Long-Short Term Memory Networks and Synthetic Data for Heavy Vehicle Rollover Prevention

Original
Long-Short Term Memory Networks and Synthetic Data for Heavy Vehicle Rollover Prevention / Perboli, G., Tota, A., Velardocchia, F.. - In: IEEE OPEN JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS. - ISSN 2687-7813. - ELETTRONICO. - 6:(2025), pp. 792-798. [10.1109/ojits.2025.3579653]

Availability:
This version is available at: 11583/3001555 since: 2025-07-04T19:31:34Z

Publisher:
Institute of Electrical and Electronics Engineers Inc.

Published
DOI:10.1109/ojits.2025.3579653

IEEE Open Journal of
Intelligent Transportation Systems

Received 18 September 2024; revised 24 December 2024; accepted 31 May 2025. Date of publication 13 June 2025; date of current version 30 June 2025.
Digital Object Identifier 10.1109/OJITS.2025.3579653

Long-Short Term Memory Networks and Synthetic Data for Heavy Vehicle Rollover Prevention

GUIDO PERBOLI [1], ANTONIO TOTA [2], AND FILIPPO VELARDOCCHIA [1]

[1]DIGEP, Politecnico di Torino, 10129 Turin, Italy
[2]DIMEAS, Politecnico di Torino, 10129 Turin, Italy

CORRESPONDING AUTHOR: F. VELARDOCCHIA (e-mail: filippo.velardocchia@polito.it)

ABSTRACT Heavy vehicle rollover plays a pivotal role in road safety scenarios. Numerous researchers addressed the topic, with particular focus on drivers related injuries. Considering the same and other connected implications, the necessity for techniques able to estimate and predict overturning eventualities appears evident. Different methodologies were explored, with notable achievements obtained by neural network-based algorithms. At the same time, their heavy requirements in terms of data needs to be addressed to allow practical applications in terms of time and costs. Consequently, exploring the interaction between simulation and experimental data becomes extremely important, motivating the methodology proposed by this paper. In details, an heavy vehicle model was designed in IPG Carmaker®, while experimental data on its physical alter ego were acquired. This led to the generation of a synthetic dataset and the collection of an empirical one. Both were used to define a Long Short-Term Memory architecture, with a dual purpose. First, as typical rollover indicator, estimate the vehicle roll angle. Second, compare the performance of the neural networks, aiming to obtain at least the same order of magnitude in terms of RMSE, MSE and MAE. The goal was to demonstrate that synthetic data can not only be used in combination with real data, but also as substitutes able to address time and cost constraints inevitably linked to the latter, allowing more efficient experiments for overtipping prevention.

INDEX TERMS LSTM networks, transportation, noise, rollover, heavy vehicles, signal processing.

I. INTRODUCTION

HEAVY vehicle rollover is a central issue in road safety due to the large use of this type of vehicles in transporting goods and people [1]. Overturning events can lead to severe consequences [2], [3], [4], [5], [6], [7], motivating research on prevention techniques [8], [9], [10], [11]. Various methodologies were proposed, including those focusing on driver safety [12], [13], [14], [15], [16], [17], [18], [19], [20]. More recently, neural networks have gained attention for predicting rollover conditions, given their capability to learn from large datasets [21], [22], [23], [24], [25]. However, these approaches often require extensive and costly on-field data. Experimental campaigns are time-consuming and expensive, making the collection of a sufficient number of real data challenging. This justifies the exploration of synthetic data to complement or replace the empirical ones [26], [27], [28], [29].

The review of this article was arranged by Associate Editor Meng Li.

In this context, the synergy between simulation and experimental data is critical. Synthetic data can be generated in complex and realistic scenarios by high-fidelity simulation tools [30], potentially approximating the behavior of a real vehicle under various maneuvers and terrains. Noise, however, inevitably affects real data. This noise, stemming from road irregularities and sensor imperfections, is often unknown and can reduce the performance of neural networks [31], [32]. While some studies resort to Kalman filters and CNNs for noise mitigation [33], [34], the role of Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) architectures, to handle time-dependent noisy signals remains less explored in heavy vehicle rollover prevention. LSTMs can effectively model temporal sequences and have demonstrated remarkable performance in similar tasks [35], [36], [37].

This work aims to address these gaps and show that synthetic data can be reliably used for training LSTM-based algorithms, not only in combination with but also as substitutes for real data, thus enabling effective and time-efficient heavy vehicle rollover prevention. Differently from existing literature where only simple test tracks or known vehicles are considered [38], highly complex simulation scenarios (e.g., the Stelvio Pass) and various empirical maneuvers were leveraged. Vertical load transfers (Load Transfer Ratios, LTRs) were also estimated, crucial rollover indicators that are challenging to measure on-field. Moreover, different noise levels in synthetic data were systematically investigated to quantify and approximate the on-board noise interference observed in real data. By doing so, a methodology to create large, noise-tailored synthetic datasets and train LSTM networks capable of dealing with realistic disturbances was provided. The innovative contribution lies, particularly, in designing a training pipeline that integrates large synthetic datasets, calibrates noise levels, and demonstrates the feasibility of achieving comparable or better results than those based on scarce and expensive real data acquisitions. The LSTM choice is further justified by preliminary tests with other models (e.g., bi-LSTM, Random Forest) that yielded unsatisfactory results.

