ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

AI-Based Predictive Maintenance for EV Chargers Using Real-World Operational and Fault-History Data

Matthew Busayo Olanrewaju[1]; Cornelius Chukwujekwu Okafor[2]; Emmanuel Ayodeji Afolabi[3]

[1]Engr Bsc, Msc, MIET, R. COREN
[2]Engr Bsc, Msc
[3]Engr Bsc, Msc, MIET

Publication Date: 2026/09/10

Abstract: Reliable electric vehicle (EV) charging infrastructure is increasingly important as charging networks expand. The study developed an explainable machine-learning framework to determine if a charger would document a maintenance-relevant fault in the next three calendar days. A Chinese dataset of 441,077 charging sessions from 92 chargers was transformed into a continuous charger-day panel with 16 operational, environmental, fault-history and rolling-workload predictors. The performance of Logistic Regression, Random Forest, XGBoost and LightGBM was compared via leakage-controlled temporal validation, operating-threshold selection on the independent validation period, and final evaluation on a later temporal holdout. Logistic Regression achieved the highest cross-validated Average Precision (0.701) and ROC-AUC (0.817). It had ROC-AUC 0.806 and Average Precision 0.750 on the temporal holdout comprised of 15,026 rows. For the validation selected threshold 0.16, the recall value was 0.875 and the F2 value was 0.762. Ablation and correlation-aware SHAP analysis highlighted the importance of recent fault history for short-horizon risk, explaining 67.77% of the total mean absolute SHAP attribution. There was a significant domain shift observed in the UK and Swedish datasets; however, the fault labels were not compatible and no direct external validation was possible. The framework offers a rigorous and interpretable basis for prioritisation of maintenance and emphasizes the necessity for the local validation and setting of thresholds dependent on costs before proceeding to new chargers and networks.

Keywords: *Electric Vehicle Charging Infrastructure; Predictive Maintenance; Fault Prediction; Explainable Artificial Intelligence; Logistic Regression; Temporal Validation.*

How to Cite: Matthew Busayo Olanrewaju; Cornelius Chukwujekwu Okafor; Emmanuel Ayodeji Afolabi (2026) AI-Based Predictive Maintenance for EV Chargers Using Real-World Operational and Fault-History Data. *International Journal of Innovative Science and Research Technology*, 11(9), 50-65. https://doi.org/10.38124/ijisrt/26sep157

I. INTRODUCTION

The increasing adoption of electric mobility is putting a demand on charging infrastructure as a key component of contemporary transport and energy systems. High charging reliability is essential not only to facilitate the rise of EV use, but also to guarantee user confidence and network availability. The charging needs vary from place to place as they are influenced by travel behaviour, access to charging points, parking facilities and the local demand dynamics (Funke et al., 2019). The deployment of EVs also has a global boost, with focus on charging stations and service reliability (International Energy Agency, 2023). The greater the usage of public and semi-public chargers, the more of a operational issue maintenance becomes as opposed to a secondary engineering issue.

EV chargers are power-electronic and electro-mechanical equipment subject to repeated switching, variation in energy throughput, environmental variations and disparate user behaviour. These conditions produce transaction and operating records that can be used for condition monitoring. Thermal and operational data have already been applied to detect abnormal behaviour in charging stations (Izquierdo Gomez et al., 2022), and autoencoder-based techniques have shown good performance in early fault detection when applied to EV supply equipment (Sakwa et al., 2024). This provides a chance to shift from reactive maintenance to methods which rely on past operating data to predict failures before they disrupt services.

The wide basis for this transition is given by predictive maintenance. Instead of using static schedules or engaging post-event intervention, it predicts the risk associated with an equipment for a given time in the future based on current and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

historical asset data (Carvalho et al., 2019; Zonta et al., 2020). The reason for machine learning being helpful in this setting is that fault occurrence may closely depend on complex relationships between input utilisation, the environment and previous behaviour of the equipment (Dalzochio et al., 2020). In the area of machinery prognostics, it is shown that all temporal information will be critical for future-state prediction since previous operating states may contain information about upcoming asset condition (Lei et al., 2018; Zhao et al., 2019). As such, transaction histories can to temporal indicators of recent workload at EV chargers and degree of fault recurrence.

Several machine learning models are already developed and optimised in EV charging analytics, especially for the charging characteristic, demand prediction, load forecasting and flexible charging (Shahriar et al., 2020; Almaghrebi et al., 2020; Sun et al., 2020; Wang et al., 2020; Sadeghian et al., 2021). Similar research in the area of electrified system also used machine learning based on battery health estimation, charging system supervisory control and management of electrical assets (Ng et al., 2020; Nademi & Zhang, 2020; Hoffmann et al., 2020). Later works expanded EV analysis using feature-rich and interpretable spatiotemporal models (Cao et al., 2024; Shang et al., 2025). However, these studies mainly project the interaction of vehicles and the charging infrastructure instead of detecting if the very charging equipment is about to fail in maintenance. This distinction matter significantly because a demand model can reach high precision while offering little insight about the short-term reliability of the charger delivering that service.

Charger utilisation is also extremely heterogeneous in real-world charging studies. In the context of lower availability and unequal uptake in public rapid-charging, Olanrewaju & Aderinko (2026a) show congruence in arrival state of charge, charging duration and partial replenishment behaviour using UK public-charging telemetry. Their results indicate that public chargers operate according to user choices, as opposed to the theoretical cycles. Using high-resolution data, Olanrewaju and Aderinko (2026b) assessed interactions between EV charging, photovoltaic generation and battery storage at the system level, revealing that peak load and energy balance responds to changing charging demand. These studies collectively show the benefit of real operation data to shed light on charger adoption and energy-system impacts. This study extends that data-driven perspective into short-horizon infrastructure fault prediction.

Research specifically examining charging-equipment condition is smaller, but also developing. Shuvo and Yilmaz (2020) associated predictive maintenance based reasoning with growing charging load of electric vehicle in distribution networks. Izquierdo Gomez et al. (2022) used thermal modelling for charger anomaly detection. Also, Sakwa et al. (2024) researched early EVSE failure to diagnose based on autoencoders. More recently, Wang et al. (2025) used an adaptive dynamic thresholds-based fault prediction for EV charging equipment. These studies demonstrate the viability of data driven charger monitoring. Nonetheless, anomaly detection is not the same as current fault detection and prospective fault prediction. Anomaly detection detects deviations from expected behaviour and predictive maintenance needs to estimate an event that has not annonated prior to the predicted label instance (Pang et al., 2021).

Three gaps are especially relevant. First, one of the prerequisites for short-horizon maintenance prediction is that it must be able to distinguish between information disclosed or observable at prediction time and events taking place after this point. This is necessary to prevent future information from leaking into feature construction, model selection or threshold tuning and inflating apparent performance. Second, fault recurrence and accumulated workload need to be analyzed together since they are different measures of charger history collected from test data. Third, they have other requirements than pure discrimination performance. The predictions that they make should be interpretable, calibrated and tested under proper temporal holdout conditions. Explainable AI consists of methods that explain which features are responsible for the model risk estimates (Arrieta et al., 2020). This is particularly pertinent in predictive maintenance where trust and operational actionability impact deployment (Cummins et al., 2024). SHAP is a principled method for model attribution (Lundberg et al. 2020), but true interpretation of Shapley values requires all inputs as correlational feature dependence must be controlled for (Aas et al., 2021).

Accordingly, this study proposes an explainable machine-learning approach that classifies the recorded maintenance-related EV-charger faults in the next 3 calendar days. The framework includes current operational and environmental factors as well as fault and workload histories of length 7, 14, and 30 days, and compares 4 machine learning algorithms in an embargoed temporal validation, locks on a separate temporal validation period, and evaluates the selected model on a later temporal holdout. Other assessments include robustness, calibration, and clustered uncertainty and correlation-aware explainability. Operational domain shift and limits to model transportability is explored using UK and Swedish charging data.

- *Aim, objectives and Research Question*

The aim of the study was to develop and evaluate an explainable short-horizon model for predicting recorded maintenance-relevant faults in EV chargers using real-world operational and fault-history data. Specific objectives were to: (1) characterize charging activity and maintenance-relevant fault patterns in the Chinese charging data set; (2) develop leakage-safe temporal predictors and compare four models for making three-day prospective fault predictions, namely Logistic Regression, Random Forest, XGBoost and LightGBM; (3) evaluate the selected model with respect to an independently chosen operating threshold, temporal holdout performance, calibration, robustness and explainability; and (4) compare the operational differences between the Chinese, UK and Swedish data sets and implications for model transportability.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

The research question in this study is formulated as follows: To what extent can historical operational and fault information be used to explain the faults of the EV-charger, which are relevant to maintenance, over the next three days? The study was designed as a predictive study with prespecified objectives and a research question, instead of a classical null-hypothesis test. Model choice and performance claims were based on temporal validation and out-of-time predictive evaluation.

