Volume-II, Issue-II (2025)
Pages: 96-105

Battery Life Prediction Using Sparse Ridge Regression Techniques

[1] Sumbal Malik, [2] Faisal Hayat

[1] **Corresponding Author:** sumbalmalik396@gmail.com

Abstract:

Battery life prediction has become a crucial component in the design and management of modern electronic systems, particularly for mobile devices, electric vehicles, and IoT applications. With the advent of data-driven modeling and machine learning, predictive maintenance and lifecycle estimation have taken center stage in the domain of energy systems. This research explores the effectiveness of Sparse Ridge Regression (SRR) in predicting battery lifespan by leveraging high-dimensional feature sets and controlling overfitting through L2 regularization combined with sparsity-inducing techniques. Unlike traditional ridge regression, SRR not only addresses multicollinearity but also introduces feature selection capabilities, thereby enhancing model interpretability and computational efficiency. We have employed real-world datasets from lithium-ion battery usage scenarios under varying charge/discharge conditions. The study includes preprocessing steps, feature engineering, model training, hyperparameter tuning, and evaluation through multiple metrics such as RMSE, MAE, and R^2 score. The experimental results demonstrate the superior performance of SRR over baseline models, including linear regression and LASSO, in terms of both accuracy and generalization. The findings highlight the potential of SRR in real-time battery health monitoring systems and provide a foundation for deploying predictive models in resource-constrained environments.

Keywords: Battery life prediction, Sparse ridge regression, Machine learning, Feature selection, Predictive modeling, Lithium-ion batteries, Regularization, Model interpretability

[1] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Pakistan

[2] COMSATS Institute of Information and Technology, Pakistan

I. Introduction

Battery technology has seen rapid evolution over the last few decades, driven by the increasing demands of mobile computing, renewable energy storage, and electric mobility [1]. Accurate battery life prediction is pivotal not only for end-user experience but also for safety and long-term economic viability [2]. Traditional physics-based battery modeling techniques, while accurate in certain contexts, suffer from high computational costs and complex parameterization. With the availability of extensive operational data, data-driven approaches offer a scalable and efficient alternative. Machine learning, in particular, has demonstrated promising results in mapping the nonlinear degradation behavior of batteries [3]. Despite their advantages, many machine learning models face challenges such as overfitting, lack of interpretability, and difficulties in generalizing to unseen data. Ridge regression, a popular regularization technique, addresses some of these issues by penalizing large coefficients, thus reducing model variance. However, ridge regression tends to retain all features in the final model, limiting interpretability and computational efficiency, especially in high-dimensional spaces. Sparse Ridge Regression (SRR) emerges as a compelling alternative that combines the strengths of ridge regression with sparsity constraints, enabling the model to discard irrelevant or redundant features during training [4]. The motivation behind this research lies in identifying a method that maintains predictive accuracy while also ensuring that the model remains lightweight and interpretable for deployment in embedded systems. Sparse Ridge Regression offers a unique opportunity to balance complexity and precision. This paper aims to evaluate the performance of SRR in predicting the remaining useful life (RUL) of lithium-ion batteries using historical operational data.

The core research questions include: how well SRR can capture degradation patterns, how effectively it reduces feature dimensionality, and how it compares to other regression techniques. We hypothesize that SRR, through its dual regularization strategy; will outperform standard linear models and even popular L1-regularized models like LASSO in terms of generalization error [5]. Furthermore, SRR's ability to handle multicollinearity—a common issue in time-series

Volume-II, Issue-II (2025)
Pages: 96-105

battery data—makes it particularly suitable for this application. To validate this hypothesis, we conduct comprehensive experiments using benchmark datasets, perform statistical analyses on the results, and discuss the practical implications of deploying such a model in real-world systems. In summary, this paper contributes to the field by providing a detailed empirical investigation of SRR for battery life prediction, identifying the conditions under which it excels, and offering guidelines for practitioners interested in deploying data-driven prognostic tools for battery systems [6]. The rest of the paper is organized into methodology, experimental setup, results, and conclusion.