In summary, the contributions are as follows:

- Combined Experimental and Simulation Framework: A heavy vehicle model in IPG Carmaker® was defined, and on-field data were collected on its physical counterpart. The simulation replicated realistic, complex tracks and maneuvers, enabling the generation of large synthetic datasets.
- Noise Handling and Estimation: By systematically adding artificial noise with various Signal-to-Noise Ratio (SNR) levels to synthetic data, real on-board noise was approximated and LSTM networks were trained to be capable of dealing with disturbances, thus identifying a plausible noise spectrum for on-field measurements.
- Rollover Indicators and Replaceability: The results show that the LSTM networks trained on noise-adjusted synthetic data can estimate the vehicle's roll angle with accuracy at least comparable to the one achieved using limited real data. Moreover, these networks also estimate LTRs, providing a more comprehensive understanding of incipient rollover conditions and reducing the need for exhaustive real-world testing.

The rest of the article is organized as follows. Section II provides a general overview of the system. Section III describes the methodology for data acquisition, synthetic data generation, noise analysis, and LSTM network design. Section IV presents and discusses the experimental results. Finally, Section V highlights conclusions and future perspectives.

II. SYSTEM OVERVIEW

The overall process is organized into two main phases: *Data Acquisition* and *Data Analysis*, as reported in Figs. 1 and 2, respectively. In the Data Acquisition phase, two complementary approaches are combined, one relying on synthetic data generation through the IPG Carmaker® simulation environment and one based on on-field measurements recorded during real heavy vehicle maneuvers. Although it is not possible to provide photos of the on-field setup due to confidentiality (the vehicle is still a prototype), the instrumentation and data collection procedures follow standard automotive testing protocols as described in Section III and Table 1.

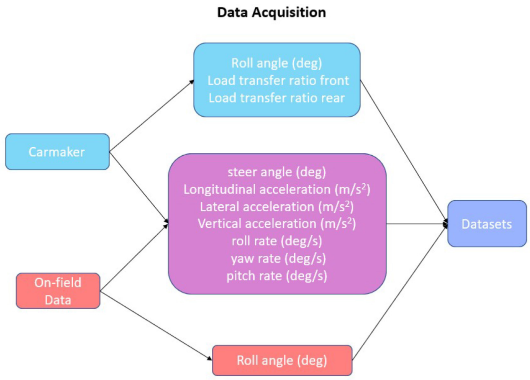


FIGURE 1. System Overview - Data acquisition.

TABLE 1. Equipment: main sensors used on board.

Type	Description
Cable Transducers	1x ID/JX-P510-30-N11-10S-N2K 1x ID/JX-P510-30-N11-10S-N4K
Rotative Position Transducers	4x Gefran PS-11-E-0-472 1x Gefran PS-20-B-0-472
Inertial Sensor + GPS	1x Kistler Correvit S-Motion DTI Type 2055A 1x Kistler GPS

The *Carmaker®* block represents the simulation environment where a high-fidelity vehicle model and complex road scenarios are built [30], allowing virtually unlimited data generation. The *On-Field data* block corresponds to real measurements acquired onboard the vehicle, including accelerations, rotation rates, and steer angle, which are time-dependent signals crucial for predicting rollover events. Both sources of data converge into the central block of Fig. 1, where the selected variables–easily measured on-board–serve as LSTM inputs. The LSTM model is chosen given its suitability for handling time-series data and capturing temporal dependencies more effectively than traditional feed-forward networks [38], [39], [40], [41].

On the output side, the roll angle is available from the physical vehicle, while the Load Transfer Ratios (LTRs) can be directly (and only) computed from the simulation. Thus, the on-field and synthetic data streams are aligned so that the LSTM-based methodologies can learn patterns relevant to incipient rollover conditions. The resulting *Datasets* block corresponds to 28 unique synthetic maneuvers and 12 empirical ones, each crafted or selected to highlight rollover-related scenarios.

The Data Analysis phase (Fig. 2) starts from the previously created datasets and structures them according to specific purposes. Here, three different types of training are

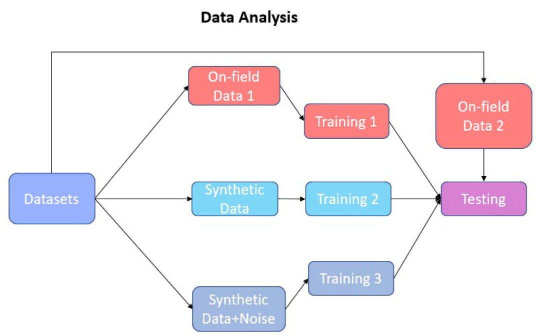


FIGURE 2. System Overview - Data Analysis.

performed to assess the performance of the LSTM networks under various conditions: one based exclusively on empirical data (*Training 1*), one on synthetic data without additional noise (*Training 2*), and one on synthetic data with artificially introduced noise (*Training 3*). This approach ensures a thorough understanding of how noise and data source affect the LSTM's predictions. The trained networks are then tested on four sets of on-field maneuvers not included in the training, evaluating their estimation accuracy through metrics such as RMSE, MSE, and MAE. This evaluation also helps to infer the on-field noise characteristics and confirm that synthetic data, appropriately adjusted for noise, can replicate or even improve upon results obtained with networks trained on real measurements.

III. METHODOLOGY

As anticipated, the methodology consists of two main phases: data collection/generation and the subsequent analysis and verification of their effectiveness in modeling heavy vehicle rollover through neural networks algorithms. The problem addressed is inherently time-dependent: the vehicle's dynamic state evolves over time, and the rollover indicators (roll angle, LTRs) depend on the temporal history of acceleration, rotation rates, and steering inputs.