II. MATERIALS AND METHODS

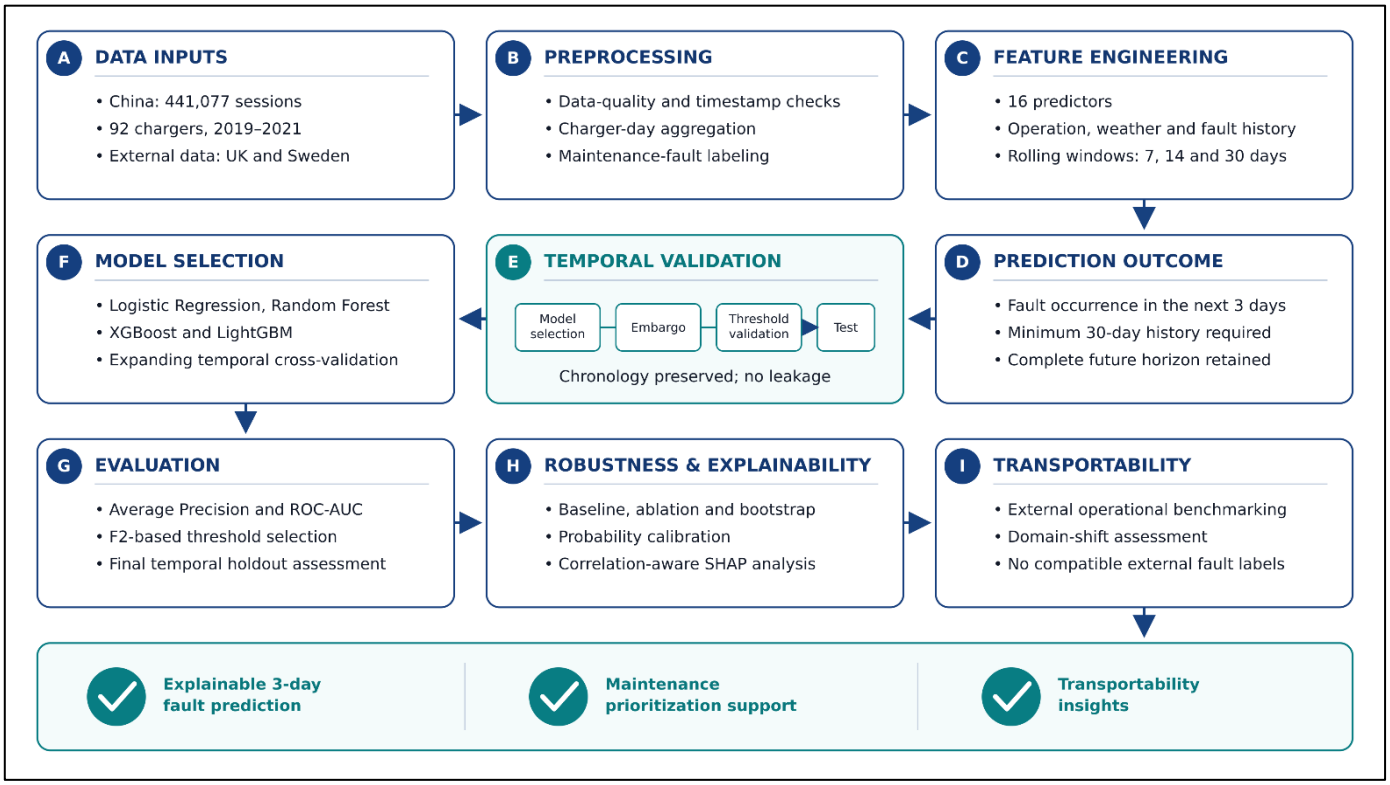


Fig 1 Methodological Framework for Explainable Short-Horizon Prediction of Maintenance-Relevant EV-Charger Faults and External Transportability Assessment.
Source: Authors' Conceptualisation.

➢ *Study Design and Data Sources*

A retrospective predictive-modelling design was used. Historical data from a charging network in China was used as the development data set, while data from charging networks in the UK and Sweden was used exclusively for external operational benchmarking and for transportability assessment. The prediction unit was the charger-day. Information available through the end of day t was used to estimate whether at least one maintenance-relevant fault would be recorded during calendar days t+1 to t+3. The analytical design maintained the sequence of the features during construction, during selection of the model, during the tuning of the thresholds, and during the final evaluation.

The Chinese dataset contained 441,077 charging sessions involving 56,115 users and 92 charging posts across 13 location categories and three districts, covering 31 December 2019 to 31 December 2021. Available variables included charger identifier, transaction energy, charging start and end times, temperature, relative humidity, rainfall and charging-termination cause. The external collection comprised 49 CSV files from CPOs in the UK and Sweden. These records contained charger and connector identifiers, session start time, delivered energy and duration. External files spanned late 2025 to early 2026, and were analysed separately due to the fact that they had different periods of data, schema, and available outcomes.

➢ *Data Preprocessing and Maintenance-Relevant Fault Definition*

The charging timestamps in Chinese text were converted to standard datetime objects and invalid values were checked. Assessment of missing values, duplicate records, negative duration and zero-duration sessions were conducted prior to modelling. The charging time was determined from the start and end times recorded. Energy, environmental and duration variables were converted to numerical form where required. The source-provided broad abnormal-event indicator was used for descriptive analysis only because it included a variety of events and was not the same as a charger specific maintenance outcome.

The recorded termination cause was used to create a rule-based maintenance-relevance taxonomy. Severity 2 included DC Output Fault and System communications failure. Severity 3 included Power conversion unit failure,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

DC switching device failure, AC switching device failure and AC Input Failure. All sessions not falling into any of these six categories were marked as zero with the maintenance-relevance indicator. Severity 3 events were also kept as an independent critical failure audit result. These categories were study-specific proxies for maintenance relevance, not manufacturer confirmed health states, direct physical degradation measurements or confirmed maintenance work orders.

- *Charger-Day Construction, Temporal Predictors and Outcome*

The data from each session was collated by charger identifier and calendar date. The following parameters were measured every day: charging session count, total delivered energy, mean energy/session, mean charging duration, mean temperature, mean relative humidity, rainfall, and counts of maintenance-relevant events. Each of these chargers was then extended between its first and last date observed to form a daily calendar. Dates that had no transactions recorded had zero sessions, delivered energy and zero fault events recorded. Mean session energy, duration and weather were not imputed for inactive dates. These dates were not assumed to be a healthy charger and no recorded transaction or fault event occurred. Only days with at least one charging transaction were used for prediction anchors.

All of the historical predictors were computed after chronological sorting in each charger. All rolling variables were shifted by one day prior to calculating otherwise prediction-day or future information would be fed into the past. Complete 7-, 14- and 30-day windows were used. For a window of length w, historical fault burden was defined as:

$$H_{c,t}^{(w)} = \sum_{j=1}^{w} F_{c,t-j}$$

Where F(c, t-j) is 1 if charger c recorded at least one fault that is relevant for maintenance in the corresponding previous day, and 0 otherwise. Historical workload was represented by 7-, 14- and 30-day mean daily energy and mean daily session frequency. The final primary predictor set contained 16 variables: current daily sessions, daily energy, mean session energy, mean duration, temperature, humidity and rainfall; previous 7-, 14- and 30-day fault days; and 7-, 14- and 30-day mean energy and session frequency. Only charger-days with a complete 30-day historical window were eligible for modelling. The main results were whether at least one fault was found that needs maintenance within the three calendar days following prediction day t. It was defined as:

$$Y_{c,t}^{(3)} = max\left\{F_{c,t+1}, F_{c,t+2}, F_{c,t+3}\right\}$$

The result was therefore 1 if any maintenance relevant fault was entered between day t+1 and day t+3, and 0 otherwise. Explicit one-, two- and three-day leads were generated before taking the maximum, and a target was assigned only when the complete three-day future horizon existed. The resulting analytical data set was 55,123 eligible charger day prediction anchors. Future charging sessions were not retained as a predictor, but only for outcome-observability sensitivity analysis.

