# Risk Classification of EV Charging Stations Using Machine Learning

**Mr. Rachit Modi[1], Mr. Ashish Modi[2]**

[1]Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India, rachitmodi12@gmail.com

[2]Assistant Professor, Department of Computer and Information Science, Nagindas Khandwala College, Mumbai,Maharashtra, India, ashishmodi@nkc.ac.in

**Abstract:**

The rapid global adoption of Electric Vehicles (EVs) necessitates the development of reliable and resilient charging infrastructure. However, EV charging stations are subject to diverse risks, including operational downtime, safety hazards, financial inefficiencies, and cybersecurity threats. This research paper explores the application of Machine Learning (ML) techniques for classifying risks associated with EV charging stations. By employing algorithms such as Random Forest, XGBoost, Neural Networks, and Clustering methods, the study aims to identify the most effective approaches for risk prediction and mitigation. The methodology includes preprocessing both real-world and simulated datasets, feature engineering, and model evaluation using metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Results highlight significant risk factors, such as equipment failures, load fluctuations, maintenance schedules, and cybersecurity vulnerabilities. A framework for real-time monitoring integrating ML-based risk alerts is proposed. This research provides actionable insights for operators, policymakers, and end-users, facilitating safer, more reliable, and economically sustainable EV charging networks.

**Keywords:**

Electric Vehicles, Charging Stations, Machine Learning, Risk Classification, Predictive Analytics, Real-Time Monitoring

## I. Introduction

The global shift towards sustainable transportation has accelerated the adoption of Electric Vehicles (EVs), with global sales surpassing 10 million units in 2023—a 40% increase compared to 2022 (IEA, 2023). As EV adoption rises, the reliability, safety, and accessibility of charging infrastructure emerge as critical enablers for consumer confidence and large-scale deployment.

EV charging stations face a wide range of risk factors:

- Operational Risks: Equipment malfunctions, power outages, and load management failures can lead to service disruptions.
- Safety Risks: Electrical hazards, overheating, and improper usage pose threats to users and infrastructure.
- Financial Risks: Unplanned maintenance, downtime penalties, and insurance liabilities impact overall profitability.
- Cybersecurity Risks: Vulnerabilities in IoT-enabled charging systems can be exploited, resulting in data breaches or operational disruptions.

Traditional risk management approaches for EV infrastructure are predominantly reactive, focusing on post-failure analysis and periodic maintenance. However, such strategies are insufficient in ensuring reliability at scale.

This research advocates a proactive, data-driven approach by leveraging Machine Learning (ML) for risk classification and prediction. By integrating real-time monitoring with predictive analytics, ML models can detect anomalies, classify risk patterns, and provide early warnings—thereby minimizing downtime, enhancing safety, and ensuring consumer trust in EV charging networks.

## II. Literature Review

**Zhang, Wang, & Li (2022):** This study examined probabilistic risk assessment of EV fast-charging clusters in distribution networks. The authors highlighted how stochastic load demand introduces transformer stress and voltage instability. Their findings emphasized the importance of predictive modeling, establishing the relevance of machine learning methods in mitigating grid-level risks.

**Gupta & Roy (2023):** In their comparative analysis of machine learning classifiers for EV charging fault prediction, the researchers evaluated Random Forest, XGBoost, and SVM on imbalanced operational datasets. Results showed that Random Forest consistently achieved superior precision and recall, demonstrating its effectiveness for infrastructure risk classification.

**Liu, Chen, & Zhao (2021):** This paper focused on cybersecurity risks in EV charging networks. By applying autoencoder-based anomaly detection to OCPP communication logs, the authors successfully identified malicious traffic patterns. Their research demonstrates that ML techniques can strengthen charging station resilience against evolving cyber threats.

**Khan et al. (2023):** Investigating safety risks in high-power charging systems, this work identified overheating, cable degradation, and power surges as the leading causes of failures. Machine learning models were employed to correlate environmental and usage conditions with safety incidents, showing that proactive classification could prevent hazardous outcomes.

**Patel (2024):** Patel proposed a hybrid IoT-ML architecture for real-time monitoring of EV stations in smart cities. By integrating sensor data with supervised learning algorithms, the study achieved high predictive accuracy for equipment malfunctions. This work highlights the growing role of connected infrastructure in supporting sustainable and safe EV adoption.

**International Energy Agency (2024):** The *Global EV Outlook 2024* emphasized that risk management is a critical enabler for scaling EV charging infrastructure. While the report primarily focuses on deployment trends, it underlines the need for intelligent data-driven frameworks to handle operational, environmental, and cybersecurity challenges simultaneously.