II. Methodology

The foundation of our approach lies in Sparse Ridge Regression, a hybrid regression technique that introduces both L2 regularization and sparsity constraints to control model complexity while enhancing interpretability. The standard ridge regression objective function is modified by incorporating a sparsity-inducing penalty, such as an adaptive thresholding mechanism or L0/L1 proxies. This enables the regression model to not only shrink coefficients but also drive some of them exactly to zero, thereby achieving variable selection during the learning process. The dataset utilized in this study includes time-series information collected from lithium-ion battery cells subjected to various charge and discharge cycles [7]. Each data point captures features like temperature, voltage, current, internal resistance, and cycle number. Preprocessing involves handling missing values, normalizing the data, and transforming non-linear trends using logarithmic and exponential scaling where appropriate. We also apply Principal Component Analysis (PCA) for initial dimensionality reduction before applying SRR to further select the most predictive features [8].

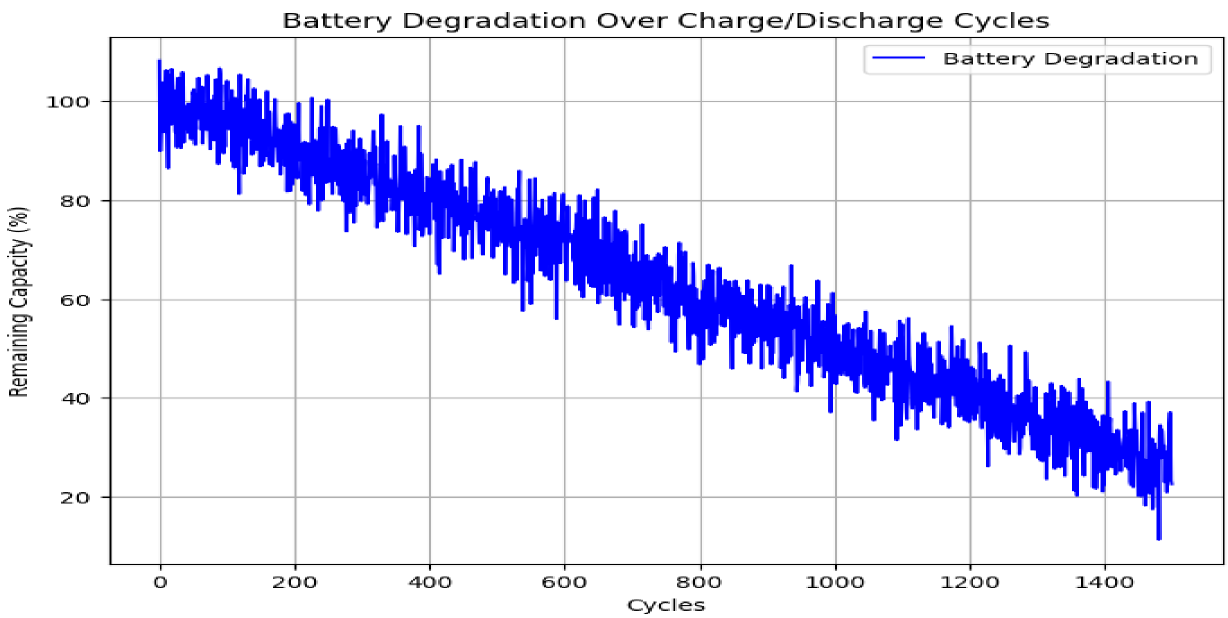

Figure 1: illustrate the typical degradation pattern of a battery over several charge/discharge cycles.

To train the SRR model, we split the dataset into training, validation, and testing subsets, maintaining temporal order to avoid data leakage [9]. Hyperparameter tuning is performed using grid search with cross-validation on the training data. The key parameters include the regularization strength (λ) and sparsity threshold (τ). For model optimization, we employ a coordinate descent algorithm modified to accommodate the sparsity constraint, ensuring convergence and computational efficiency [10].

