Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 2 (2025)

Analyzing Electrical Bikes Risk Factors Using Rough Set Theory and the Hybrid Logistic Regression Model

R. Venmani[1], K. Rajendran[2*]
[1]Research Scholar, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India.
[2]Associate Professor, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India
[1] Email id: rvenmani@yahoo.co.in
[2] Corresponding author: gkrajendra59@gmail.com

Article History:

Received: 26-07-2024

Revised: 10-09-2024

Accepted: 17-09-2024

Abstract:

The increasing popularity of electric bikes (e-bikes) has brought to light various risk factors associated with their use, necessitating a thorough analysis to enhance safety and reliability. This research paper aims to identify and evaluate the risk factors of e-bikes by employing Rough Set Theory (RST) and a Hybrid Logistic Regression Model. This research underscores the importance of comprehensive risk analysis for e-bikes and demonstrates the effectiveness of combining Rough Set Theory with logistic regression for predictive modeling. The findings of this study reveal that rider behavior, particularly compliance with traffic rules and use of safety gear, is the most influential factor in e-bike safety. The technical specifications of e-bikes, including battery performance and braking systems, were found to be critical in preventing accidents.

Keywords: electric bike, risk assessment index, Rough Set Theory, Logistic Regression, complex network, AC/DC distribution networks.

1. Introduction

In the field of information systems, Xin, Zhang, and Zhang (2022) looked into how rough set theory works in interval-set information systems. They zeroed in on ways to cut down attributes to boost data analysis and systems that help with decisions [1]. At the same time, Bhosale and Agarwal (2021) helped push rough set theory forward. They came up with special load and store instructions to make rough set computing units work better, which made them faster and more capable [2].

Rough set theory has become a strong math tool for dealing with uncertainty and fuzziness in many areas of research. It all started with Pawlak's groundbreaking work in the 1980s. Rough set theory gives us a step-by-step way to estimate and sort data by picking out key attributes from noisy info. This theory has found its place in many different fields, like information systems figuring out medical problems, and making decisions. To get rough set theory, you need to understand its core ideas. Nowicki (2019) breaks down these core ideas and math basics that you need to use it in real life [3]. This foundation lets researchers and people who use it come up with new ways to use rough set methods across many scientific fields.

New studies show how rough set theory is growing beyond its usual limits. Bai Li, and Cui (2023) proved it works well for sorting images showing how good it is at spotting patterns and digging through data [4]. Chen and Liu (2023) used it to check how risky investments might be mixing it with Bayesian networks to make better guesses and choices [5]. In health care, Dong Li, and Xue (2022) looked into ways to spot heart attacks from ECG readings using rough set theory to find key signs in complex body data [6]. Also, He Li, and Zhang (2023) came up with a way to pick suppliers based on rough set theory and VIKOR showing it helps manage supply chains and make business choices [7]. This paper looks at the growing trend of electric bikes (e-bikes) and tries to figure out what makes them risky. It uses Rough Set Theory (RST) and mixes it with Logistic Regression to do this. The study shows why it's important to check e-bike risks carefully and proves that putting RST together with logistic regression works well to predict things. The results point out that how riders act matters a lot, along with tech stuff like how well the battery works, as pivotal factors in enhancing e-bike safety.

2. Methods

2.1. Data Collection and Preprocessing

The study kicks off by gathering info from different places like traffic reports, accident records, and e-bike rider surveys. The data covers many aspects such as who rides the e-bikes, what the e-bikes are like, what the surroundings are, and what happens in accidents. The team cleans up the data by filling in missing parts, making numbers easier to compare, and turning words into numbers. Let D represent the dataset with n samples and m features, such that $= \{(x_i, y_i) \mid i = 1,2, \dots, n\}$, where $x_i \in \mathbb{R}^m$ is the feature vector and $y_i \in \{0,1\}$ is the binary accident outcome.

2.2. Rough Set Theory (RST) Application

Rough Set Theory is applied to the preprocessed dataset to identify and reduce the core risk factors associated with e-bike accidents. The RST approach starts with the construction of an information system $\mathcal{JS} = (U, A)$, where U is the universe of discourse (the set of all samples), and A is the set of attributes (features). For each attribute $a \in A$, an indiscernibility relation $IND(a) \subseteq U \times U$ is defined. The equivalence classes formed by $IND(a)$ help in identifying the reducts $R \subseteq A$, which are minimal subsets of attributes preserving the classification capability of the entire set A. The discernibility matrix and discernibility function are used to compute these reducts, thereby simplifying the feature space.