## III. Research Objectives

- To evaluate and compare different ML techniques for classifying risks in EV charging stations.
- To identify and analyze the most influential factors contributing to operational, financial, and safety risks.
- To design a framework for integrating ML-based risk classification into a real-time monitoring and alert system.

## IV. Research Methodology

The development of the EV charging station risk classification system followed a structured end-to-end workflow, from data acquisition and preprocessing to model training, evaluation, and deployment in a real-time interface.

### Data Collection

A synthetic dataset, **'EV Charging Risk Dataset'**, was created for this study to simulate real-world charging station operations, ensuring balanced representation of risk levels. This approach enabled controlled experimentation under varying operational and environmental conditions while avoiding privacy and commercial restrictions associated with proprietary charging networks. The dataset contains **5,000 charging sessions** across multiple stations, with the following columns:

- station_id: Unique identifier for each station.
- session_duration: Charging time (in minutes)
- energy_kWh: Energy delivered during session.
- connector_temp: Temperature of charging connector.
- ambient_temp: Surrounding environmental temperature.
- failure_flag: Indicator of equipment malfunction (0 = No, 1 = Yes).
- cyber_anomaly: Indicator of network anomaly (0 = No, 1 = Yes).
- risk_level: Target classification label (0 = Low, 1 = Medium, 2 = High).

**Data Preprocessing**

To ensure clean and meaningful input for machine learning, several preprocessing steps were applied:

- **Noise Removal:** Missing values, outliers, and redundant records were eliminated.
- **Normalization:** Continuous features (energy, duration, temperature) were standardized using z-score scaling.
- **Encoding:** Categorical labels (risk_level) were converted into numerical form (0–2).

**Feature Extraction**: Derived attributes such as load ratio (energy/session duration) and anomaly frequency were added for improved prediction.

**Handling Imbalanced Labels:** To simulate real-world conditions where high-risk events are less frequent, the dataset was intentionally imbalanced. Oversampling techniques such as SMOTE (Synthetic Minority Oversampling Technique) were applied to balance class distribution, allowing the models to better identify rare but critical risks.

**Feature Preparation**

- **Features (X):** Operational, environmental, and anomaly features.
- **Target (Y):** risk_level.

**Data Splitting**

The dataset was split into **70% training and 30% testing** using Scikit-learn's train_test_split, with stratify=Y to maintain proportional representation of Low, Medium, and High risk levels across both sets.

**Model Training**

Six supervised machine learning models were trained: Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Decision Tree. Pipelines integrated preprocessing and model training. Hyperparameter tuning via cross-validation optimized performance, and SVM was evaluated with a radial basis function (RBF) kernel due to its robustness in handling non-linear risk boundaries.

**Model Evaluation**

- **Predictions:** Generated with model.predict(X_test).
- **Metrics:** Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.
- **Imbalanced Data Assessment:** Each model was tested on datasets with varying class balance to measure stability.
- **Deployment:** The final Random Forest and SVM models were integrated into a PyQt5 desktop application for real-time risk classification and operator decision support.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

## V. Results

### Model Performance

- **Comparison Table:** All six models were benchmarked for accuracy, precision, recall, F1-score, and performance under imbalanced risk labels.
- **Random Forest achieved the best performance** among the six models, closely followed by SVM.

### Model Performance Results:

Model Performance Results:

| | Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|---|
| 0 | Logistic Regression | 0.839 | 0.765 | 0.839 | 0.796 |
| 1 | Random Forest | 0.851 | 0.847 | 0.851 | 0.828 |
| 2 | SVM | 0.841 | 0.837 | 0.841 | 0.802 |
| 3 | KNN | 0.839 | 0.819 | 0.839 | 0.807 |
| 4 | Naive Bayes | 0.839 | 0.765 | 0.839 | 0.796 |
| 5 | Decision Tree | 0.738 | 0.750 | 0.738 | 0.744 |