- *Leakage-Controlled Temporal Validation Design*

Modelling dataset was partitioned chronologically, not randomly. The last temporal holdout started on 1 July 2021. Since the result was three days into the future, 28-30 June 2021 was also a final embargo period between development and holdout data. The model development was then split into a model-selection period (30 January 2020 to 13 March 2021), a three-day internal embargo period (14–16 March) and a reserved threshold-validation period (17 March to 27 June 2021). This structure supported the latest outcome that led to a model-selection label coming before the start of threshold validation, and the latest threshold-validation outcome occurring before the temporal holdout began.

Within the model-selection period, three expanding calendar-based time-series folds were created. Each fold completed training early enough that the entire outcome window (t+1 to t+3) was complete before the end of the validation window. Target window overlaps between training and validation were avoided because this was prevented by the horizon-specific embargo. Future-time generalisation in the observed fleet was thus evaluated in the final holdout. It was not meant to be a leave-charger-out test of entirely unseen equipment.

- *Model Development, Threshold Selection and Performance Assessment*

Four supervised classifiers were compared: Logistic Regression, Random Forest, XGBoost and LightGBM. The selected set was used to compare an interpretable regularised linear model with nonlinear ensemble models which are commonly employed in predictive maintenance (Carvalho et al., 2019; Susto et al., 2015). Within the logistic regression modelling pipeline, the scaling parameters were standardised to ensure that they were only learned from the data used for the training samples. Leakage-safe temporal search was used to select hyperparameters. Twenty sampled configurations were assessed for each of the tree-based models and all 18 Logistic Regression configurations were evaluated. The primary model-selection criterion was Average Precision, as the outcome was imbalanced, with Average ROC-AUC being the secondary discrimination measure. Optimisation for threshold and final holdout evaluation were performed after the model family and hyperparameters were locked. Only the threshold validation set was used to test classification thresholds between 0.05 and 0.95 in steps of 0.01. The operating threshold maximised the F2 score, which gives greater weight to recall than precision:

$$F_2 = \frac{5PR}{4P + R}$$

Where P is precision and R is recall. If two thresholds produced the same F2, recall and then precision were used as tie-breakers. The traditional 0.50 was kept for comparison. Final performance measures included accuracy, precision, recall, specificity, F1, F2, ROC-AUC, Average Precision,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

confusion-matrix counts and alert rate. Probability calibration was assessed using the Brier score, log loss, a ten-bin quantile calibration curve and expected calibration error. No calibration was done on the last final temporal holdout.

➢ *Robustness and Uncertainty Analyses*

Several prespecified sensitivity analyses were conducted. The first was a comparison to a simple baseline using seven days of fault history, where a fault in the past seven days resulted in a positive alert. Second, there was a limit on the number of days allowed to include outcome-observability sensitivity analysis, limiting the number of prediction anchors with at least one charging session on the next three days. Future exposure was only used to identify this evaluation set and was not included in the model training. Third, a complete 16-feature model was compared with a three-variable fault-history-only and 13-variable operational/environmental model without fault history using fixed-hyperparameter ablation. A reduced-collinearity sensitivity model was used to select representative predictors from highly overlapping families rolling windows. The uncertainty of final holdout performance was estimated using 1,000 charger-cluster bootstrap replications. Charger identifiers were sampled with replacement and all observations for each sampled charger were kept together, this maintained the dependence of observations within charger. Percentile-based 95% confidence intervals were obtained from the 2.5th and 97.5th percentiles of the bootstrap distributions.

➢ *Predictor Dependence and Explainability*

The Spearman rank correlation was calculated for pairwise predictor dependence, and significant correlations (absolute values of 0.70 or greater) were considered to be high. This audit was significant because 7-, 14- and 30-day histories may overlap and share a large amount of information. Model associations, in the form of standardised coefficients and odds ratios per one-standard-deviation increase, were reported, not causal effects for the selected Logistic Regression model.

Global explanation used SHapley Additive exPlanations (Lundberg et al., 2020). To avoid assuming feature independence, following the rationale of Aas et al. (2021), several predictors were strongly correlated, so the SHAP values were calculated using the conditional correlation approach, which uses LinearExplainer and a linear imputation masker. A reproducible sample of 3,000 temporal-holdout observations was explained using 1,000 development observations as background data. Numerical reconstruction of the Logistic Regression decision function was verified by an additivity audit. SHAP values were not interpreted as causal estimates, but as model attributions.

➢ *Operational Benchmarking and Reproduction Outside the System*

The UK and Swedish files were audited separately before being compared. The session records needed to have valid charger identifier, connector identifier, a parseable start time, and a non-negative energy, and a non-negative duration. Duplicate sessions were detected by country, operator, charger, connector, start time, delivered energy and duration [rounded to 3 decimal places]. Only the first identical record was retained. Then external fault-field availability was evaluated prior to any attempt of model transfer. Direct external predictive validation was not performed as the external datasets were not compatible with the future maintenance-fault outcome and some contained fewer predictors than others.

The primary external comparison was on charger-day observation; and the granularity sensitivity analysis was performed using connector-day aggregation. To prevent domination by the most frequent daily charger, the mean daily session, mean daily energy, energy per session and session-weighted mean duration were computed at the device level and summarised in regions. Operational domain shift was identified through standardised mean differences from the Chinese development domain and through the percentage of external charger-days below, within and above the 1st - 99th percentile range of the Chinese operational domain. Uncertainty at the device level was calculated using 2,000 bootstrap replications. These comparisons were only descriptive and did not assume the nature of a causal country effect.

All analyses were performed in Python using a fixed random seed of 42. The final fitted model, locked threshold, predictions, analytical audit tables, software versions and metadata were saved programmatically. The serialized model was reloaded to ensure that all the original holdout probabilities were reproduced exactly.

III. RESULTS

➢ *Dataset Characteristics and Data Quality*

The main Chinese charging dataset was 441,077 charging sessions containing 24 original variables. The records represented 56,115 users, 92 charging posts, 13 location categories and three districts, spanning 31 December 2019 to 31 December 2021. Data-quality assessment showed no missing values, duplicate records or invalid timestamps. Following timestamp standardisation, charging duration was calculated for all sessions. The mean charging duration was 54.70 minutes (SD = 94.68); median = 33.28 minutes and interquartile range (IQR) = 13.32 to 59.65 minutes. There were no negative durations noted, but five zero-duration sessions were retained in the source data audit.

The source-provided broad abnormal-event indicator determined 81,934 sessions (18.58%) for abnormal and 359,143 sessions (81.42%) for normal. The charging process was completed normally in 203,201 sessions (46.07%), and 155,942 sessions (35.35%) were terminated by the user. Other charging-pile faults (20,114; 4.56%), DC output faults (14,813; 3.36%) and vehicle communication failures (13,936; 3.16%) were the most common among the equipment and communication-related categories. The broad abnormal indicator was used solely for descriptive exploration and was not used as predictive-maintenance outcome.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