A. DATA COLLECTION AND GENERATION

The first phase involves obtaining empirical and synthetic datasets specifically designed to highlight incipient rollover conditions. On-field measurements were acquired on-board an instrumented heavy vehicle with the following characteristics: 3230 mm wheelbase, 1710 mm track, 7000 kg total mass, and a vertical layout leading to a high center of gravity (2100 mm maximum and 345 mm minimum height under differential). These parameters increase rollover risk. The vehicle operated under different scenarios and terrains, using standard automotive-grade sensors. Table 1 details the sensors employed, including cable transducers, rotary potentiometers, and an IMU with GPS integration. While the IMU can provide estimates of roll angle, more direct measurements were obtained using cable transducers on the suspensions. This measurement configuration ensures time-series data acquisition, where each variable $x_i(t)$ is indexed by time t.

This structure led to the acquisition of 12 maneuvers, from which 8 were selected for the data analysis phase. However, the small quantity of empirical data (around 50000 pieces of information per variable for the total of the final 4 maneuvers used in training) poses limitations in properly modeling the time-dependent signals for rollover prediction [29]. Thus, an alternative to costly and time-consuming on-field experiments is data generation via IPG Carmaker®, a simulation platform capable of replicating various vehicle types, roads, and maneuvers [30]. A heavy vehicle model was designed and parameterized to match the real test vehicle's dynamics. As a result, 28 synthetic maneuvers were generated, and 16 of these were selected to form a large-scale, uniform training set (each maneuver lasted 500 seconds and was sampled every 0.01, for a total of 50000 data points per variable, 800000 total information).

To address the critical challenge of noise affecting on-board instrumentation, a second version of these 16 synthetic datasets was created by adding controlled white Gaussian noise at different Signal-to-Noise Ratios (SNRs), ranging from 0 dB to 50 dB. This type of disturbance is very commonly used to approximate road surface vibrations [42], [43]. Since real on-board measurements inherently contain unknown noise levels, this approach allowed us to approximate real noise conditions, thereby enabling an informed comparison between empirical and synthetic data in the subsequent analysis phase.

B. MODEL AND THEORETICAL BACKGROUND

Predicting rollover indicators from time-series data requires a model capable of handling temporal dependencies. Let $x(t) = [x_1(t), x_2(t), \ldots, x_7(t)]$ represent the input vector at time t, including accelerations, rotation rates, and steering angle. The goal is to estimate outputs $y(t) = [\text{Roll angle}(t), \text{LTR}_{\text{front}}(t), \text{LTR}_{\text{rear}}(t)]$ (see Table 2). It has to be severely underscored that LTRs are outputs only with synthetically generated data (it is not possible to acquire this variables empirically). Formally, the function of interest is:

$$\hat{y}(t) = f(x(t), x(t-1), \ldots, x(0)), \tag{1}$$

where the outputs depend on the entire history of inputs.

Recurrent Neural Networks (RNNs) are well-suited for modeling such temporal relationships. However, standard RNNs suffer from vanishing or exploding gradients when dealing with long sequences. LSTM networks overcome these issues by introducing a memory cell and gating mechanisms:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{2}$$
$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{3}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{4}$$
$$\tilde{c}t = \tanh(W_c x_t + U_c h_{t-1} + b_c), \tag{5}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{6}$$
$$h_t = o_t \odot \tanh(c_t), \tag{7}$$

TABLE 2. LSTM: inputs and outputs.

Type	Description	Unit of Meaurement
Inputs	Longitudinal Acceleration	m/s^2
	Lateral Acceleration	m/s^2
	vertical Acceleration	m/s^2
	Roll rate	$^\circ/s$
	Yaw rate	$^\circ/s$
	Pitch rate	$^\circ/s$
	Steer Angle	$^\circ$
Outputs	Roll Angle	$^\circ$
	Load Transfer Ratio Front	–
	Load Transfer Ratio Rear	–

where σ is the logistic sigmoid, i_t, f_t, o_t are the input, forget, and output gates, respectively, c_t is the cell state, and h_t the hidden state at time t. This architecture can retain information over long time intervals, a crucial feature for rollover detection, where subtle changes in vehicle dynamics accumulate over time.

While other models (e.g., bi-LSTM, Random Forest) were preliminarily tested, they did not achieve satisfactory accuracy for the examined scenario. Random Forest models, commonly suited for classification tasks, struggled with the continuous, time-dependent regression problem. Bi-LSTM networks did not improve results and increased model complexity. Thus, the LSTM architecture, supported by strong literature evidence in time-series analysis [35], [36], [37], [38], [39], [40], [41], was selected as the baseline model.

C. METHODOLOGICAL STEPS AND LSTM CONFIGURATION

All maneuvers (both real and synthetic) were conducted/designed to highlight incipient rollover conditions. Although LTRs cannot be measured on-field, their inclusion in the outputs from synthetic data enriches the training, allowing the LSTM model to learn more comprehensive rollover dynamics and opening potential future research developments.

Before training, data standardization was performed. Each LSTM model's architecture follows Table 3, where the number of hidden units (100) and other hyperparameters were chosen after empirical testing to balance training time and accuracy. More than 100 hidden units did not yield improvements, while fewer led to decreased performance. Similarly, adaptive methods (Adam), stepwise learning rate reduction, and gradient clipping were applied to ensure stable training.

