framework rigorously.

References

[1] Ministry of Road Transport and Highways (MoRTH), "Road accidents in India," Online report, 2021. [Online]. Available: https://morth.nic.in/sites/default/files/RA_2021_Compressed.pdf

[2] S. Aoki, K. Sezaki, N. J. Yuan, and X. Xie, "An early event detection technique with bus GPS data," in *25th ACM SIGSPATIAL GIS*, 2017.

[3] X. Fei and O. Gkountouna, "Spatiotemporal clustering in urban transportation: a bus route case study in washington DC," *SIGSPATIAL Special*, vol. 10, no. 2, pp. 26–33, 2018.

[4] X. Fei, O. Gkountouna, D. Pfoser, and A. Züfle, "Spatiotemporal bus route profiling using odometer data," in *27th ACM SIGSPATIAL GIS*, 2019, pp. 369–378.

[5] L. Jiang, X. Chen, and W. He, "SafeCam: Analyzing intersection-related driver behaviors using multi-sensor smartphones," in *IEEE PerCom*, 2016, pp. 1–9.

[6] J. Yu *et al.*, "Senspeed: Sensing driving conditions to estimate vehicle speed in urban environments," *IEEE Transactions on Mobile Computing*, vol. 15, no. 1, pp. 202–216, 2015.

[7] L. Yan and H. Shen, "TOP: optimizing vehicle driving speed with vehicle trajectories for travel time minimization and road congestion avoidance," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 2, pp. 1–25, 2019.

[8] M. X. Hoang, Y. Zheng, and A. K. Singh, "FCCF: Forecasting citywide crowd flows based on big data," in *24th ACM SIGSPATIAL*, 2016.

[9] Z. Zhou, Y. Wang, X. Xie, L. Chen, and H. Liu, "RiskOracle: A minute-level citywide traffic accident forecasting framework," in *AAAI*, vol. 34, no. 01, 2020, pp. 1258–1265.

[10] Y. Yin *et al.*, "Multimodal deep learning for robust road attribute detection," *ACM Transactions on Spatial Algorithms and Systems*, vol. 9, no. 4, pp. 1–25, 2023.

[11] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and mobile application for driver behavior analysis and accident prevention," *IEEE transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2427–2436, 2019.

[12] D. Vij and N. Aggarwal, "Smartphone based traffic state detection using acoustic analysis and crowdsourcing," *Applied Acoustics*, vol. 138, pp. 80–91, 2018.

[13] S. Tiwari, R. Bhandari, and B. Raman, "Roadcare: a deep-learning based approach to quantifying road surface quality," in *ACM COMPASS*, 2020, pp. 231–242.

[14] R. Mandal *et al.*, "Exploiting multi-modal contextual sensing for city-bus's stay location characterization: Towards sub-60 seconds accurate arrival time prediction," *ACM Trans. Internet Things*, vol. 4, 2023.

[15] W. Zhong *et al.*, "A reliability-aware vehicular crowdsensing system for pothole profiling," *ACM IMWUT*, vol. 3, no. 4, pp. 1–26, 2019.

[16] K. R. Opara *et al.*, "Road roughness estimation through smartphone-measured acceleration," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 2, pp. 209–220, 2021.