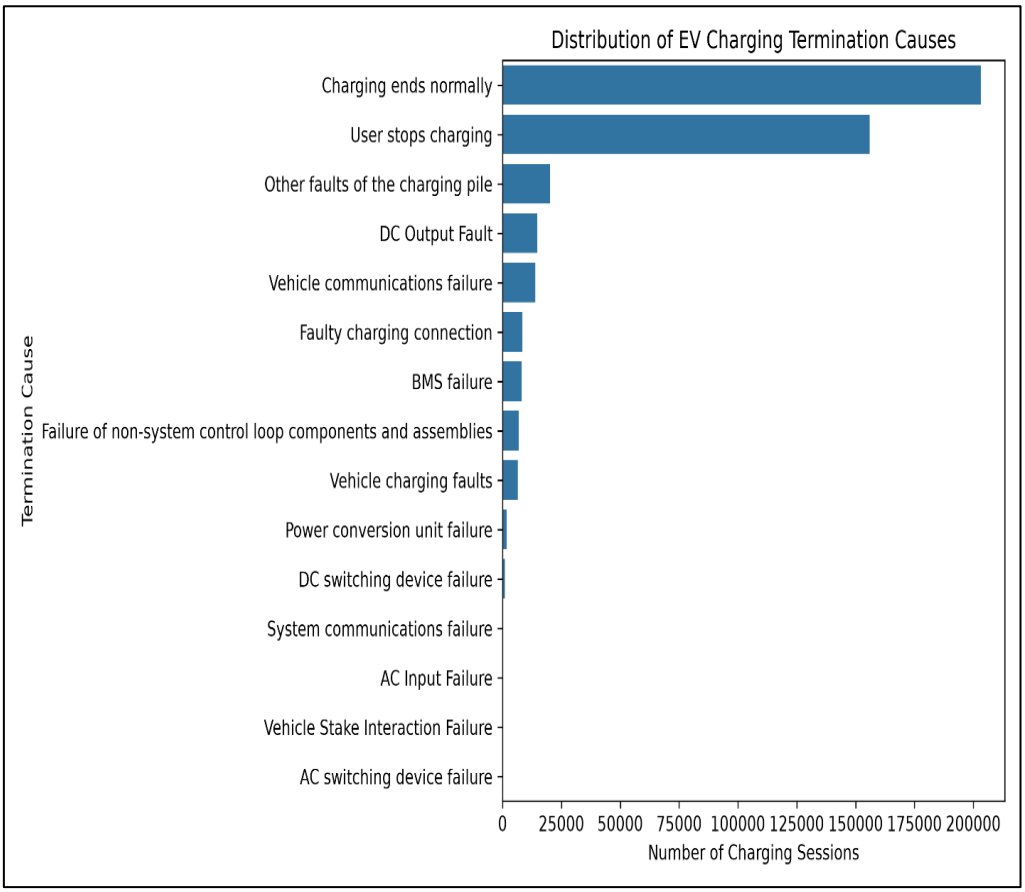


Fig 2 Distribution of EV Charging Termination Causes

Table 1 Dataset Characteristics and Data Quality

Characteristic	Result
Charging sessions	441077
Unique users	56115
Charging posts	92
Location categories	13
Districts	3
Study period	31 Dec 2019–31 Dec 2021
Missing values	0
Duplicate records	0
Invalid timestamps	0
Mean charging energy, kWh	16.44
Median charging energy, kWh	15.15
Mean charging duration, min	54.7
SD charging duration, min	94.68
Median charging duration, min	33.28
Charging duration IQR, min	13.32–59.65
Negative-duration sessions	0
Zero-duration sessions	5
Broad abnormal-event sessions	81,934 (18.58%)

Note. The broad is_abnormal source variable was employed for descriptive exploratory analysis but was not the same with the rule-based maintenance-relevant fault outcome that was modelled.

➢ *Definition of Maintenance-Relevant Fault and Construction of Predictive Dataset*

The prespecified rule-based maintenance-relevance taxonomy resulted in the identification of 17,790 maintenance-relevant fault events, of which 2,815 were categorized as critical equipment-failure events. The maintenance-relevant category consisted of the following categories: DC output faults, system communication failures, power conversion unit failures, DC switching-device failures, AC input failures, and AC switching-device failures. The largest category was DC output faults (14,813 events).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

There were 57,239 charger-days observed on 92 chargers, with aggregation of transactions. The expanded table to a continuous charger-specific calendar contained 66,715 calendar rows, with 9,476 days of no charging transactions. These zero-transaction dates were intentionally provided to maintain continuity of the calendar and not necessarily indicative of good charger states. The mean percentage of days for which charging activity was observed across chargers was 85.85%.

1,842 observations were dropped after applying the restriction on prediction anchors (observed charger-days) with a full historical window of 30 days. Additional 274 observations were removed due to not having a full three-day target horizon. The modelling dataset consisted of 55,123 observations of charger-day, with 16,538 (30.00%) having at least one fault of interest during the calendar days t+1 to t+3 and 38,585 (70.00%) having no fault of interest during these calendar days. There were only 638 prediction anchors without any charging session recorded in the three days that followed (1.16%).

Table 2 Maintenance-Relevant Fault Taxonomy and Construction of the Predictive Cohort

Panel / category	Classification	N
A. Maintenance-relevant event taxonomy		
DC Output Fault	Maintenance severity 2	14813
System communications failure	Maintenance severity 2	162
Power conversion unit failure	Maintenance severity 3	1653
DC switching device failure	Maintenance severity 3	1058
AC Input Failure	Maintenance severity 3	99
AC switching device failure	Maintenance severity 3	5
Total maintenance-relevant events	Severity ≥2	17790
Critical failure events	Severity 3	2815
B. Predictive dataset construction		
Observed charger-days	Initial daily aggregation	57239
Continuous charger-calendar rows	Dense daily calendar	66715
No-transaction calendar days	Within observed charger lifespan	9476
Rows with complete 30-day history	After history restriction	55397
Final modelling rows	Complete history + 3-day target	55123
Negative primary targets	No fault in (t+1:t+3)	38,585 (70.00%)
Positive primary targets	≥1 fault in (t+1:t+3)	16,538 (30.00%)
Zero future-session windows	Sensitivity audit	638 (1.16%)

Note. Severity designations are a study specific, rule-based proxy for maintenance relevance, and should not be interpreted as a formal manufacturer or industry severity designations.

➢ *Temporal Model Selection and Predictive Performance*

The leakage-controlled model-selection period consisted of 31,179 observations, during which temporal model selection was carried out. Three calendar-based validation folds were defined with horizon-specific embargoes to ensure that outcomes used for training labels were earlier than the validation intervals. Logistic Regression achieved the highest mean cross-validated Average Precision (AP = 0.701, SD = 0.041) and mean ROC-AUC (0.817, SD = 0.012), compared with XGBoost (AP = 0.685), LightGBM (AP = 0.684) and Random Forest (AP = 0.664). Logistic Regression was therefore selected, and preceeded threshold optimisation and temporal holdout evaluation.

Table 3 Leakage-Safe Temporal Model Selection and Descriptive Holdout Discrimination

Model	CV AP, mean ± SD	CV ROC-AUC, mean ± SD	Holdout AP	Holdout ROC-AUC
Logistic Regression	0.701 ± 0.041	0.817 ± 0.012	0.75	0.806
XGBoost	0.685 ± 0.028	0.813 ± 0.009	0.749	0.804
LightGBM	0.684 ± 0.027	0.812 ± 0.009	0.75	0.805
Random Forest	0.664 ± 0.043	0.798 ± 0.019	0.745	0.801

Note. AP = Average Precision. Holdout results for non-selected models are descriptive only; they were not used to change the model selected by development-period cross-validation.

A total of 8,444 observation values were recorded during the reserved threshold-validation period, of which 27.40% were faulty. The selected model had a precision of 0.787, recall of 0.481, F2 of 0.522 at the conventional threshold of 0.50. The operating threshold was set to 0.16 by maximising F2 during the reserved validation period, resulting in recall of 0.836, F2 of 0.714, precision of 0.450 and specificity of 0.615. The corresponding alert rate was 50.86%. Then the threshold was locked without any changes.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

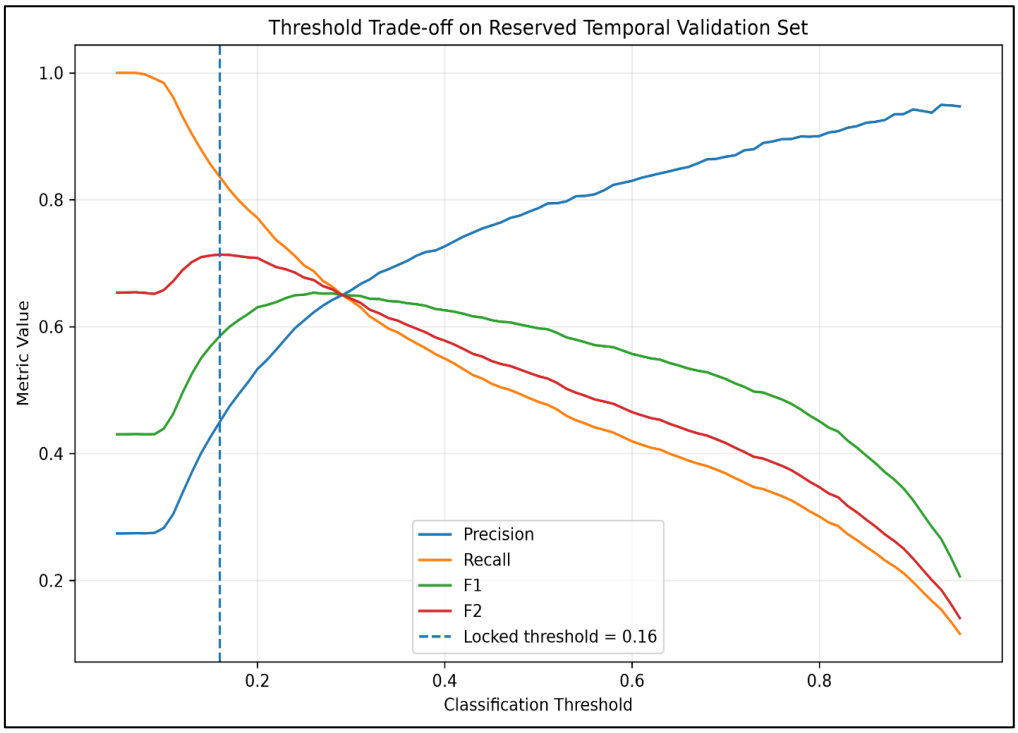


Fig 4 Threshold Trade-off on Reserved Temporal Validation Set

The out-of-time temporal holdout contained 15,026 charger-days from 1 July to 28 December 2021 and had a higher outcome prevalence of 37.41%. The 92 chargers in the holdout were all present in the development data, so this evaluation is actually a generalisation for future time and not generalisation for unseen chargers. At the locked threshold of 0.16, Logistic Regression achieved ROC-AUC 0.806, AP 0.750, recall 0.875, precision 0.501, specificity 0.480 and F2 0.762. The confusion matrix comprised 4,510 true negatives, 4,895 false positives, 700 false negatives and 4,921 true positives. Charger-level cluster bootstrap analysis using 1,000 replications produced a 95% CI of 0.769–0.836 for ROC-AUC and 0.667–0.805 for AP; the recall CI was 0.835–0.905 and the F2 CI was 0.712–0.804. The bootstrap was successfully completed for all 1,000 replications.