It is brought again to attention that, for the empirical-data-trained networks, only the roll angle is estimated since LTRs are not directly available from on-field measurements. In contrast, the synthetic-data-trained networks can estimate both roll angle and LTRs. After training, the LSTM models were tested on previously unseen on-field maneuvers to assess their prediction accuracy. Prior to the final comparison, a preliminary validation (Table 4) confirmed that the network trained on synthetic data without additional noise could accurately reconstruct roll angle and LTRs on unseen synthetic maneuvers. This step ensures that the chosen architecture and training process are sufficiently robust before moving on to real-world comparisons.

TABLE 3. LSTM: Architecture.

Type	Description	Characteristics
Layers	Sequence input	Number of Features (7)
	LSTM	Number of hidden units (100)
	Fully connected	Number of Responses (3)
	Regression	
Hyperparameters	Adam	Adaptive moment estimation
	MaxEpochs	850
	GradientThreshold	1
	InitialLearnRate	0.01
	LearnRateDropPeriod	425
	LearnRateDropFactor	0.2

TABLE 4. RMSE, MSE, MAE for preliminary algorithm validation, networks trained on simulation data estimating simulated data on Stelvio and Bernina tracks, other than a FishHook maneuver.

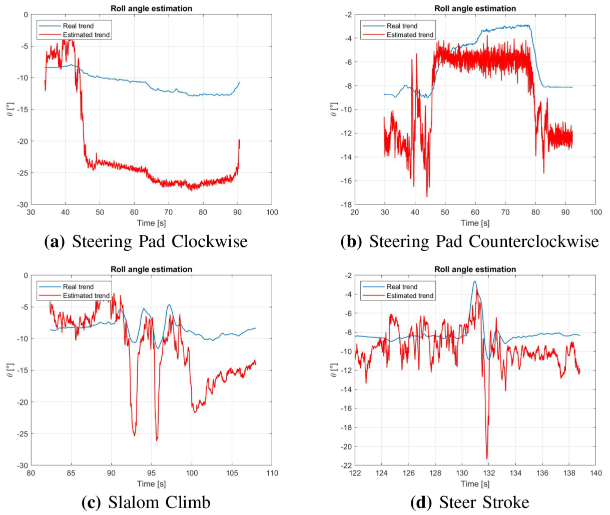
Maneuver	Error	Values
Bernina	RMSE Roll	0,0718 deg
	MSE Roll	0,0052 deg
	MAE Roll	0,0033 deg
	RMSE LTR Front	0,0196 deg
	MSE LTR Front	3,857e-04 deg
	MAE LTR Front	0,0040 deg
	RMSE LTR Rear	0,0442 deg
	MSE LTR Rear	0,0020 deg
	MAE LTR Rear	0,0220 deg
Stelvio	RMSE Roll	0,1096 deg
	MSE Roll	0,0120 deg
	MAE Roll	0,0695 deg
	RMSE LTR Front	0,0139 deg
	MSE LTR Front	1,926e-04 deg
	MAE LTR Front	0,0074 deg
	RMSE LTR Rear	0,0571 deg
	MSE LTR Rear	0,0033 deg
	MAE LTR Rear	0,0351 deg
FishHook	RMSE Roll	0,5540 deg
	MSE Roll	0,3069 deg
	MAE Roll	0,1295 deg
	RMSE LTR Front	0,0237 deg
	MSE LTR Front	5,595e-04 deg
	MAE LTR Front	0,0075 deg
	RMSE LTR Rear	0,0190 deg
	MSE LTR Rear	3,609e-04 deg
	MAE LTR Rear	0,0088 deg

With these validations complete, the next step is comparing the performance of LSTM models trained on real data and those trained on synthetic data, including noise-adjusted variants. This comparison, discussed in the Results section, focuses on 4 on-field maneuvers not present within the training datasets: Steering Pad at Increasing Speed Clockwise (Steering Pad Clockwise), Steering Pad at Increasing Speed Counterclockwise (Steering Pad Counterclockwise), Slalom Climb (Slalom Climb), and Left Steer Stroke (Steer Stroke).

IV. RESULTS

A. RESULTS: ALGORITHMS TRAINED ON REAL DATA

The AI algorithms trained on real data, following the testing phase, return the outcomes reported in Figure 3. In these plots, "Time" represents the temporal progression of the maneuver, i.e., each data point corresponds to a sampling instant. This allows us to observe how the roll angle evolves during each driving scenario. It is underscored that the starting point is not zero. This is intended to highlight just the maneuver of interest, not all the preparation to execute the same, consequently having a more evident representation of the incipient rollover condition.



(a) Steering Pad Clockwise (b) Steering Pad Counterclockwise (c) Slalom Climb (d) Steer Stroke

FIGURE 3. Roll angle estimate with empirical data trained neural networks.

Observing Figure 3, the model predictions show notable discrepancies from the measured roll angle, especially for the Steering Pad maneuver in the Clockwise direction and for the Slalom Climb. The differences are likely due to the limited and uneven real-data training set, as these maneuvers were not well represented during training. Moreover, the measurement noise and variability in the on-field data–arising from sensor precision, road conditions, and other external factors–further complicate the learning task. Since the networks trained exclusively on empirical data are inherently constrained by the scarcity and specificity of the maneuvers, their ability to generalize to new conditions is limited. Additionally, as mentioned, no LTRs estimation is possible here because these indicators are not directly measurable on-board.