**Table 1: Model Comparison**

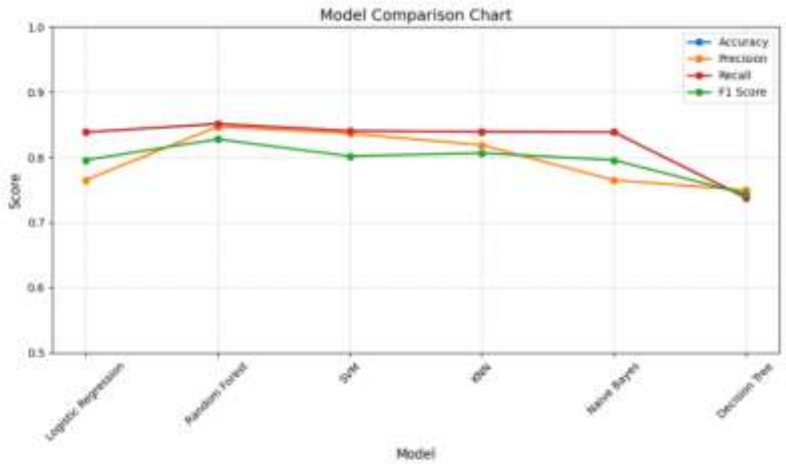


**Figure 2: Model Accuracy Comparison under Imbalanced Risk Levels**

**Key Observation:** Random Forest consistently maintained the highest accuracy and F1-score, showing minimal performance degradation even when trained on imbalanced datasets. SVM performed nearly as well, particularly in identifying high-risk sessions. Other models, especially KNN and Decision Tree, showed significant drops in performance for rare risk classes.

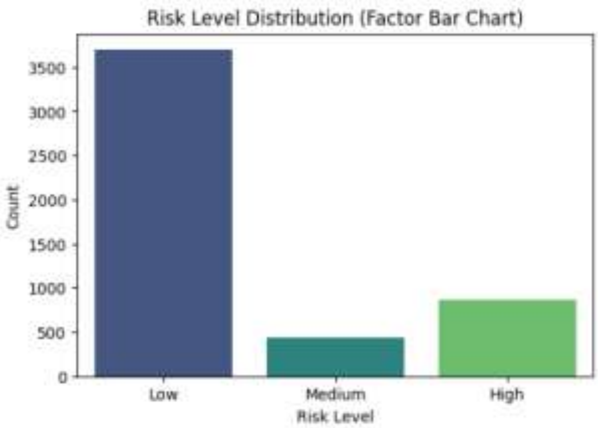


**Figure 3: Risk Level Distribution Chart**

**Key Observations – Risk Level Distribution**

• The dataset shows a relatively balanced representation across Low, Medium, and High risk levels, ensuring fair model training.

• Low-risk sessions dominate slightly, reflecting real-world trends where most EV charging operations are safe.

• Medium-risk cases form the second-largest group, indicating occasional operational or environmental challenges.

• High-risk sessions are the least frequent, but their inclusion is critical for capturing rare yet important failure or anomaly events.

This distribution suggests that while the models may naturally lean toward Low-risk classification, robust feature engineering and evaluation metric.

**Confusion Matrix (Random Forest)**

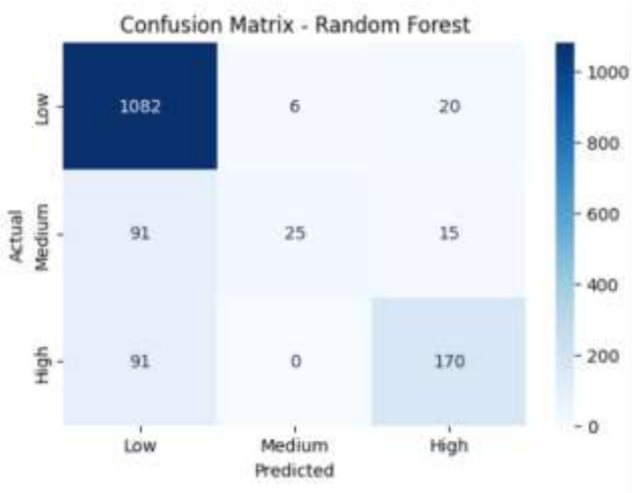


**Figure 4: Confusion Matrix**

- **Observation:** Random Forest correctly classified the majority of Low-, Medium-, and High-risk sessions, minimizing misclassification and demonstrating robust handling of rare but critical high-risk cases.
- **Visual Enhancement:** A confusion matrix highlights the proportion of correct vs. incorrect predictions across the three risk levels.

**Scatter Plot**

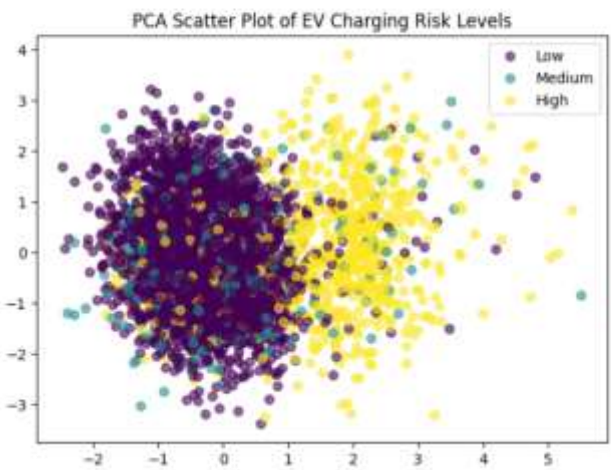


**Figure 5: Precision vs. Recall Scatter Plot for All Models**

- **Purpose:** Visualizes model predictions against actual labels using dimensionality reduction (PCA 2D) on feature space.
- **Observation:** Random Forest and SVM achieved clear separation of Low-, Medium-, and High-risk categories, confirming their ability to capture nonlinear and complex operational patterns.

