Data-driven Semi-supervised Machine Learning with Surrogate Safety Measures for Abnormal Driving Behavior Detection

Zhang, Lanxin; Dong, Yongqi; Farah, Haneen; Zgonnikov, Arkady; van Arem, Bart

Publication date
2023

Document Version
Final published version

Citation (APA)
Zhang, L., Dong, Y., Farah, H., Zgonnikov, A., & van Arem, B. (2023). *Data-driven Semi-supervised Machine Learning with Surrogate Safety Measures for Abnormal Driving Behavior Detection*. Poster session presented at the 35th Annual Conference of International Cooperation on Theories and Concepts in Traffic safety (ICTCT), Catania, Italy.

Data-driven Semi-supervised Machine Learning with Surrogate Safety Measures for Abnormal Driving Behavior Detection

Lanxin Zhang[#], Yongqi Dong[#,*], Haneen Farah, Arkady Zgonnikov, Bart van Arem
zhanglx2018@hotmail.com

Background

- Road traffic safety is a worldwide issue, in which human factors contribute to 92.9% of car accidents.
- Abnormal driving behaviors are major road safety risks.
- ML approaches show promise in detecting these behaviors.
- Most studies, however, rely on fully supervised ML methods, needing substantial labeled data.
- Furthermore, Surrogate Safety Measures (SSMs) as a proactive road safety assessment, has never been fully utilized.
- SSMs help detect excessive crash risk and understand crash-precipitating conditions.

Aim

The main aims of this study are to:

- develop a novel data-driven approach for detecting abnormal driving behaviors using real-world naturalistic driving data.
- utilize semi-supervised machine learning, combining self-supervised training with partly labeled data, to enhance the accuracy and efficiency of the detection system.
- introduce SSMs as input features to improve the performance of the detection model.

The framework of the proposed pipeline

- **Identified Abnormal Driving Behavior in CitySim Dataset**
  - Rapid acceleration and emergency brake behavior
  - Rapid lane-changing behavior
  - Close lane-changing behavior
- **Surrogate safety measures (2D-Time-To-Collision)**
- **Baseline models**
  - Isolation Forest
  - Robust Covariance
- **Hierarchical Extreme Learning Machines based Semi-Supervised Machine Learning**

Evaluation Metrics

- **Accuracy**
- **Precision**
- **True Positive Rate**
- **F1-Meassure**
- **Recall**
- **False Positive Rate**