Contrarily, the Steering Pad Counterclockwise and Steer Stroke maneuvers exhibit better alignment between predicted and actual roll angle. This suggests that these maneuvers are more closely represented in the training set or represent more stable conditions that the LSTM can more easily learn. It underlines the importance of ensuring maneuver diversity and data uniformity during the collection phase to improve model robustness.

To quantify these observations, Table 5 reports the RMSE, MSE, and MAE for each maneuver. As anticipated, the highest errors are for the Steering Pad Clockwise and Slalom Climb maneuvers, confirming the need for more representative and comprehensive training data if we were to rely solely on real measurements. Increasing the variety and similarity of maneuvers in the training set or standardizing maneuver durations and sampling rates could help reduce these errors. However, such approaches come with significant time and cost implications [29].

B. RESULTS: ALGORITHMS TRAINED ON SYNTHETIC DATA

To address the limitations identified with the real-data-trained models, the performance of LSTM networks trained on synthetic datasets is examined. Using IPG Carmaker®, it is possible to generate large-scale, diversified datasets that capture a broad range of vehicle behaviors. However, synthetic data alone are not enough: real on-board measurements contain noise, and to emulate these conditions, controlled white Gaussian noise is introduced at various Signal-to-Noise Ratios (SNRs, from 0 to 50 dB).

Tables 6, 7, and 8 showcase the results obtained by training the LSTM on synthetic data augmented with different SNR levels. The key finding is that an SNR of 10 dB yields the best performance. At this noise level, the synthetic data closely resemble the conditions found in real measurements, enabling the model to generalize better to the on-field tests. Introducing too much noise (0 dB) degrades performance, while too little noise (20 dB, 30 dB, 40 dB, 50 dB) makes the model susceptible to overfitting "clean" signals and struggling when faced with the inherent imperfections of real-world measurements.

For completeness, Table 9 reports results where synthetic data contain no artificially introduced noise. As expected, in this scenario, performance is even worse than at high SNR levels, confirming that some level of controlled noise

TABLE 5. RMSE, MSE, MAE associated to roll angle estimate with empirical data trained neural networks.

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	12,52 *deg*
	MSE Roll	156,9 *deg*
	MAE Roll	11,66 *deg*
Steering Pad Counterclockwise	RMSE Roll	3,288 *deg*
	MSE Roll	10,81 *deg*
	MAE Roll	2,867 *deg*
Slalom Climb	RMSE Roll	5,710 *deg*
	MSE Roll	32,60 *deg*
	MAE Roll	4,528 *deg*
Steer Stroke	RMSE Roll	2,329 *deg*
	MSE Roll	5,423 *deg*
	MAE Roll	1,857 *deg*

TABLE 6. RMSE, MSE, MAE, tests on real data and training with synthetic data, SNR of 0dB (left) and 10dB (right).

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	2,601 *deg*
	MSE Roll	6,764 *deg*
	MAE Roll	2,184 *deg*
Steering Pad Counterclockwise	RMSE Roll	4,439 *deg*
	MSE Roll	19,70 *deg*
	MAE Roll	3,353 *deg*
Slalom Climb	RMSE Roll	3,391 *deg*
	MSE Roll	11,50 *deg*
	MAE Roll	2,735 *deg*
Steer Stroke	RMSE Roll	3,550 *deg*
	MSE Roll	12,61 *deg*
	MAE Roll	3,256 *deg*

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	1,945 *deg*
	MSE Roll	3,781 *deg*
	MAE Roll	1,059 *deg*
Steering Pad Counterclockwise	RMSE Roll	2,140 *deg*
	MSE Roll	4,580 *deg*
	MAE Roll	1,770 *deg*
Slalom Climb	RMSE Roll	1,743 *deg*
	MSE Roll	3,040 *deg*
	MAE Roll	1,098 *deg*
Steer Stroke	RMSE Roll	1,801 *deg*
	MSE Roll	3,243 *deg*
	MAE Roll	0,953 *deg*

TABLE 7. RMSE, MSE, MAE, test on real data and training with synthetic data, SNR equal to 20dB (left) and 30dB (right).

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	1,953 *deg*
	MSE Roll	3,812 *deg*
	MAE Roll	1,054 *deg*
Steering Pad Counterclockwise	RMSE Roll	3,165 *deg*
	MSE Roll	10,02 *deg*
	MAE Roll	2,370 *deg*
Slalom Climb	RMSE Roll	2,218 *deg*
	MSE Roll	4,918 *deg*
	MAE Roll	1,529 *deg*
Steer Stroke	RMSE Roll	2,037 *deg*
	MSE Roll	4,149 *deg*
	MAE Roll	1,230 *deg*

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	2,402 *deg*
	MSE Roll	5,768 *deg*
	MAE Roll	1,864 *deg*
Steering Pad Counterclockwise	RMSE Roll	6,238 *deg*
	MSE Roll	38,91 *deg*
	MAE Roll	4,778 *deg*
Slalom Climb	RMSE Roll	3,332 *deg*
	MSE Roll	11,10 *deg*
	MAE Roll	2,334 *deg*
Steer Stroke	RMSE Roll	3,006 *deg*
	MSE Roll	9,037 *deg*
	MAE Roll	2,363 *deg*

TABLE 8. RMSE, MSE, MAE, test on real data and training with synthetic data, SNR equal to 40dB (left) and 50dB (right).