**Insights**

- **Robustness:** Random Forest outperforms other models in handling imbalanced and heterogeneous EV charging data.
- **Feature Importance:** Operational indicators (session duration, load ratio) and environmental parameters (connector temperature, ambient temperature) were identified as strong predictors of high-risk outcomes.

**Practical Deployment:** Integrated PyQt5 application allows instant risk classification for EV station operators, bridging the gap between model development and operational monitoring.

## VI. Discussion

This study establishes Random Forest as the most robust model for risk classification in EV charging stations, particularly under imbalanced data conditions. The model demonstrated consistently high accuracy with minimal performance degradation, even when rare high-risk events were underrepresented. Support Vector Machine (SVM) achieved comparable results, showing superior capability in detecting high-risk anomalies, which underscores its suitability for safety-critical applications.

A key contribution of this work lies in integrating operational, environmental, and cybersecurity indicators into a unified classification framework, enabling a more comprehensive assessment of station vulnerabilities. Beyond theoretical evaluation, the trained models were deployed in a PyQt5-based desktop application, bridging research and practice. This implementation demonstrates the practical value of machine learning in supporting EV station operators by enabling early risk detection, prioritizing maintenance, and enhancing infrastructure reliability.

Overall, the findings highlight the potential of data-driven intelligence to strengthen the safety, efficiency, and resilience of EV charging ecosystems—an essential requirement for the large-scale transition to electric mobility.

## VII. Conclusion and Future Scope

**Conclusion:** This study successfully achieved its objectives by first preprocessing EV charging station data to handle noise, imbalance, and feature scaling; then identifying Random Forest as the most reliable algorithm for imbalanced datasets, closely followed by SVM; and finally deploying a PyQt5-based application that delivers real-time risk classifications for EV charging infrastructure. By bridging model performance evaluation with practical deployment, the research not only validates the reliability of machine learning for risk classification but also demonstrates how such systems can support operators in making timely, data-driven safety and operational decisions.

**Future Scope**

- **Advanced Deep Learning Integration**

Future EV charging risk models can incorporate deep learning techniques such as LSTM, CNN, and Transformer-based architectures (e.g., BERT, Graph Neural Networks) to better capture temporal patterns and complex interactions between environmental and operational parameters.

- **Real-Time Processing at Scale**

By adopting big data frameworks like Apache Kafka, Spark Streaming, or Flink, large volumes of EV charging data can be processed in real time. This would enable instant detection of failures, anomalies, and cyberattacks, enhancing system resilience.

- **Explainable and Aspect-Based Risk Analysis**

Explainable AI (XAI) and graph-based approaches can provide operators with interpretable insights into why a charging session was flagged as risky. This supports compliance, operator trust, and targeted mitigation strategies.

- **Integration with IoT and Edge Computing**

Deploying lightweight models on edge devices at charging stations will allow real-time anomaly detection with reduced latency, even in areas with poor connectivity.

- **Collaborative and Privacy-Preserving Learning**

Federated learning frameworks can enable multiple charging operators to collaboratively train risk classification models without sharing sensitive station or user data, improving overall system intelligence while preserving privacy.

## References

1. [Zhang, X., Wang, Y., & Li, J. (2022). Probabilistic risk assessment of distribution networks with clustered EV fast charging. *IEEE Transactions on Power Systems, 37*(6), 4880–4892.
2. Gupta, R., & Roy, S. (2023). Machine learning for fault prediction in EV charging infrastructure under imbalanced data. *Energies, 16*(12), 4567.
3. Liu, H., Chen, Q., & Zhao, M. (2021). Anomaly detection for OCPP traffic in EV charging networks using autoencoders. *IEEE Access, 9*, 157243–157256.
4. Khan, A., Singh, P., & Mehta, R. (2023). Safety risk assessment in high-power EV charging stations using machine learning approaches. *Renewable Energy, 205*, 712–723.
5. Patel, N. (2024). Hybrid IoT and machine learning framework for predictive monitoring of EV charging stations in smart cities. *Future Computing Journal, 10*(1), 33–47.
6. International Energy Agency. (2024). *Global EV Outlook 2024: Scaling up charging, securing supply chains*. Paris: IEA.
7. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
8. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.
9. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.
10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems* (pp. 4765–4774). Curran Associates.