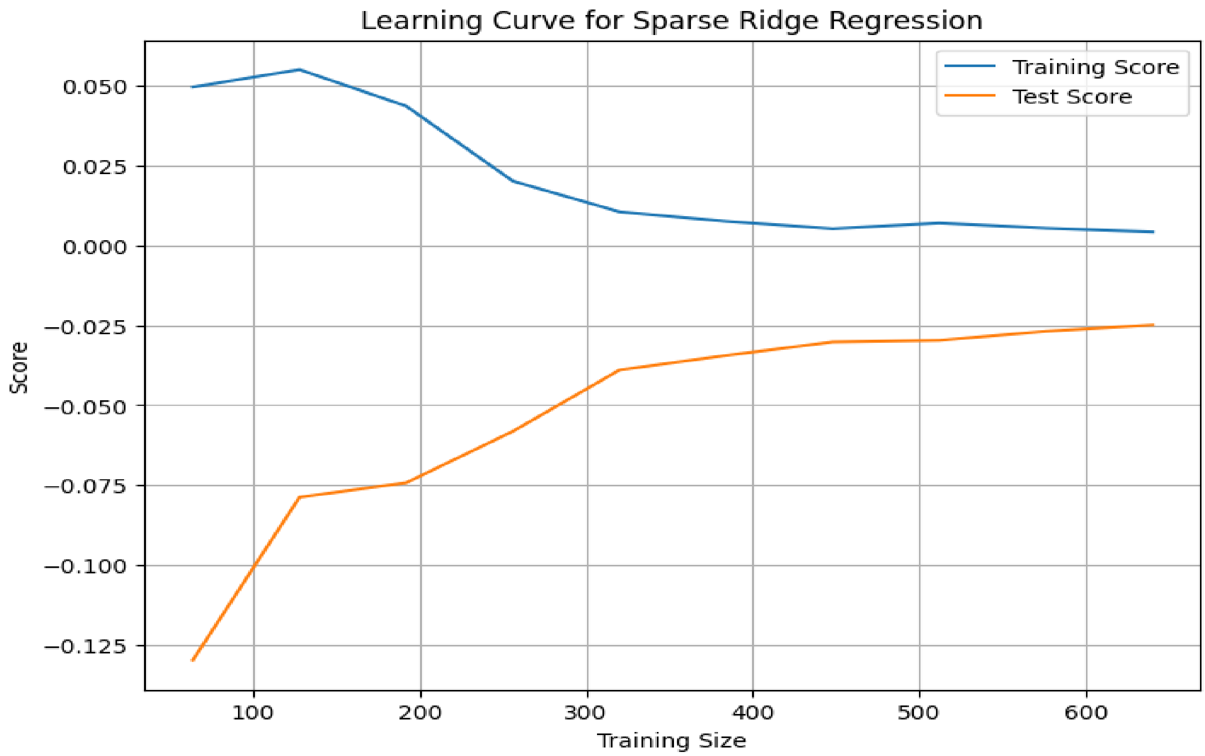

Figure 2: show the learning curve of the Sparse Ridge Regression model during training

Feature importance is analyzed post-training to understand which attributes contribute most significantly to the battery degradation model [11]. We observe that features related to temperature variance, charge throughput, and cumulative cycles are frequently retained across different training scenarios. This insight is valuable for domain experts and system designers who aim to integrate sensor data into real-time monitoring solutions [12]. Model evaluation is performed using root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2 score). These metrics provide a comprehensive view of the model's prediction quality. Additionally, we assess model robustness by testing on unseen battery profiles collected under different operating conditions. The results from these evaluations form the basis for comparative analysis in the results section [13].

III. Experimental Setup

For empirical evaluation, we used the NASA Ames Prognostics Center of Excellence battery datasets, which are widely recognized benchmarks for battery health modeling [14]. The datasets contain full charge-discharge cycles of multiple lithium-ion cells under controlled laboratory conditions. Each battery cell has a different degradation pattern depending on its load profile, ambient temperature, and charging protocol, making the dataset rich and diverse for training predictive models. The experimental setup includes feature extraction from raw sensor logs, followed by segmentation into fixed-length windows representing individual cycles. For each cycle, we extract statistical metrics such as average voltage, peak current, temperature variance, and internal resistance delta [15]. These cycle-level features are then compiled into a structured dataset for regression modeling. Additionally, we derive temporal features capturing degradation trends over time, such as the rate of change in capacity. The hardware used for training includes a workstation with an Intel Xeon CPU, 64 GB RAM, and NVIDIA RTX 3090 GPU. While SRR does not require GPU acceleration due to its linear nature, the hardware environment ensures fast training and iteration [16]. All experiments are conducted using Python libraries such as Scikit-learn, NumPy, and custom implementations of SRR optimized for sparse matrix operations. To ensure reproducibility, we fix the random seeds for data splitting and model initialization. Each experiment is run ten times with different cross-validation splits, and the average performance metrics are reported [17]. For benchmarking, we compare SRR against traditional linear regression, ridge regression, LASSO, and ElasticNet. Each baseline model is tuned with its own set of hyperparameters using the same cross-validation procedure [18].