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	2,953 deg
	MSE Roll	8,719 deg
	MAE Roll	2,335 deg
Steering Pad Counterclockwise	RMSE Roll	3,507 deg
	MSE Roll	12,30 deg
	MAE Roll	2,788 deg
Slalom Climb	RMSE Roll	3,757 deg
	MSE Roll	14,12 deg
	MAE Roll	2,377 deg
Steer Stroke	RMSE Roll	2,970 deg
	MSE Roll	8,821 deg
	MAE Roll	1,916 deg

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	3,518 deg
	MSE Roll	12,37 deg
	MAE Roll	2,873 deg
Steering Pad Counterclockwise	RMSE Roll	4,427 deg
	MSE Roll	19,60 deg
	MAE Roll	3,437 deg
Slalom Climb	RMSE Roll	2,706 deg
	MSE Roll	7,320 deg
	MAE Roll	1,733 deg
Steer Stroke	RMSE Roll	3,258 deg
	MSE Roll	10,62 deg
	MAE Roll	1,875 deg

TABLE 9. RMSE, MSE, MAE, test on real data and training with synthetic data, no SNR).

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	3,357 deg
	MSE Roll	11,27 deg
	MAE Roll	2,942 deg
Steering Pad Counterclockwise	RMSE Roll	5,317 deg
	MSE Roll	28,27 deg
	MAE Roll	4,252 deg
Slalom Climb	RMSE Roll	5,597 deg
	MSE Roll	31,33 deg
	MAE Roll	4,300 deg
Steer Stroke	RMSE Roll	3,610 deg
	MSE Roll	13,03 deg
	MAE Roll	2,668 deg

TABLE 10. RMSE, MSE, MAE, tests on real data and training with synthetic data, SNR of 10dB (left), and real data (right).

Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	1,945 deg
	MSE Roll	3,781 deg
	MAE Roll	1,059 deg
Steering Pad Counterclockwise	RMSE Roll	2,140 deg
	MSE Roll	4,580 deg
	MAE Roll	1,770 deg
Slalom Climb	RMSE Roll	1,743 deg
	MSE Roll	3,040 deg
	MAE Roll	1,098 deg
Steer Stroke	RMSE Roll	1,801 deg
	MSE Roll	3,243 deg
	MAE Roll	0,953 deg

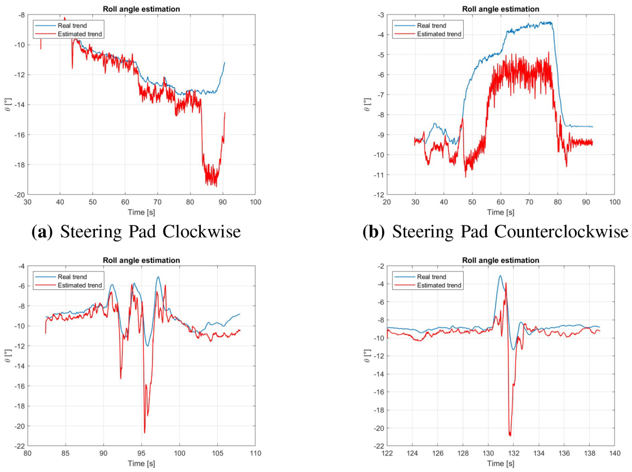
Maneuver	Error	Values
Steering Pad Clockwise	RMSE Roll	12,52 deg
	MSE Roll	156,9 deg
	MAE Roll	11,66 deg
Steering Pad Counterclockwise	RMSE Roll	3,288 deg
	MSE Roll	10,81 deg
	MAE Roll	2,867 deg
Slalom Climb	RMSE Roll	5,710 deg
	MSE Roll	32,60 deg
	MAE Roll	4,528 deg
Steer Stroke	RMSE Roll	2,329 deg
	MSE Roll	5,423 deg
	MAE Roll	1,857 deg

is essential to improving realism and thus generalization to real data.

Finally, comparing the best synthetic-based results (SNR = 10 dB) with those obtained from the real-data-trained network (Table 10) highlights the advantage of using large, noise-calibrated synthetic datasets. While in principle collecting more real data could close this gap, it would be impractical and expensive. Thus, a combination of simulation and controlled noise introduction emerges as an effective solution for training robust rollover prediction models.

As shown in Figure 4, the predicted roll angle for the synthetic-data-trained LSTM at SNR = 10 dB closely follows the real measurements, even for the challenging maneuvers. Moreover, although not shown for brevity, the synthetic-data-trained models can also output LTRs. With appropriate validation, these additional outputs could further enhance the understanding of rollover conditions.

In summary, these results confirm that carefully engineered synthetic datasets, combined with realistic noise modeling, enable the LSTM models to achieve greater accuracy and robustness compared to those relying solely on a limited set of real measurements. This approach offers a cost-effective and flexible method to train and test advanced neural-based methodologies for rollover prevention in heavy vehicles.



(a) Steering Pad Clockwise (b) Steering Pad Counterclockwise (c) Slalom Climb (d) Steer Stroke

FIGURE 4. Roll angle estimate with synthetic data trained neural networks, SNR 10dB.