Table 4 Performance of the Selected Logistic Regression Model Across Validation and Temporal Holdout Sets

Metric	**Validation, threshold 0.50**	**Validation, threshold 0.16**	**Holdout, threshold 0.50**	**Holdout, threshold 0.16**
Accuracy	0.822	0.675	0.767	0.628
Precision	0.787	0.45	0.787	0.501
Recall	0.481	0.836	0.516	0.875
Specificity	0.951	0.615	0.916	0.48
F1	0.597	0.585	0.623	0.638
F2	0.522	0.714	0.554	0.762
Alert rate	0.168	0.509	0.246	0.653
ROC-AUC	0.828	0.828	0.806	0.806
Average Precision	0.719	0.719	0.75	0.75

Note. The 0.16 threshold was selected exclusively on the reserved temporal validation set by maximising F2 and remained fixed for the temporal holdout. The threshold should be described as sensitivity-oriented, not as an economically optimal maintenance threshold.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

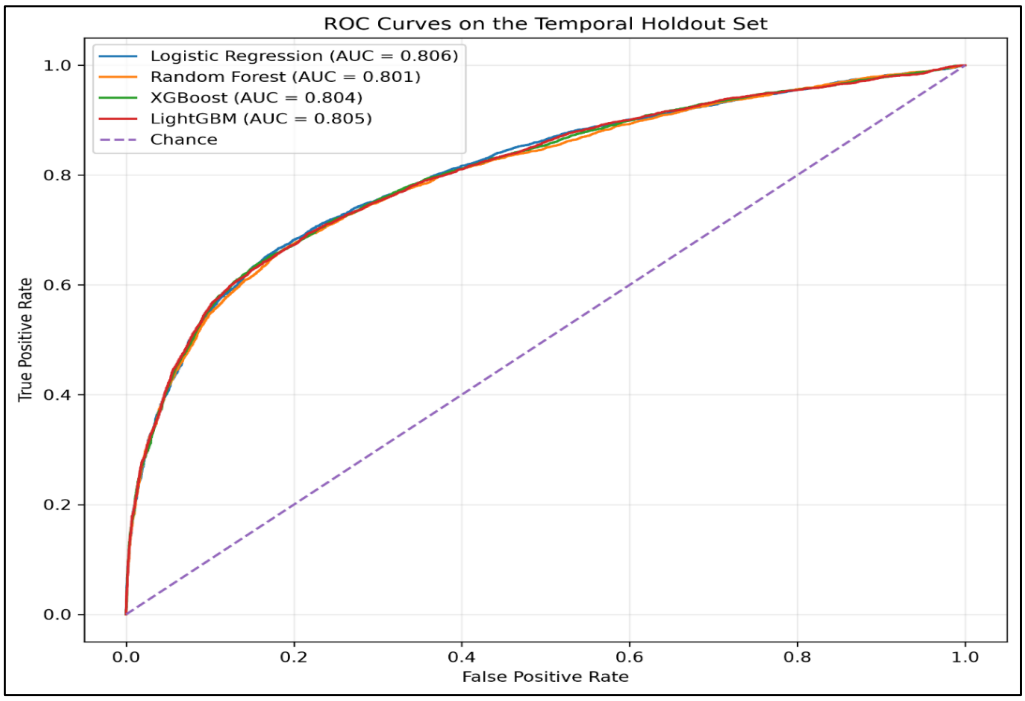


Fig 5 ROC Curves on the Temporal Holdout Set

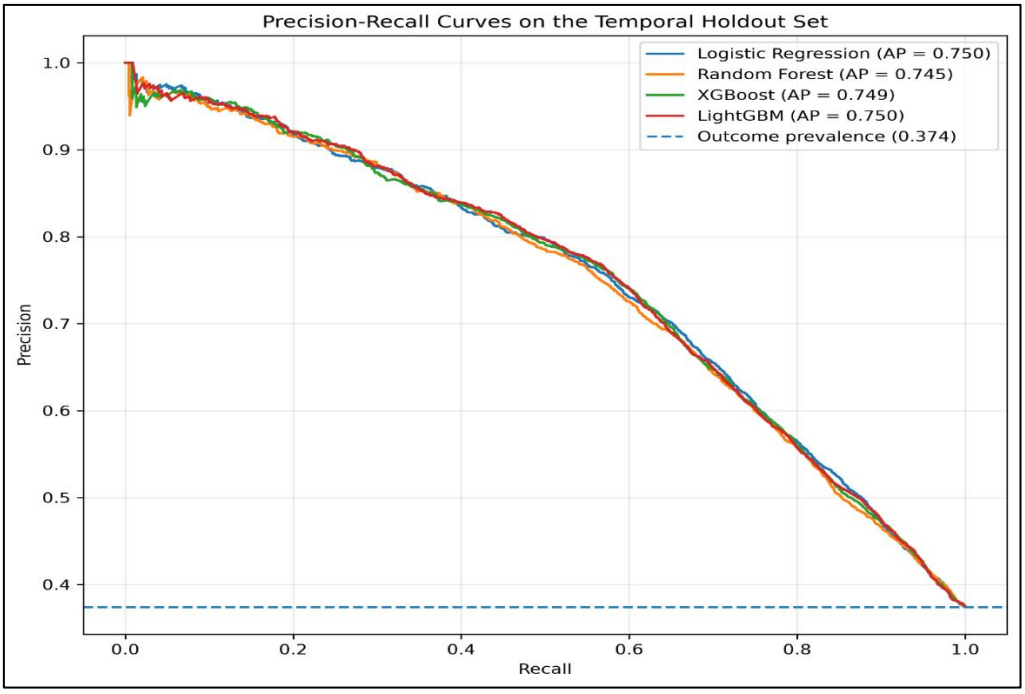


Fig 6 Precision-Recall Curves on the Temporal Holdout Set

➢ *Robustness, Calibration and Model Explainability*

This selected model significantly outperformed a simple seven-day persistence based on fault history. When comparing to the full Logistic Regression model whose performance was 0.806 ROC-AUC and 0.750 AP, the persistence baseline performed at 0.769 ROC-AUC and 0.672 AP. Excluding the 66 temporal-holdout observations with no charging session during the subsequent three days produced almost unchanged performance: ROC-AUC increased from 0.806 to 0.806, AP from 0.750 to 0.752 and F2 from 0.762 to 0.763. This suggested that the main performance estimates were not based on windows that would never be exposed to charging again in the future. Historical fault burden was responsible for the majority of the predictive information through feature-set ablation. The model using only the three historical fault-day variables had an AP of 0.744 and ROC-AUC of 0.804 while the 16-feature model had an AP of 0.750 and an ROC-AUC of 0.806. In contrast, the operational and environmental variables without any fault history had an AP of 0.506 and ROC-AUC of 0.617. A reduced-collinearity 7-feature sensitivity model yielded similar AP (0.735) and ROC-AUC (0.800), indicating that the main predictive signal was robust even though there was significant correlation between the rolling-window variables.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

Table 5 Baseline Comparison and Sensitivity Analyses

Analysis	No. of features / rows	ROC-AUC	Average Precision	Brier score	Log loss
Primary 16-feature model	16	0.806	0.75	0.167	0.51
Fault-history-only model	3	0.804	0.744	0.168	0.512
Operational/environmental only	13	0.617	0.506	0.229	0.658
Reduced-collinearity model	7	0.8	0.735	0.171	0.521
Seven-day persistence baseline	Simple score	0.769	0.672	—	—
Future-exposure sensitivity	14,960 rows	0.806	0.752	0.167	—

Calibration assessment yielded a Brier score of 0.167, log loss of 0.510 and expected calibration error of 0.0397. For several intermediate risk bins, the predicted risk was lower than the observed risk, and in the highest predicted risk bins, the predicted risk was slightly higher than the observed risk; no recalibration was performed for the temporal holdout.