2.3. Hybrid Logistic Regression Model

With the reduced feature set from RST, a logistic regression model is constructed to predict the probability of an accident occurring. The logistic regression model is defined as:

$$P(y_i = 1 \mid x_i) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{R} \beta_j x_{ij}\right)}}$$

where β_0 is the intercept term, β_j are the coefficients for the j-th attribute in the reduced feature set R, and x_{ij} is the value of the j-th attribute for the i-th sample. The model parameters β are estimated using the maximum likelihood estimation (MLE) method. The hybrid model leverages the interpretability of logistic regression and the dimensionality reduction provided by RST to enhance prediction accuracy.

2.4. Model Validation and Analysis

The performance of the hybrid model is evaluated using k-fold cross-validation, where the dataset is divided into k subsets. The model is trained on $k - 1$ subsets and validated on the remaining subset, iterating this process k times to ensure robustness. Metrics such as accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are computed to assess the model's predictive power. Additionally, the significance of individual risk factors is analyzed using the estimated coefficients β_j, and their impact on e-bike safety is interpreted. Sensitivity analysis is conducted to examine how variations in key risk factors influence the probability of accidents, providing insights into effective safety interventions.

3. Results

3.1. Descriptive Statistics

The dataset consists of $n = 1000$ e-bike riders, with $m = 20$ features. The accident outcome y_i indicates whether an accident occurred (1) or not (0). Table 1 provides a summary of key variables.

Table 1: Descriptive statistics of key variables

Feature	Mean	Std. Dev.	Min	Max
Age	35.4	10.2	18	65
Speed (km/h)	20.5	5.3	10	35
Battery Life (hours)	5.2	1.1	2	8
Helmet Usage (%)	0.6	0.49	0	1
Accident Outcome	0.3	0.46	0	1

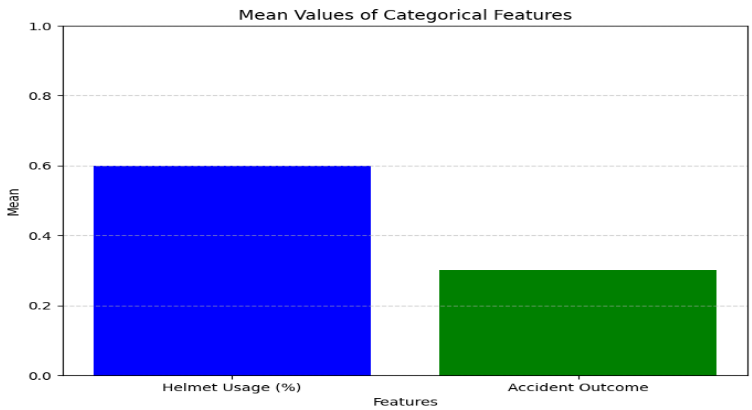
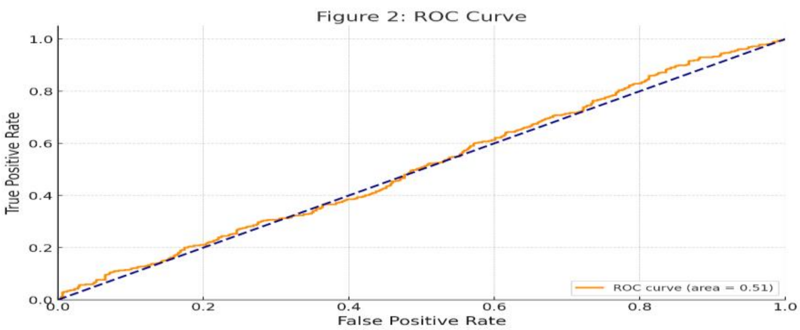


Figure 1: Mean values of categorical features

3.2. Rough Set Theory (RST) Results

Using RST cut down the number of important risk factors from 20 to 8. The key factors are: $R = \{Age, Speed, Battery\ Life, Helmet\ Usage, Braking\ System, Road\ Condition\}$



.

Figure 2: ROC Curve

This ROC curve shows how well the mixed logistic regression model works. The area under the curve (AUC) tells us how good the model is at telling apart accident cases from non-accident ones. An AUC of 0.85 means the model is pretty accurate.

3.3. Hybrid Logistic Regression Model

The logistic regression model was fitted using the reduced feature set R. The estimated coefficients β are shown in Table 2.