V. CONCLUSION AND FUTURE DEVELOPMENTS

The original objective, the comparison of LSTM algorithms based on empirical and simulated data was achieved. The estimation of three key parameters to judge the incipient rollover condition of the vehicle was obtained, leading to study roll angle and LTRs of the two axles. Given the data available and the methodologies adopted, this comparison resulted in significantly superior performance by LSTM-type neural networks based on synthetic data compared to LSTM networks with same architecture based on real data, suggesting that the first can be used as substitutes of the second. This is due to the virtually unlimited numerousness of the synthetic datasets and the handling, at a synthetic level, of the noise characterizing in-vehicle measurements. Certainly more data can be acquired at empirical level to improve the outcomes of LSTM networks based on real data. At the same time, this would require probably excessive costs in terms of capitals and time. In details, the LSTM networks trained with artificially implemented noise on synthetic data present better results in terms of accuracy (RMSE, MSE, MAE) in comparison with the ones attained by LSTM networks based on empirical data. The synthetic data based algorithms, in addition, provide an estimate of the noise affecting on-board measurement other than the trends of the LTRs of the heavy vehicles, not attainable by neural networks simply based on empirical data. These results suggest that the vehicle model used in the simulation phase is now sufficiently accurate to effectively approximate the behavior of the real vehicle, implying that the simulation environment itself (IPG CarMaker®) is more than effective in its representation of reality. This implicate that it should be possible, other than generally preventing the heavy vehicle rollover, realistically simulate the effects of rollover accidents on the driver without being limited by data.

In summary, the outcomes are exceptionally positive. Future developments can include a further refinement of the dataset and in-vehicle testing to verify real time operation of the designed networks.

REFERENCES

[1] T. Zhu, X. Yin, B. Li, and W. Ma, "A reliability approach to development of rollover prediction for heavy vehicles based on SVM empirical model with multiple observed variables," *IEEE Access*, vol. 8, pp. 89367–89380, 2020.

[2] Y. Sellami, H. Imine, A. Boubezoul, and J.-C. Cadiou, "Rollover risk prediction of heavy vehicles by reliability index and empirical modelling," *Veh. Syst. Dyn.*, vol. 56, no. 3, pp. 385–405, 2018.

[3] J. L. Evans, S. A. Batzer, and S. B. Andrews, "Evaluation of heavy truck rollover accidents," in *Proc. 19th Int. Tech. Conf. Enhanced Safe. Veh. (ESV)*, 2005, pp. 1–31.

[4] J. L. Evans, S. A. Batzer, S. B. Andrews, and R. M. Hooker, "Evaluation of heavy truck rollover crashworthiness," *Transp*, 2006, to be published.

[5] *Bilan Social Annuel du Transport Routier de Marchandises*, SOeS, Mexico City, Mexico, 2015.

[6] *Traffic Safety Facts 2016: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System, NHTSA*, US Dept. Transp., Washington, DC, USA, 2017.

[7] X. Wu, X. Ge, S. Luo, and H. Huang, "Research on vehicle rollover and control," in *Proc. 2nd IEEE Int. Conf. Adv. Comput. Control (ICACC)*, 2010, pp. 510–514.

[8] Q. Ren, M. Xu, and X. Yan, "An investigation of heterogeneous impact, temporal stability, and aggregate shift in factors affecting the driver injury severity in single-vehicle rollover crashes," *Accident Anal. Prevention*, vol. 200, Jun. 2024, Art. no. 107562.

[9] C. Chen, G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian, "Investigating driver injury severity patterns in rollover crashes using support vector machine models," *Accid. Anal. Prevent.*, vol. 90, pp. 128–139, May 2016.

[10] G. Azimi, A. Rahimi, H. Asgari, and X. Jin, "Severity analysis for large truck rollover crashes using a random parameter ordered logit model," *Accid. Anal. Prevent.*, vol. 135, Feb. 2020, Art. no. 105355.

[11] A. J. Anarkooli, M. Hosseinpour, and A. Kardar, "Investigation of factors affecting the injury severity of single-vehicle rollover crashes: A random-effects generalized ordered probit model," *Accid. Anal. Prevent.*, vol. 106, pp. 399–410, Sep. 2017.

[12] F. Braghin, F. Cheli, R. Corradi, G. Tomasini, and E. Sabbioni, "Active anti-rollover system for heavy-duty road vehicles," *Veh. Syst. Dyn.*, vol. 46, no. 1, pp. 653–668, 2008.

[13] A. Tarko, T. Hall, M. Romero, and C. G. L. Jiménez, "Evaluating the rollover propensity of trucks—A roundabout example," *Accid. Anal. Prevent.*, vol. 91, pp. 127–134, Jun. 2016.

[14] M. Ghazali, M. Durali, and H. Salarieh, "Path-following in model predictive rollover prevention using front steering and braking," *Veh. Syst. Dyn.*, vol. 55, no. 1, pp. 121–148, 2017.

[15] L. Li, Y. Lu, R. Wang, and J. Chen, "A three-dimensional dynamics control framework of vehicle lateral stability and rollover prevention via active braking with MPC," *IEEE Trans. Ind. Electron.*, vol. 64, no. 4, pp. 3389–3401, Apr. 2017.

[16] T. Zhu, C. Zong, B. Wu, and Z. Sun, "Rollover warning system of heavy duty vehicle based on improved TTR algorithm," *Jixie Gongcheng Xuebao/J. Mech. Eng.*, vol. 47, no. 10, pp. 88–94, 2011.

[17] H. Li, Y. Zhao, H. Wang, and F. Lin, "Design of an improved predictive LTR for rollover warning systems," *J. Brazil. Soc. Mech. Sci. Eng.*, vol. 39, no. 10, pp. 3779–3791, 2017.