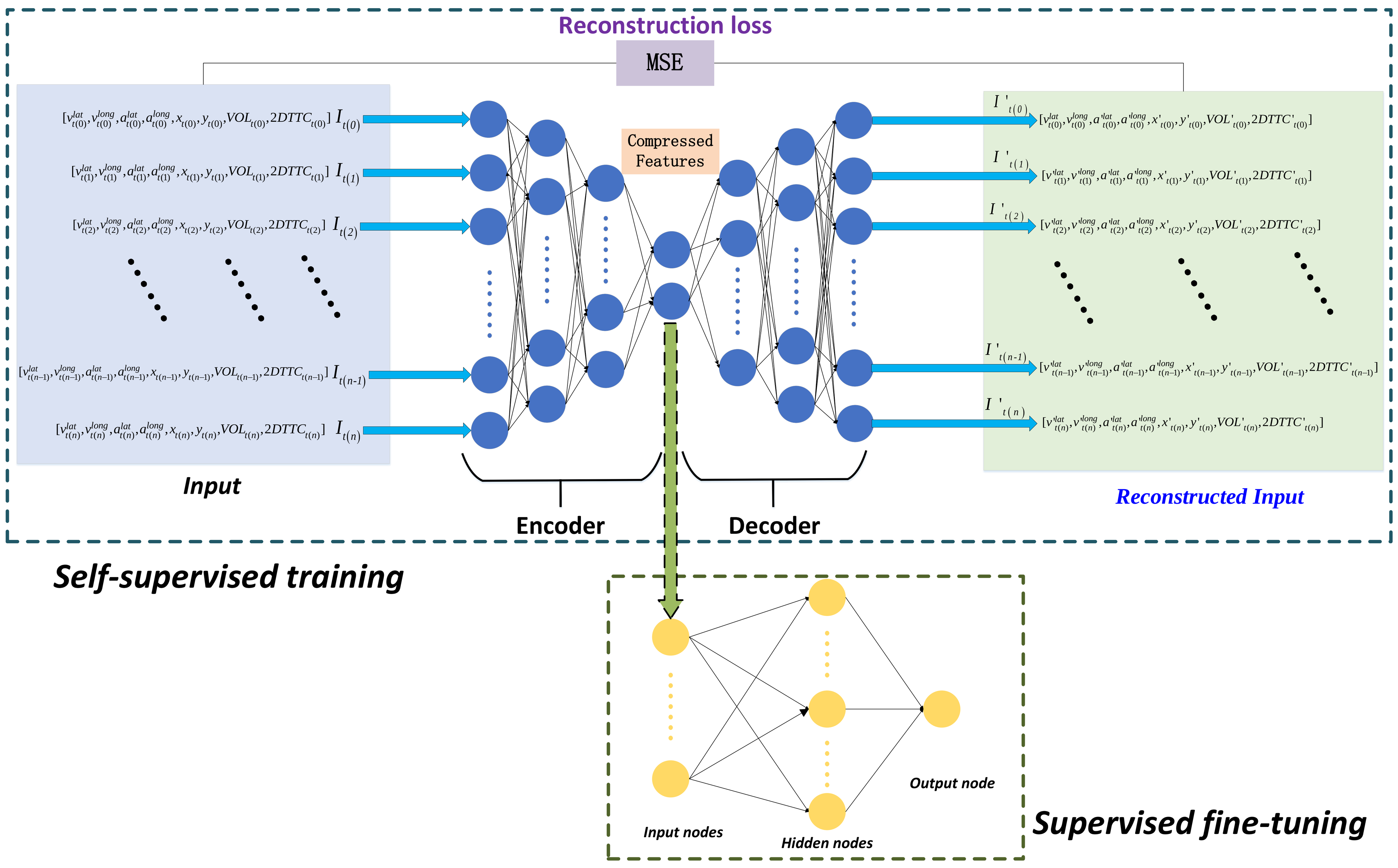

Figure 1. The Framework of HELM-based Semi-supervised Machine Learning Method.

Results

Features Setting	Features Input
1	coordinates/velocity/angle
2	coordinates/velocity/angle/acceleration/distance
3	coordinates/velocity/angle/2D time-to-collision

Model	Setting	Accuracy	Precision	Recall	F1-Score	FPR	TPR
Robust Covariance	1	0.3337	0.7628	0.1779	0.3735	0.1745	0.1779
	2	0.3348	0.7702	0.1767	0.3762	0.1663	0.1767
	3	0.9570	0.9487	0.9973	0.9028	0.1701	0.9973
Isolation Forest	1	0.5789	0.8766	0.5185	0.4680	0.2303	0.5185
	2	0.4387	0.8673	0.3080	0.4219	0.1487	0.3080
	3	0.9615	0.9517	1.0000	0.9131	0.1600	1.0000
HELM	1	0.9471	0.9349	1.0000	0.8766	0.2196	1.0000
	2	0.9614	0.9561	0.9949	0.9144	0.1440	0.9949
	3	**0.9958**	**0.9963**	0.9983	**0.9913**	**0.0118**	**0.9983**