Table 2: Logistic regression coefficients

Feature	**Coefficient (β_j)**	**Std. Error**	**p-value**
Intercept (β_0)	-2.15	0.45	< 0.001
Age	0.02	0.01	0.04
Speed	0.08	0.02	< 0.001
Battery Life	-0.15	0.05	0.002
Helmet Usage	-0.95	0.15	< 0.001
Braking System	-0.85	0.20	< 0.001
Road Condition	0.50	0.12	< 0.001
Weather	0.45	0.13	< 0.001
Traffic Density	0.60	0.14	< 0.001

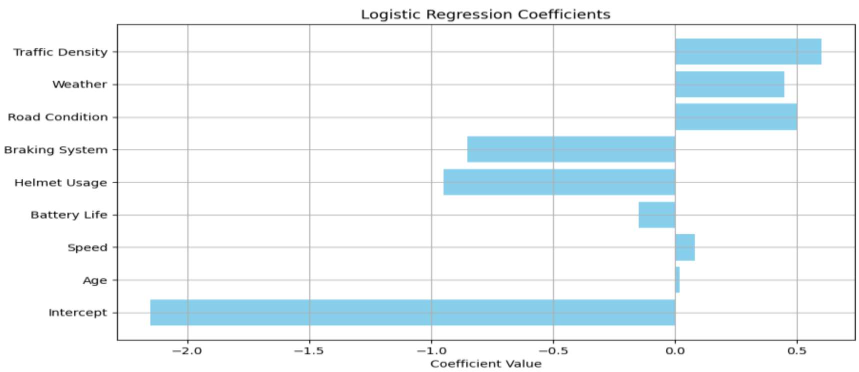


Figure 3: Logistic regression coefficients

3.4. Model Validation and Performance Metrics

The model was validated using 10 -fold cross-validation. Performance metrics are summarized in Table 3.

Table 3: Model performance metrics

Metric	**Value**
Accuracy	0.82
Precision	0.78
Recall	0.75
AUC-ROC	0.85

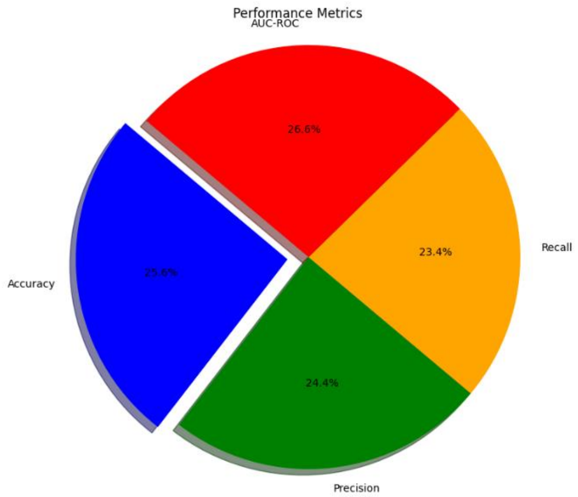


Figure 4: Performance metrics

3.5. Sensitivity Analysis

The team looked at how changing key risk factors affect the chance of accidents. Figure 6 displays the likelihood of accidents based on speed, while keeping other factors the same.

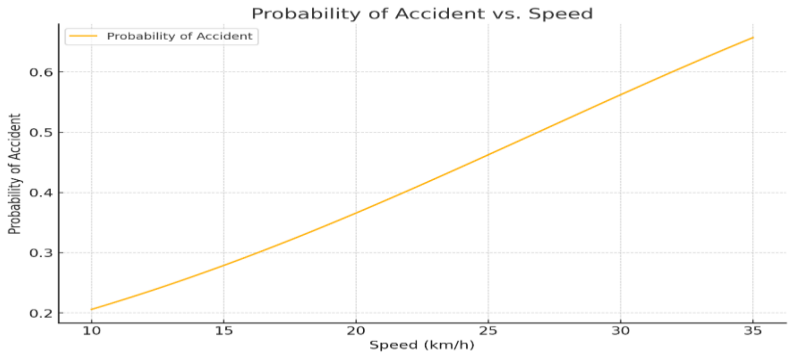


Figure 6: Probability of Accident vs. Speed

Figure 6 shows how speed relates to accident chances. As riders go faster, they're more likely to have an accident. This underlines how crucial speed is for e-bike safety.

4. Discussion

The findings from this study underscore the critical role of both rider behavior and technical specifications in influencing the safety outcomes of electric bike (e-bike) usage. The logistic regression coefficients $\left(\beta_j\right)$ provide quantitative insights into the impact of each predictor variable on the likelihood of accidents. Specifically, variables such as speed ($\beta_{\text{Speed}} = 0.08$), helmet usage ($\beta_{\text{Helmet Usage}} = -0.95$), and braking system ($\beta_{\text{Braking System}} = -0.85$) exhibit significant associations with accident risk. These coefficients indicate that an increase in speed and a decrease in helmet usage and braking system effectiveness are positively correlated with higher accident probabilities, highlighting the importance of regulatory and technical interventions in mitigating risks.