A predictor-correlation audit highlighted 31 feature pairs with absolute Spearman correlation ≥ 0.70, especially those features which overlap 7-, 14- and 30-days workload. Hence, the individual regression coefficients were considered as the associations of the models and not as independent causal effects. The strongest standardized coefficient was observed for the previous 30-day fault-day burden (standardised coefficient = 0.914; OR per 1 SD = 2.49), followed by the previous seven-day fault burden (OR = 1.46) and previous 14-day fault burden (OR = 1.22).

The correlation-aware SHAP analysis confirmed the significance of the fault history. The historical fault-burden domain accounted for 67.77% of the total mean absolute SHAP attribution, while the current operational state, historical energy workload, environment and historical session workload made up 8.65%, 8.18%, 7.80% and 7.59%, respectively. The three highest-ranking individual features were previous 30-day, 14-day and seven-day fault days.

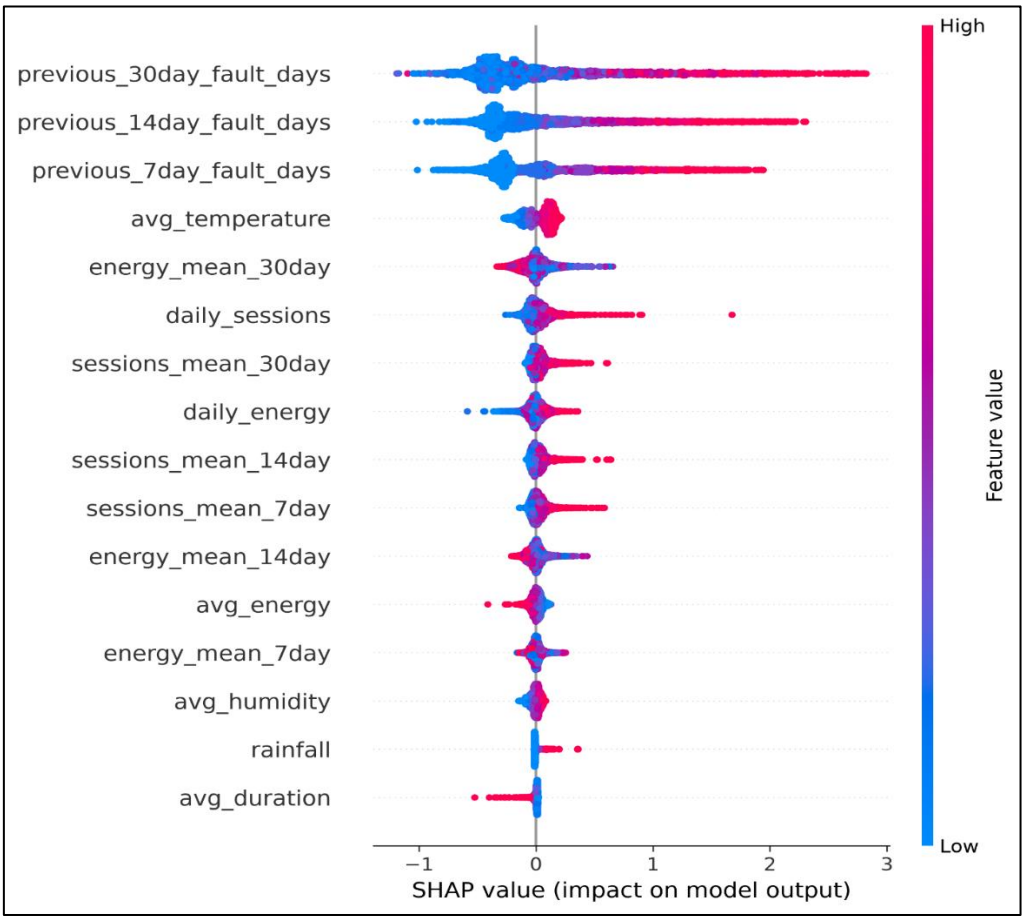


Fig 7 SHAP Summary Plot of Feature Contributions to EV Charger Degradation Prediction

➢ *External Operational Benchmarking and Transportability*

The external UK/Swedish collection consisted of 49 CSV files with a total of 23,124 merged records. There were 22,854 session-like records, of which 18,261 were valid with parsed non-negative energy and duration values. A total of 17,491 charging sessions were deduplicated, with 770 duplicate session keys being detected through cross-file duplicate auditing. These comprised 16,494 UK sessions from 26 chargers and 997 Swedish sessions from eight chargers.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

The aggregation resulted in 2,060 UK charger-days and 426 Swedish charger-days. However, the available external health tables contained no observations for the warning, medium-impact-error or high-impact-error fields. Thus, it was not possible to build a network with an outcome that fits within the Chinese three-day maintenance-fault constraint. Only four of the 16 model variables were directly available, six workload features were constructible from session histories and six variables, comprising the three fault-history and three weather predictors, were unavailable. Predictive validation by direct external was not conducted.

However, significant domain shift was found at the device level, with operational profiles. In the Chinese model-development domain, the mean number of daily sessions per charger was 6.66, whereas in the UK it was 7.97 and in Sweden 2.05. Mean daily energy was 108.99 kWh in China, 217.40 kWh in the UK and 174.89 kWh in Sweden. The mean energy per session was also significantly different, in China at 15.97 kWh and in the UK at 27.42 kWh, and in Sweden at 81.93 kWh, whereas the mean session duration was lower for the external: 30.11 minutes in the UK and 34.51 minutes in Sweden, as compared to 52.26 minutes in China.

Relative to China, the standardized mean differences for UK devices were 0.41 for daily session frequency, 1.40 for daily energy, 4.23 for energy per session and −1.69 for duration. Corresponding Swedish differences were −2.16, 0.82, 2.83 and −1.28. Range-coverage analysis was also applied, revealing that 96.36% of charger-days for UK charge and 30.05% of charger-days for Sweden fall within the range of 1st – 99th percentile for average energy per session as defined by the Chinese, and that 69.95% of the Swedish sample exceeds the Chinese upper reference limit. The results show significant covariate shift at the operational level, and highlight the importance of locally labelled fault data to claim predictive transportability.

Table 6 Device-Level Operational Benchmarking Across Chinese, UK and Swedish Charging Datasets

Operational domain	Devices	Median active days	Mean daily sessions, 95% CI	Mean daily energy, kWh, 95% CI	Energy/session, kWh, 95% CI	Mean duration, min, 95% CI
China, State Grid	92	453	6.66 (6.04–7.26)	108.99 (96.96–121.32)	15.97 (15.26–16.71)	52.26 (48.51–56.26)
United Kingdom	26	86.5	7.97 (6.68–9.24)	217.40 (184.28–251.49)	27.42 (26.87–28.03)	30.11 (28.82–31.39)
Sweden,	8	52.5	2.05 (1.63–2.55)	174.89 (114.79–236.71)	81.93 (59.86–102.61)	34.51 (30.06–39.26)

Note. Confidence intervals were obtained using device-level bootstrap resampling. Comparisons describe operational domain shift and must not be interpreted as causal country effects. The external data did not contain a compatible maintenance-fault outcome and therefore were not used for direct predictive validation.

IV. DISCUSSIONS

- *Main Findings and Short-Horizon Fault-Prediction Performance*

This research aimed to build an interpretable machine-learning model to forecast the occurrence of a maintenance-relevant fault at an EV charger within the next three calendar days. There are several important findings. First, a relatively simple regularised Logistic Regression model performed better during temporal model selection than Random Forest, XGBoost and LightGBM. Second, the model maintained good discrimination in the later temporal holdout with a ROC-AUC of 0.806 and Average precision of 0.750. Third, recent fault history accounted for most of the predictive information. Finally, significant differences in the setups and operation of the datasets between China, UK and Sweden revealed that transportability of the model should not be taken for granted without locally labeled fault data.