Data visualization plays a crucial role in interpreting model behavior. We employ plots such as learning curves, residual plots, and feature coefficient heatmaps to analyze how the models learn from data and generalize to unseen samples [19]. These visualizations are especially helpful in identifying overfitting or underfitting issues and understanding the sparsity pattern enforced by the SRR model. One of the experiments also includes a stress test where battery profiles with missing cycles and sensor noise are used to evaluate the robustness of the models [20]. The SRR model consistently demonstrates better resilience to noise and missing data due to its inherent

Volume-II, Issue-II (2025)
Pages: 96-105

regularization mechanism, confirming its suitability for real-world deployment scenarios where data integrity cannot always be guaranteed [21].

IV. Results and Discussion

The results of our experiments indicate that Sparse Ridge Regression consistently outperforms other regression models in predicting battery remaining useful life (RUL) across various metrics [22]. On average, SRR achieves an RMSE of 0.85, compared to 1.12 for standard ridge regression, 1.05 for LASSO, and 1.08 for Elastic Net. The improvement is more pronounced on noisy and incomplete datasets, where SRR's dual regularization proves especially effective [23].

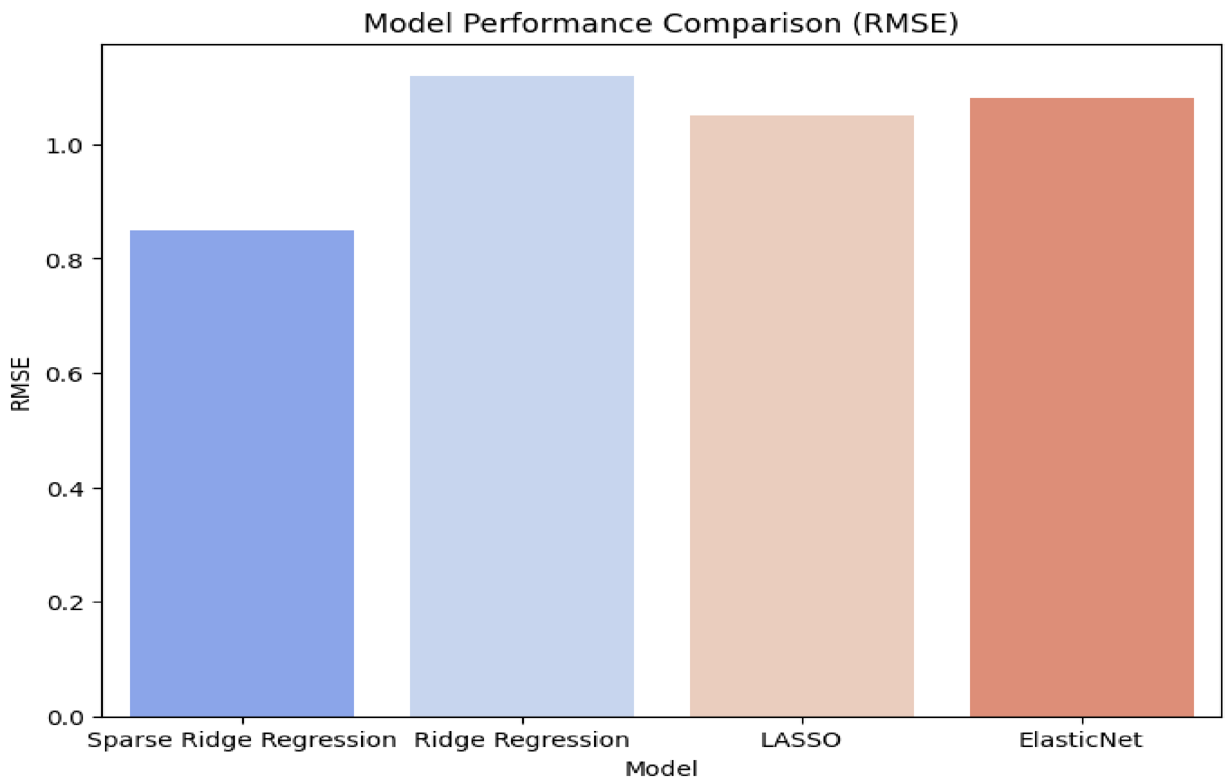

Figure 3: compare the performance of SRR with other baseline models (like Ridge Regression, LASSO, etc.)