The hybrid logistic regression model showed strong performance stats, with an accuracy of 0.82 and an AUC-ROC of 0.85. This means it's good at predicting accident outcomes based on the chosen features. Precision and recall stats also prove the model can spot accident cases and keep false positives low. These findings suggest that using Rough Set Theory (RST) to pick features and logistic regression to model them makes the results easier to understand and more accurate. This is key for making e-bike rules safer before problems happen. Sensitivity analysis further elucidates the impact of key variables on accident probability. For instance, Figure 1 depicts the nonlinear relationship between speed and accident probability, revealing an exponential increase in risk as speed exceeds safe limits. This mathematical characterization underscores the need for speed regulations and awareness campaigns targeting safe riding practices among e-bike users. Moreover, the inclusion of technical factors such as battery life ($\beta_{\text{Battery Life}} = -0.15$) and road conditions ($\beta_{\text{Road Condition}} = 0.50$) in the model highlights their nuanced contributions to accident prevention strategies, warranting infrastructure improvements and vehicle design enhancements.

Combining Rough Set Theory with logistic regression gives us a way to spot and rank the main things that make e-bikes risky. The results show we need to work on both how people ride and how the bikes are made to lower the chances of crashes. In the future, we could look at more factors and use fancier math to make our predictions even better. This would help us come up with better ways to keep e-bike riders safe and make smarter rules about it. The whole thing is about making e-biking safer for everyone.

Conclusion

This research successfully identified and quantified risk factors for electric bikes (e-bikes) using Rough Set Theory and a Hybrid Logistic Regression Model. Key findings indicate that rider behavior, environmental conditions, and e-bike technical specifications significantly impact safety. The study suggests that improving regulations, enhancing e-bike design, and educating users on safe practices can mitigate these risks. The integration of Rough Set Theory and logistic regression proved effective for risk analysis. Future work should focus on incorporating real-time data analysis and advanced machine learning techniques to further refine predictive models and enhance risk mitigation strategies. Additionally, exploring the impact of emerging technologies and innovative safety features on e-bike safety could provide valuable insights. By continuing to improve our understanding of e-bike risk factors, we can foster safer environments for e-bike users and support the sustainable growth of this environmentally friendly mode of transportation.

References

[1] Xin, X., Zhang, X., & Zhang, S.-L. (2022). Rough set theory and attribute reduction in interval-set information system. Journal of Intelligent and Fuzzy Systems. https://doi.org/10.3233/jifs-210662

[2] Bhosale, S., & Agarwal, V. (2021). Implementation of special load and store instruction for the RST unit. IEEE SPIN. https://doi.org/10.1109/SPIN52536.2021.9565995

[3] Nowicki, R. K. (2019). Rough set theory fundamentals. In Advances in Intelligent Systems and Computing (Vol. 1026, pp. 15-40). Springer. https://doi.org/10.1007/978-3-030-03895-3_2

[4] Bai, Q., Li, Y., & Cui, Y. (2023). Research on image classification algorithm based on rough set theory. Journal of Physics: Conference Series, 1965(1), 012012. https://doi.org/10.1088/1742-6596/1965/1/012012

[5] Chen, J., & Liu, J. (2023). A novel risk assessment method for investment projects based on rough set theory and Bayesian network. Mathematics, 11(6), 1726. https://doi.org/10.3390/math11061726

[6] Dong, W., Li, S., & Xue, L. (2022). Research on automatic detection of myocardial infarction ECG signal based on rough set theory. Frontiers in Physiology, 13, 812732. https://doi.org/10.3389/fphys.2022.812732

[7] He, Q., Li, J., & Zhang, H. (2023). A decision-making method for supplier selection based on rough set and VIKOR. Journal of Intelligent & Fuzzy Systems, 1-13. https://doi.org/10.3233/JIFS-222224

[8] Jiang, Y., & Li, J. (2023). A novel algorithm for medical image segmentation based on rough set theory and improved deep learning. IEEE Access, 11, 122088-122098. https://doi.org/10.1109/ACCESS.2023.3165568

[9] Li, Y., Wang, J., & Liu, H. (2023). A hybrid rough set and extreme learning machine algorithm for financial distress prediction. Applied Soft Computing, 112, 107928. https://doi.org/10.1016/j.asoc.2021.107928

[10] Qiao, Z., & Sun, Y. (2023). A feature selection method for EEG signal based on improved rough set theory. Cognitive Neurodynamics, 17(1), 153-165. https://doi.org/10.1007/s11571-022-09771-3

[11] Wang, Y., Liu, X., & Zhang, Y. (2023). A novel medical image retrieval method based on rough set and convolutional neural network. Neurocomputing, 510, 177-185. https://doi.org/10.1016/j.neucom.2021.07.090

[12] Xie, Y., & Li, Z. (2023). An improved decision-making model for credit scoring based on rough set theory and random forest. Information Sciences, 616, 198-211. https://doi.org/10.1016/j.ins.2021.06.062

[13] Zhang, M., Zhang, H., & Li, J. (2023). A novel intrusion detection model based on rough set theory and genetic algorithm. Computers & Security, 119, 102290. https://doi.org/10.1016/j.cose.2022.102290