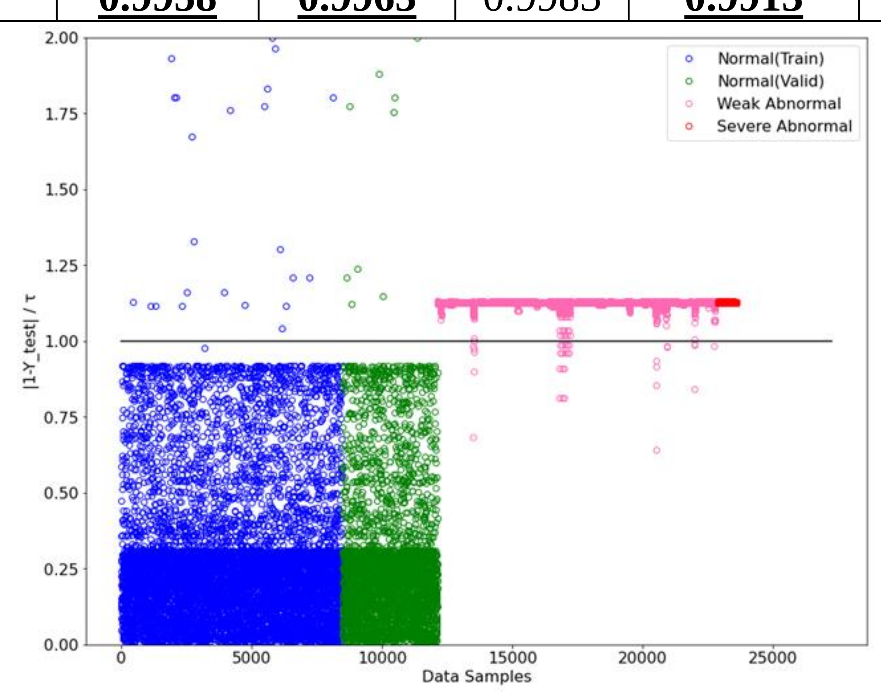

Figure 2. Scatter visualization of the result obtained by Semi-supervised HELM

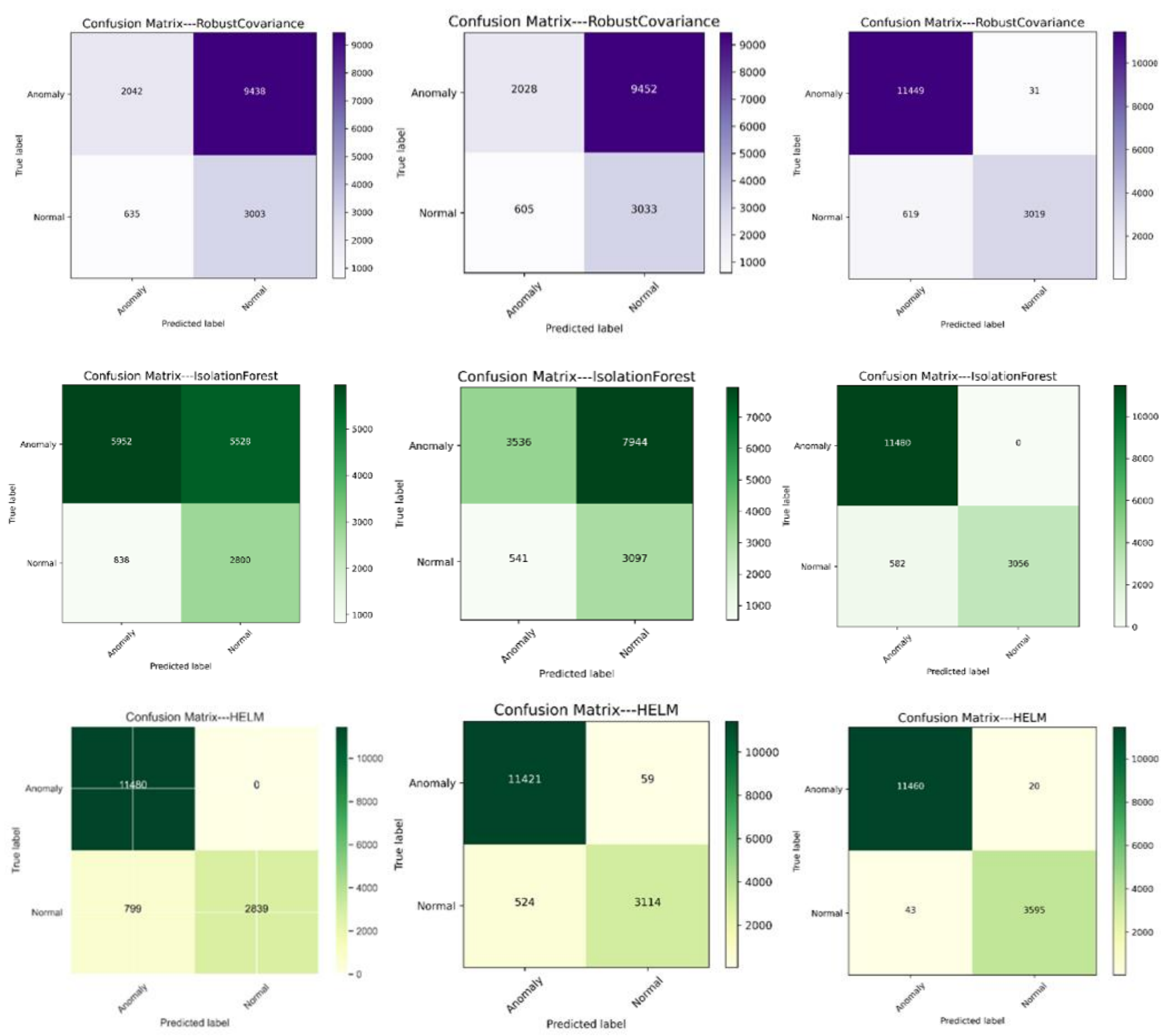

Figure 3. Visualization of models' performance under Setting 1, 2, and 3

Conclusions

- The proposed semi-supervised Hierarchical Extreme Learning Machines (HELM) model achieved the best performance among other models.
- The study emphasized the crucial role of SSMs, particularly 2D-TTC. The incorporation of the 2DTTC SSM significantly enhanced detection accuracy by over 5% compared to baseline settings.
- The semi-supervised approach cannot detect different kinds of anomalies (e.g., rapid/close lane-changing) which can be a future research direction.

TU Delft · NWO · SAMEN · Traffic and Transportation Safety Lab