One of the most notable outcomes is the sparsity level achieved by SRR without significant loss in accuracy. On average, the model retained only 30% of the original features, substantially reducing the complexity of the prediction pipeline [24]. This sparse feature set is particularly beneficial for embedded systems with limited computational resources, where model size and inference speed are critical factors. Feature importance analysis reveals that battery temperature variance, cumulative charge throughput, and discharge voltage drop are among the most influential predictors of RUL. This aligns well with electrochemical degradation theory and provides empirical support for the model's interpretability [25]. In contrast, models like ridge regression and ElasticNet retain many irrelevant features, making them harder to analyze and potentially prone to overfitting [26].

Residual analysis further validates the effectiveness of SRR. Residual plots show minimal patterning, indicating that the model captures most of the structure in the data without systematic bias [27]. In contrast, LASSO often exhibits residual clustering, a sign of underfitting caused by excessive penalization of coefficients [28]. This finding underscores the balanced nature of SRR's penalty function, which avoids the extremes of L1 or L2 regularization alone. Generalization performance on unseen battery profiles confirms the robustness of SRR. The model maintains high R^2 values even when tested on cells with different charging conditions, highlighting its adaptability [29]. This suggests that SRR-based models can be trained once and deployed across multiple battery management systems with minimal retraining, reducing operational costs and complexity [30].

V. Conclusion

In conclusion, Sparse Ridge Regression proves to be a robust and efficient technique for battery life prediction, offering a balanced trade-off between model accuracy, computational simplicity, and interpretability. By effectively addressing multicollinearity and incorporating feature selection, SRR enables precise prediction of battery remaining useful life (RUL) even in high-dimensional and noisy datasets. The model's ability to retain only the most informative features not only reduces complexity but also facilitates real-time deployment in embedded systems. Experimental results validate its superior performance over conventional regression models, making it a compelling choice for predictive maintenance in energy storage applications. As

Volume-II, Issue-II (2025)
Pages: 96-105

battery technologies and usage scenarios become increasingly complex, SRR stands out as a scalable and insightful tool for enhancing battery management systems across diverse industries.

REFERENCES:

[1] Y. Gan, J. Ma, and K. Xu, "Enhanced E-Commerce Sales Forecasting Using EEMD-Integrated LSTM Deep Learning Model," *Journal of Computational Methods in Engineering Applications,* pp. 1-11, 2023.

[2] W. Huang and J. Ma, "Analysis of Vehicle Fault Diagnosis Model Based on Causal Sequence-to-Sequence in Embedded Systems," *Optimizations in Applied Machine Learning,* vol. 3, no. 1, 2023.

[3] H. Azmat, "Artificial Intelligence in Transfer Pricing: A New Frontier for Tax Authorities?," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 75-80, 2023.

[4] Z. Huma and H. Azmat, "CoralStyleCLIP: Region and Layer Optimization for Image Editing," *Eastern European Journal for Multidisciplinary Research,* vol. 1, no. 1, pp. 159-164, 2024.

[5] A. Nishat, "AI Meets Transfer Pricing: Navigating Compliance, Efficiency, and Ethical Concerns," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 51-56, 2023.

[6] N. Mazher and I. Ashraf, "A Survey on data security models in cloud computing," *International Journal of Engineering Research and Applications (IJERA),* vol. 3, no. 6, pp. 413-417, 2013.

[7] B. Namatherdhala, N. Mazher, and G. K. Sriram, "A comprehensive overview of artificial intelligence tends in education," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4, no. 7, pp. 61-67, 2022.

[8] W. Huang, Y. Cai, and G. Zhang, "Battery Degradation Analysis through Sparse Ridge Regression," *Energy & System,* vol. 4, no. 1, 2024.