The predictive performance is consistent with the general findings that operating data from the past can be leveraged to predict equipment faults, not just identify anomalies after they happen. Research in predictive maintenance has become more focussed on data-driven approaches, which gain patterns from past operating states and failure events (Carvalho et al., 2019). The importance of explicitly temporal prediction is also consistent with Zonta et al. (2020), who identified time-based predictive-maintenance approaches as an important research challenge. In the present study, this matter is addressed by providing a prediction anchor that is separated from the next three-day outcome window and by keeping chronological separation throughout the model development process.

This is a significant difference in EV charging studies. Izquierdo Gómez et al. (2022) employed thermal modelling and machine learning to detect anomalous charger behavior, and Sakwa et al. (2024) applied autoencoder and LSTM models for early anomaly and fault estimation from charging-unit power profiles. While these methods show the benefit of data-driven condition monitoring, they also show that the analytical tasks are different than prospective short horizon classification. More recently, Wang et al. (2025) developed an adaptive-threshold framework for charging-equipment fault prediction using environmental, operating and fault-related factors. In this regard, the present study joins the trend of moving away from contemporaneous anomaly detection toward predicting future events of interest for maintenance.

The selection of Logistic Regression is also noteworthy. When equipment behaviour involves nonlinear interactions, complex models are often supposed to perform better than simpler algorithms. In this context, deep and ensemble learning has been the subject of extensive studies on machine-health monitoring (Zhao et al., 2019). But, greater model complexity does not guarantee better temporal generalisation.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

In this study, XGBoost and LightGBM produced holdout discrimination very close to Logistic Regression, but neither improved the prespecified temporal model-selection criterion. This result indicates that the predictive structure was largely obtained from the engineered historical features. It also demonstrates the importance of considering models under the same leakage-controlled validation design, instead of picking an algorithm for its complexity.

- *Fault Recurrence was the Dominant Short-Horizon Risk Signal*

The ROC-AUC of the model using only the three fault-history variables was 0.804, while the complete 16-feature model had a ROC-AUC of 0.806 and an AP of 0.750. The ROC-AUC and AP obtained by the operational and environmental variables without any fault history were 0.617 and 0.506, respectively. These findings indicate that recent recurrence of maintenance-relevant faults was much more informative for three-day risk than charging volume, energy throughput, duration or weather alone.

This outcome is in line with one of the core principles of condition-based maintenance: that information about the past condition of equipment can be stored in its recent history. Historical condition data are usually converted to health indicators prior to future-state prediction in machinery prognostics (Lei et al., 2018). Predictive-maintenance frameworks also rely on the present and past state of a system to identify the existing patterns of normal operation from the patterns of development of failure (Carvalho et al., 2019). For EV infrastructure specifically, Hsu et al. (2023) identified prognostics and health management as an important component of intelligent charging-system maintenance and highlighted the need for robust measurements and condition indicators.

However, care must be taken when interpreting the finding. Repeated termination codes could be indicative of continuous or reoccurring failure processes, not necessarily physical failure. The model thus forecasts maintenance relevant fault events as opposed to component degradation measured directly by sensors or maintenance inspections. This distinction also helps to understand why investigators shouldn't report previous fault-day counts as actual measurements of degradation. They are not mechanistic and their value is predictive.

The small improvement from the history only model to the full model is also informative. The addition of operational and environmental variables was not without value, but was limited based on the fault history of the recent years. This differs from studies of EV charging demand, where operational, temporal and external features can play a major role in predicting utilisation. For example, Cao et al. (2024) improved charging-demand forecasts through enhanced external-feature selection, while Shang et al. (2025) used spatiotemporal information to predict charging occupancy, volume and duration. The contrast shows that while predictors can be useful in charging demand, they don't necessarily provide the same information when it comes to short-horizon charger fault risk.

- *Explainability and the Value of Model Transparency*

Explainability is important because there is a practical implication whenever the maintenance warning appears. Depending on the type of charger, operators may have to determine which charger to investigate, how quickly to act, and if the alert is worth removing the equipment from service. Thus, XAI has been suggested as a way to enhance transparency and accountability in situations where machine-learning outputs affect operational decisions (Arrieta et al., 2020). Cummins et al. (2024) also highlighted trust and interpretability as key aspects in the context of explainable predictive maintenance.

The correlation-aware SHAP analysis gave a clear explanation on the model level. The total mean absolute SHAP attribution was 67.77% for historical fault burden. The three most influential individual features were previous 30-day, 14-day and seven-day fault days. This result is consistent with the ablation analysis, and reinforces the interpretation that most of the near-term predictive information comes from fault recurrence.

Correlation-aware explanations were significant since 31 pairs of predictors had an absolute Spearman correlation of 0.70 or higher. Standard SHAP procedures can be misleading even for linear models when strongly dependent predictors are used as independent predictors (Aas et al., 2021). The SHAP results in this study are still considered model attributions instead of cause. They explain the way the fitted model shares the predictive information between the observed variables. They do not prove that by increasing any of the operational variables a charger fault would occur directly. This separation is in line with the general XAI literature, where this separation is made between explaining the action of the model and inferring a cause (Arrieta et al., 2020).

In EV charging research, apart from maintenance, explainable machine learning is already applied. Ullah et al. (2023), for instance, applied SHAP to interpret charging-station choice predictions, and Shang et al. (2025) integrated SHAP-based interpretation into spatiotemporal charging-demand modelling. The present study is an extension of this interpretability principle to the charger fault prediction.

- *Operating Threshold, Calibration and Maintenance Implications*

Higher sensitivity was achieved in the temporal holdout with the locked threshold of 0.16. Recall reached 0.875 and F2 reached 0.762, but precision was 0.501 and the alert rate was 65.3%. This is a trade-off that is important. The threshold was deliberately chosen to help recall rather than be an economically optimal maintenance policy.

Predictive-maintenance decisions involve competing costs. An unobserved developing fault can result in downtime or disruption of service, and too many alerts can make it hard to inspect. Susto et al. (2015) demonstrated that the maintenance classifiers and prediction horizons need to be understood in the context of a practical situation regarding a trade-off between unexpected failures and maintenance

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

activities. The 0.16 threshold in the present study is therefore best viewed as a screening threshold for prioritising potentially high-risk charger-days. It would need an operator specific cost model with consequences associated with a false negative, consequences associated with a false positive, consequences associated with inspection effort and consequences associated with charger downtime.

There are also reasons for caution based on the results of the calibration. The model achieved a Brier score of 0.167 and an expected calibration error of 0.0397, but there were some intermediate ranges of probabilities with underestimation of the observed risk. There was also a higher positive-event prevalence in the later temporal holdout (37.4%), compared to the prevalence during threshold validation (27.4%). This change in time implies that discrimination can be of any value even if its chance estimates are drifting. A deployed system would therefore need regular checking of the system and re-calibration based on new labelled operating data as appropriate.

- *Transportability Across Charging Environments*

The external analysis found significant differences in operation between China, the United Kingdom and Sweden. Mean energy per session was 15.97 kWh in the Chinese reference domain, 27.42 kWh in the UK and 81.93 kWh in Sweden. The mean charging time also varied significantly. These differences illustrate that the distribution of input resources for a maintenance model may vary significantly between charging ecosystems.

This is in line with the global evidence that EV charging conditions fluctuate by country and infrastructure context. Funke et al. (2019) demonstrated that charging-infrastructure needs are heavily influenced by the local framework conditions, and that the results of a study in one country are not necessarily reproducible in another country. The context-dependent nature of charging data is also reflected in recent charging-demand studies that focus heavily on the temporal, spatial and operational patterns observed locally (Cao et al., 2024; Shang et al., 2025).

The UK and Swedish data, however, did not include a labelled maintenance-fault outcome. Six of the 16 Chinese model predictors were also unavailable externally, including fault-history and weather variables. Thus, external predictive validation was not feasible. The external datasets were then not used to determine that the model generalises internationally, but rather used to show evidence of change in the operational domain and transporting restrictions. This distinction is particularly significant since all 92 chargers in the Chinese holdout occurred during the development of the model. The current evidence is therefore supportive of the future time prediction of an observed fleet only and not prediction of unseen chargers or networks.

- *Strengths, Limitations and Future Research*

The study has a number of strengths. An analysis was conducted on over 441,000 real-world charging sessions and transformed into a prospective charger-day prediction problem. To prevent temporal leakage, there were horizon-specific embargoes, separate model selection and threshold validation periods, and a later temporal holdout. Performance was assessed using discrimination, calibration, cluster-bootstrap confidence intervals, a persistence baseline, feature ablation and exposure sensitivity. These steps give more complete information than reporting accuracy or ROC-AUC.

