Monitoring Bicycle Safety through GPS data and Deep Learning Anomaly Detection

Shumayla Yaqoob*, Salvatore Cafiso#, Giacomo Morabito*, Giuseppina Pappalardo#

*Department of Electrical, Electronic, Computer and Telecommunication Engineering, University of Catania Italy
email: shumayla.yaqoob@phd.unict.it, giacomo.morabito@unict.it

#Department of Civil Engineering and Architecture, University of Catania Italy
email: dcafiso@unict.it, giuseppina.pappalardo1@unict.it

Keywords: cyclist safety, GPS trajectory data, Convolutional Neural Network, Surrogate measure.

1 INTRODUCTION

Cycling has always been considered a sustainable and healthy mode of transport. Moreover, during Covid-19 period, cycling was further appreciated by citizens as an individual opportunity of mobility. As a counterpart of the growth in the number of bicyclists and of riding kilometres, bicyclist safety has become a challenge as the unique road transport mode with an increasing trend of crash fatalities in EU (Figure 1) [1] [2].

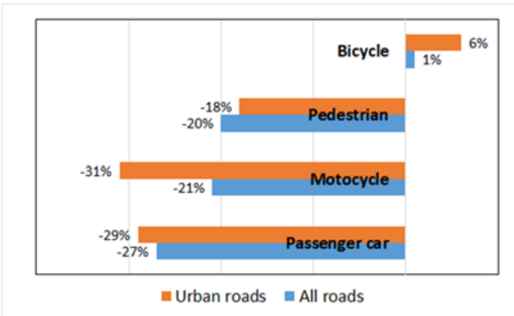

Figure 1: Trends 2010-2018 of Fatalities in crashes involving cyclists and other transport modes. Source: [2]

When compared to the traditional road safety network screening, availability of suitable data for crashes involving bicyclists is more difficult because of underreporting and traffic flow issues. In such framework, new technologies and digital transformation in smart cities and communities is offering new opportunities of data availability which requires also different approaches for collection and analysis.

An experimental test was carried out to collect data from different users with an instrumented bicycle equipped with Global Navigation Satellite Systems (GNSS) and cameras. A panel of experts was asked to review the collected data to identify and score the severity of the safety critical events (CSE) reaching a good consensus.

Anyway, manual observation and classification of CSE is a time consuming and unpractical approach when large amount of data must be analysed. Moreover, due to the complex correlation between pre-crash driving behaviour and due to high dimensionality of the data, traditional statistical methods might not be appropriate in this context. Deep learning-based model have recently gained significant attention in the literature for time series data analysis and for anomaly detection, but generally applied to vehicles' mobility and not to micro-mobility.

We present and discuss data requirements and treatment to get suitable information from the GNSS devices, the development of an experimental framework where convolutional neural networks (CNN) is applied to integrate multiple GPS data streams of bicycle kinematics to detect the occurrence of a CSE.

2 RELATED WORK AND PROBLEM FORMUATION

Deep learning-based algorithms, especially Convolutional Neural Networks (CNN) are widely used for anomaly detection in time series data [3]. Recently, deep learning and Convolutional Neural Networks (CNN) have been applied in road safety studies [4] and driving style analysis [5]. The convolutional autoencoders (CAE) allowed the extraction of valuable information from large quantities of complex and heterogeneous data, showed fast convergence speed due to the convolutional layers, and provides better performance to achieve volume, variability and velocity (i.e. big Data) [6]. The limitations of the existing works include that the observational studies applied on bicyclist safety mainly relay on traffic conflict techniques for video tracking from fixed positions. Few studies used trajectory data to identify SCE. To the best of our knowledge, this is the first work extending the use of deep learning CNN to extract features of the riding style of bicyclists from GNSS data and to detect anomaly events relevant for road safety assessment.

3 METHODOLOGY

3.1 Dataset Preparation

Dataset preparation includes both collection and treatment of GNSS data to extract features suitable to train and test the CNN model. For data collection, the source is an instrumented bicycle with GPS (Global Positioning System) and HD video system (Video Vbox Lite). Once data was recorded, different Python routines were applied to 1) improve the data quality, 2) interpolate for smoothing [7], 3) calculate derived parameters and 4) create the data set for training and testing the CNN. In the present application, speed (S) and heading (H) define the recorded time series in the GPS data, while longitudinal acceleration (LA), travelled Distance (D) and heading rate (HR), transversal acceleration (TA) and combined acceleration (CA) are derived.

3.2 BeST-DAD: the Proposed CNN application for Anomaly Detection

We call the complete scheme proposed for anomaly detection in the scenario of interest: '**B**icycl**e** **S**afety **t**hrough **D**eep learning-based **A**nomaly **D**etection' (BeSt-DAD). Best-DAD employs a 1-D Convolutional Autoencoder (CAe) as depicted in Figure 2. The input consists of a sequence of data samples, $X_1, X_2, \ldots$, generated at a frequency of 10 Hz and filtered as discussed in the previous Section 3.1. The generic X_i is a 6-tuple of values, i.e., $X_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}]$ which represent the speed, heading, heading-rate, longitudinal acceleration, transversal acceleration and combined acceleration, respectively, and thus, the input data is 2-dimensional in nature, as shown in Figure 2. Nevertheless, we flatten the input data and consider it as a 1-dimensional sequence of type:

$$x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26} \ldots \quad (1)$$

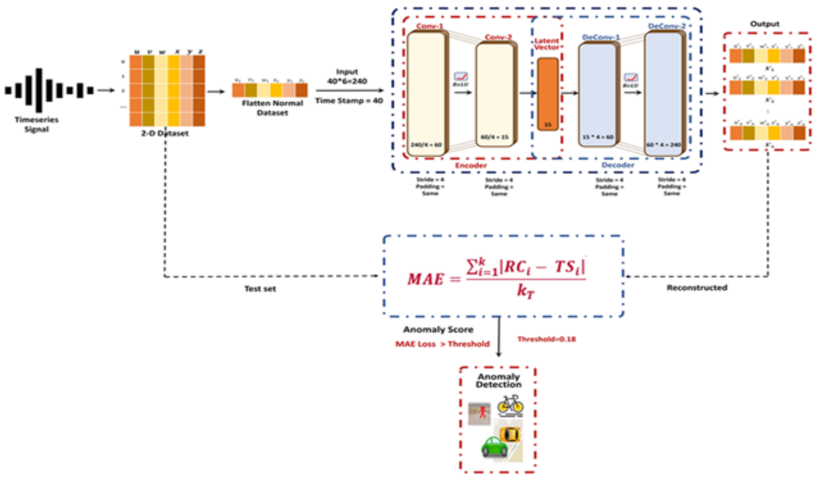

Figure 2: BeST-DAD scheme

4 RESULT AND DISCUSSION

For results, two scenarios have been selected for training and testing of the model: *A1)* Training and Testing CNN for each user, by using 80% of data for training and 20% of data for testing. *A2)* Training with the data collected by considering the entire dataset related to only one user and testing with the data from all the other users. Results for the validation scenarios (D, E, F), in terms of Recall, Precision and F-score, are presented and compared with the reference results for the two training approaches (A1, A2), for different thresholds (T) and time window sizes (TW). Results confirm the best performance for selected values of T and TW. Moreover, the application of the Svitzky Golay Filter (SGF) to the high-frequency time-variability of the cycling data improved mainly the Precision in the classification with a notable reduction of FPs. Finally, it is worth noting that merging speed and heading parameters resulted the main factor to improved performance.

5 CONCLUSION

Cyclists are vulnerable road users and their safety is still a challenge and a serious issue to be addressed. The effectiveness of the Artificial Intelligence technology and positive validation of the results with real data, makes the approach promising for the identification of location with potential hazard for cyclists in the wide urban road network by using mobility data that can be easily collected in smart cities and communities.

Performance evaluation of BeST-DAD for different model settings demonstrates that adding direction information (heading, heading rate, transversal acceleration) to the only speed parameters (speed, longitudinal acceleration), improved remarkably the capability of the model to detect anomalies. Data filtering by using SGF played an important role in reducing the FPs, as well.

The most relevant limitation highlighted by the study is the high number of FPs anyhow produced by the classification technique. The low precision is mainly related the actual phenomenon which is characterized by the cycling natural waving and speed variability and the occurrence of other factors other than CSE requiring sudden changes in riding (e.g. traffic signal, potholes, etc.)..

Finally, it is worthy to remark that CSE identification and spatial clustering may be used as first level network wide road safety assessment to be followed by targeted inspection to identify the actual need of remedial actions and priority as depicted by the most recent EU directive for Road Infrastructure Safety Management [8].

REFERENCES

[1] R. Feleke, S. Scholes, M. Wardlaw, and J. S. Mindell, “Comparative fatality risk for different travel modes by age, sex, and deprivation,” *J. Transp. Heal.*, vol. 8, no. April 2017, pp. 307–320, 2018, doi: 10.1016/j.jth.2017.08.007.

[2] “EU Commision Road Safety – Key figures,” *EU Commision*, 2020. https://ec.europa.eu/transport/road_safety/index_en.

[3] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, “Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series,” *IEEE Trans. Syst. Man, Cybern. Syst.*, pp. 1–11, 2020, doi: 10.1109/tsmc.2020.2968516.

[4] J. Zhao, P. Liu, C. Xu, and J. Bao, “Understand the impact of traffic states on crash risk in the vicinities of Type A weaving segments: A deep learning approach,” *Accid. Anal. Prev.*, vol. 159, no. June, p. 106293, 2021, doi: 10.1016/j.aap.2021.106293.

[5] A. Bichicchi, R. Belaroussi, A. Simone, V. Vignali, C. Lantieri, and X. Li, “Analysis of road-user interaction by extraction of driver behavior features using deep learning,” *IEEE Access*, vol. 8, pp. 19638–19645, 2020, doi: 10.1109/ACCESS.2020.2965940.

[6] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” *Pattern Recognit.*, vol. 61, pp. 650–662, 2017, doi: 10.1016/j.patcog.2016.06.008.

[7] P. Li, M. Abdel-Aty, and J. Yuan, “Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on GPS trajectory data,” *Accid. Anal. Prev.*, vol. 150, no. November 2020, p. 105924, 2021, doi: 10.1016/j.aap.2020.105924.

[8] Directive (EU) 2019/1936 of the European Parliament and of the Council, of 23 October 2019 amending Directive 2008/96/EC on road infrastructure safety management.