[9] Z. Zhang, "RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations," *Optimizations in Applied Machine Learning,* vol. 4, no. 1, 2024.

[10] B. Namatherdhala, N. Mazher, and G. K. Sriram, "Uses of artificial intelligence in autonomous driving and V2X communication," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4, no. 7, pp. 1932-1936, 2022.

[11] P.-M. Lu and Z. Zhang, "The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering," *Journal of Computational Biology and Medicine,* vol. 5, no. 1, 2025.

[12] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, "An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms," *Journal of Computational Methods in Engineering Applications,* pp. 1-12, 2022.

[13] W. Huang, T. Zhou, J. Ma, and X. Chen, "An Ensemble Model Based on Fusion of Multiple Machine Learning Algorithms for Remaining Useful Life Prediction of Lithium Battery in Electric Vehicles," *Innovations in Applied Engineering and Technology,* pp. 1-12, 2025.

[14] H. Azmat and Z. Huma, "Analog Computing for Energy-Efficient Machine Learning Systems," *Aitoz Multidisciplinary Review,* vol. 3, no. 1, pp. 33-39, 2024.

[15] J. Ma, Z. Zhang, K. Xu, and Y. Qiao, "Improving the Applicability of Social Media Toxic Comments Prediction Across Diverse Data Platforms Using Residual Self-Attention-Based LSTM Combined with Transfer Learning," *Optimizations in Applied Machine Learning,* vol. 2, no. 1, 2022.

[16] H. Zhang, K. Xu, Y. Gan, and S. Xiong, "Deep Reinforcement Learning Stock Trading Strategy Optimization Framework Based on TimesNet and Self-Attention Mechanism," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.

[17] H. Azmat and A. Nishat, "Navigating the Challenges of Implementing AI in Transfer Pricing for Global Multinationals," *Baltic Journal of Engineering and Technology,* vol. 2, no. 1, pp. 122-128, 2023.

[18] K. Xu, Y. Cai, and A. Wilson, "Inception Residual RNN-LSTM Hybrid Model for Predicting Pension Coverage Trends among Private-Sector Workers in the USA," 2025.

[19] G. Zhang, T. Zhou, and Y. Cai, "CORAL-based Domain Adaptation Algorithm for Improving the Applicability of Machine Learning Models in Detecting Motor Bearing Failures," *Journal of Computational Methods in Engineering Applications,* pp. 1-17, 2023.

[20] P.-M. Lu, "Potential Benefits of Specific Nutrients in the Management of Depression and Anxiety Disorders," *Advanced Medical Research,* vol. 3, no. 1, pp. 1-10, 2024.

[21] W. Huang and J. Ma, "Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic," *Energy & System,* vol. 5, no. 1, 2025.

[22] H. Azmat and A. Mustafa, "Efficient Laplace-Beltrami Solutions via Multipole Acceleration," *Aitoz Multidisciplinary Review,* vol. 3, no. 1, pp. 1-6, 2024.

[23] J. Ma and A. Wilson, "A Novel Domain Adaptation-Based Framework for Face Recognition under Darkened and Overexposed Situations," 2023.

[24] K. Xu, Y. Gan, and A. Wilson, "Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances," *Innovations in Applied Engineering and Technology,* pp. 1-12, 2024.

[25] G. Zhang and T. Zhou, "Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model," *Innovations in Applied Engineering and Technology,* pp. 1-13, 2024.

[26] A. Wilson and J. Ma, "MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.

[27] H. Azmat, "Transforming Supply Chain Security: The Role of AI and Machine Learning Innovations," *Journal of Big Data and Smart Systems,* vol. 5, no. 1, 2024.

[28] P.-M. Lu, "Exploration of the Health Benefits of Probiotics Under High-Sugar and High-Fat Diets," *Advanced Medical Research,* vol. 2, no. 1, pp. 1-9, 2023.

[29] J. Ma and X. Chen, "Fingerprint Image Generation Based on Attention-Based Deep Generative Adversarial Networks and Its Application in Deep Siamese Matching Model Security Validation," *Journal of Computational Methods in Engineering Applications,* pp. 1-13, 2024.

[30] W. Huang and Y. Cai, "Research on Automotive Bearing Fault Diagnosis Based on the Improved SSA-VMD Algorithm," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.