[18] X. Zhang, Y. Yang, K. Guo, J. Lv, and T. Peng, "Contour line of load transfer ratio for vehicle rollover prediction," *Veh. Syst. Dyn.*, vol. 55, no. 11, pp. 1748–1763, 2017.

[19] T. P. Wenzel and M. Ross, "The effects of vehicle model and driver behavior on risk," *Accid. Anal. Prevent.*, vol. 37, no. 3, pp. 479–494, 2005.

[20] C. S. Parenteau et al., "Field relevance of a suite of rollover tests to real-world crashes and injuries," *Accid. Anal. Prevent.*, vol. 35, no. 1, pp. 103–110, 2003.

[21] M. F. Aljarrah, M. A. Khasawneh, and A. A. Al-Omari, "Investigating key factors influencing the severity of drivers injuries in car crashes using supervised machine learning techniques," *J. Eng. Sci. Technol. Rev.*, vol. 12, no. 4, pp. 15–27, 2019.

[22] C. Badgujar, S. Das, D. M. Figueroa, D. Flippo, and S. Welch, "Deep neural networks to predict autonomous ground vehicle behavior on sloping terrain field," *J. Field Robot.*, vol. 40, no. 4, pp. 919–933, 2023.

[23] X. Chen, W. Chen, L. Hou, H. Hu, X. Bu, and Q. Zhu, "A novel data-driven rollover risk assessment for articulated steering vehicles using RNN," *J. Mech. Sci. Technol.*, vol. 34, no. 5, pp. 2161–2170, 2020.

[24] J. G. Guzmán, L. P. González, J. P. Redondo, M. M. M. Martínez, and M. J. L. Boada, "Real-time vehicle roll angle estimation based on neural networks in IoT low-cost devices," *Sensors*, vol. 18, no. 7, p. 2188, 2018.

[25] Z. Zhu, R. Torisu, J. Takeda, E. Mao, and Q. Zhang, "Neural network for estimating vehicle behaviour on sloping terrain," *Biosyst. Eng.*, vol. 91, no. 4, pp. 403–411, 2005.

[26] P.-P. Chao, R.-Y. Zhang, Y.-D. Wang, H. Tang, and H.-L. Dai, "Warning model of new energy vehicle under improving time-to-rollover with neural network," *Meas. Control*, vol. 55, nos. 9–10, pp. 1004–1015, 2022.

[27] K. Treetipsounthorn et al., "Vehicle rollover detection in tripped and untripped rollovers using recurrent neural networks," *Adv. Sci., Technol. Eng. Syst.*, vol. 5, no. 6, pp. 228–238, 2020.

[28] B. Zhu, Q. Piao, J. Zhao, and L. Guo, "Integrated chassis control for vehicle rollover prevention with neural network time-to-rollover warning metrics," *Adv. Mech. Eng.*, vol. 8, no. 2, pp. 1–13, 2016.

[29] M. M. Baldi, G. Perboli, and R. Tadei, "Driver maneuvers inference through machine learning," in *Proc. Int. Workshop Mach. Learn., Optim., Big Data*, 2016, pp. 182–192.

[30] A. Tota, L. Dimauro, F. Velardocchia, G. Paciullo, and M. Velardocchia, "An intelligent predictive algorithm for the anti-rollover prevention of heavy vehicles for off-road applications," *Machines*, vol. 10, p. 835, Sep. 2022.

[31] G. Perboli and E. Arabnezhad, "A machine learning-based DSS for mid and long-term company crisis prediction," *Expert Syst. Appl.*, vol. 174, Feb. 2021, Art. no. 114758.

[32] S. Sharma, J. Henderson, and J. Ghosh, "CERTIFAI: A common framework to provide explanations and Analyse the fairness and robustness of black-box models," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, New York, NY, USA, 2020, pp. 166–172.

[33] X. Huang, H. Huang, J. Wu, M. Yang, and W. Ding, "Sound quality prediction and improving of vehicle interior noise based on deep convolutional neural networks," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113657.

[34] M. Y. Shams, T. Abd El-Hafeez, and E. Hassan, "Acoustic data detection in large-scale emergency vehicle sirens and road noise dataset," *Expert Syst. Appl.*, vol. 249, Sep. 2024, Art. no. 123608.

[35] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Procedia CIRP*, vol. 99, pp. 650–655, May 2021.

[36] X.-H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, 2019.

[37] F. Velardocchia, G. Perboli, and A. Vigliani, "Analysis of heavy vehicles rollover with artificial intelligence techniques," in *Proc. Int. Conf. Mach. Learn., Optim., Data Sci.*, 2023, pp. 294–308.

[38] L. P. González, S. S. Sánchez, J. Garcia-Guzman, M. J. L. Boada, and B. L. Boada, "Simultaneous estimation of vehicle roll and sideslip angles through a deep learning approach," *Sensors*, vol. 20, no. 13, pp. 1–18, 2020.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.

[41] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[42] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Min. Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[43] Y. Liu and D. Cui, "Collaborative model analysis on ride comfort and handling stability," *J. Vibroeng.*, vol. 21, no. 6, pp. 1724–1737, 2019.

[44] T. Lenkutis, A. Čerškus, N. Šešok, A. Dzedzickis, and V. Bučinskas, "Road surface profile synthesis: Assessment of suitability for simulation," *Symmetry*, vol. 13, no. 1, pp. 1–14, 2021.