There are also significant limitations with the study. This results in a rule-based proxy that is based on historical charging termination codes. Not a manufacturer specified health index, physical degradation or a maintenance work order. Repeat events can then be interpreted as either fault episodes that have not been resolved or periods of repeated fault events. The analysis also lacks component-level measurements such as voltage, current waveforms, insulation condition and internal temperature. These are among the most common measurements used in condition monitoring and prognostics (CM&P) (Lei et al., 2018; Zhao et al., 2019).

Another limitation is that of generalisation. The temporal holdout assesses subsequent observations from a set of chargers already developed. The UK and Swedish datasets are helpful for the operational shift but do not provide evidence of predictive performance due to the lack of compatible fault labels. The framework should therefore be tested prospectively in future studies using unseen chargers, sites and operators. They should be correlated with charging logs and maintenance work orders and confirmed fault episodes. The external recalibration should also be considered prior to deployment in a new operating environment.

However, maintenance decisions should also be assessed in future work and not just prediction. A cost sensitive operating threshold may include inspection cost, missed fault cost, downtime and service availability. Other electrical and thermal measurements may be used to increase fault specificity. Lastly, a prospective deployment would enable investigators to find out if alerts for short time duration really lower unplanned downtime or enhance the effectiveness of maintenance work. The shift from predictive to proven maintenance benefits is a significant step forward for smart Electric Vehicle charging infrastructure.

V. CONCLUSIONS

This study introduces and tests an explainable machine-learning framework that uses real-world operational data for short-horizon forecasting of maintenance-relevant faults in EV charging infrastructure. The analysis introduced current operating conditions, environmental variables, recent fault history and historical workload and converted transaction level charging records into a charging day, thus maintaining the chronological order for prospecting prediction.

Regularised Logistic Regression had the highest performance in leakage-controlled temporal model selection of the four evaluated algorithms. The model's ROC-AUC on the later temporal holdout was 0.806 and Average Precision was 0.750. The recall of the validation-selected threshold of 0.16 was 0.875 and the F2 was 0.762 which showed good

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

sensitivity to short-horizon fault risk. The alert rate associated with this, however, is about 65%, meaning the operational deployment would require threshold adjustment according to the relative costs of missed fault and inspection and unnecessary maintenance action.

One important conclusion was that most of the predictive information was in the recent fault history. The fault-history-only model performed near the complete 16-feature model, and the correlation-aware SHAP analysis revealed that the model's attribution was primarily to fault history burden. This indicates that the fault codes related to maintenance are a significant short term warning and could form a practical basis for prioritising the inspection of chargers.

The study also showed how important it is to be careful with the transportability of the model. Marked differences in operational performance were identified between the Chinese, UK and Swedish charging datasets, and the external datasets contained no common maintenance-fault outcome and some predictors that were required for the final model were unavailable. The external analysis, therefore, confirms evidence of domain shift of operation, but is not an external predictive validation.

REFERENCES

[1]. Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, *298*, 103502. https://doi.org/10.1016/j.artint.2021.103502

[2]. Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K., & Alahmad, M. (2020). Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods. *Energies*, *13*(16), 4231. https://doi.org/10.3390/en13164231

[3]. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[4]. Cao, T., Xu, Y., Liu, G., Tao, S., Tang, W., & Sun, H. (2024). Feature-enhanced deep learning method for electric vehicle charging demand probabilistic forecasting of charging station. *Applied Energy*, *371*, 123751. https://doi.org/10.1016/j.apenergy.2024.123751

[5]. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

[6]. Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. *IEEE Access*, *12*, 57574–57602. https://doi.org/10.1109/ACCESS.2024.3391130

[7]. Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry*, *123*, 103298. https://doi.org/10.1016/j.compind.2020.103298

[8]. Funke, S. Á., Sprei, F., Gnann, T., & Plötz, P. (2019). How much charging infrastructure do electric vehicles need? A review of the evidence and international comparison. *Transportation Research Part D: Transport and Environment*, *77*, 224–242. https://doi.org/10.1016/j.trd.2019.10.024

[9]. Ge, X., Shi, L., Fu, Y., Muyeen, S. M., Zhang, Z., & He, H. (2020). Data-driven spatial-temporal prediction of electric vehicle load profile considering charging behavior. *Electric Power Systems Research*, *187*, 106469. https://doi.org/10.1016/j.epsr.2020.106469

[10]. Gilanifar, M., & Parvania, M. (2021). Clustered multi-node learning of electric vehicle charging flexibility. *Applied Energy*, *282*, 116125. https://doi.org/10.1016/j.apenergy.2020.116125

[11]. Hoffmann, M. W., Wildermuth, S., Gitzel, R., Boyaci, A., Gebhardt, J., Kaul, H., Amihai, I., Forg, B., Suriyah, M., Leibfried, T., Stich, V., Hicking, J., Bremer, M., Kaminski, L., Beverungen, D., zur Heiden, P., & Tornede, T. (2020). Integration of novel sensors and machine learning for predictive maintenance in medium voltage switchgear to enable the energy and mobility revolutions. *Sensors*, *20*(7), 2099. https://doi.org/10.3390/s20072099

[12]. Hsu, Y.-M., Ji, D.-Y., Miller, M., Jia, X., & Lee, J. (2023). Intelligent maintenance of electric vehicle battery charging systems and networks: Challenges and opportunities. *International Journal of Prognostics and Health Management*, *14*(3). https://doi.org/10.36001/ijphm.2023.v14i3.3130

[13]. Izquierdo Gómez, P., Barragán Moreno, A., Lin, J., & Dragičević, T. (2022). Data-driven thermal modelling for anomaly detection in electric vehicle charging stations. In *2022 IEEE Transportation Electrification Conference & Expo (ITEC)* (pp. 1005–1010). IEEE. https://doi.org/10.1109/ITEC53557.2022.9813767

[14]. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

[15]. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

[16]. Nademi, H., & Zhang, B. (2020). A learning-based supervisory control architecture for electric vehicle charging system paired with energy storages. In *2020 IEEE Transportation Electrification Conference &*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

Expo (ITEC) (pp. 621–625). IEEE. https://doi.org/10.1109/ITEC48692.2020.9161572

[17]. Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J., & Seh, Z. W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence, 2*, 161–170. https://doi.org/10.1038/s42256-020-0156-7

[18]. Olanrewaju, M. B., & Aderinko, H. A. (2026a). *State-of-charge driven charging behaviour at public EV chargers in the United Kingdom: An empirical study* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.6872978

[19]. Olanrewaju, M. B., & Aderinko, H. A. (2026b). Data-driven optimization of hybrid solar photovoltaic (PV) and battery energy storage systems (BESS) for electric vehicle charging using a hybrid data approach. *International Journal of Innovative Science and Research Technology, 11*(6), 2830–2847. https://doi.org/10.38124/ijisrt/26jun1757

[20]. Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys, 54*(2), Article 38, 1–38. https://doi.org/10.1145/3439950

[21]. Sakwa, M., Nespoli, A., Matrone, S., Leva, S., Guerini, A., Demartini, A., & Ogliari, E. (2024). Electric vehicle supply equipment monitoring and early fault detection through autoencoders. *Sustainable Energy, Grids and Networks, 40*, 101497. https://doi.org/10.1016/j.segan.2024.101497

[22]. Shahriar, S., Al-Ali, A. R., Osman, A. H., Dhou, S., & Nijim, M. (2020). Machine learning approaches for EV charging behavior: A review. *IEEE Access, 8*, 168980–168993. https://doi.org/10.1109/ACCESS.2020.3023388

[23]. Shang, Y., Li, D., Li, Y., & Li, S. (2025). Explainable spatiotemporal multi-task learning for electric vehicle charging demand prediction. *Applied Energy, 384*, 125460. https://doi.org/10.1016/j.apenergy.2025.125460

[24]. Shuvo, S. S., & Yilmaz, Y. (2020). Predictive maintenance for increasing EV charging load in distribution power system. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (pp. 1–6). IEEE. https://doi.org/10.1109/SmartGridComm47815.2020.9303021

[25]. Sun, D., Ou, Q., Yao, X., Gao, S., Wang, Z., Ma, W., & Li, W. (2020). Integrated human-machine intelligence for EV charging prediction in 5G smart grid. *EURASIP Journal on Wireless Communications and Networking, 2020*, Article 139. https://doi.org/10.1186/s13638-020-01752-y

[26]. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics, 11*(3), 812–820. https://doi.org/10.1109/TII.2014.2349359

[27]. Ullah, I., Liu, K., Yamamoto, T., Zahid, M., & Jamal, A. (2023). Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. *Travel Behaviour and Society, 31*, 78–92. https://doi.org/10.1016/j.tbs.2022.11.006

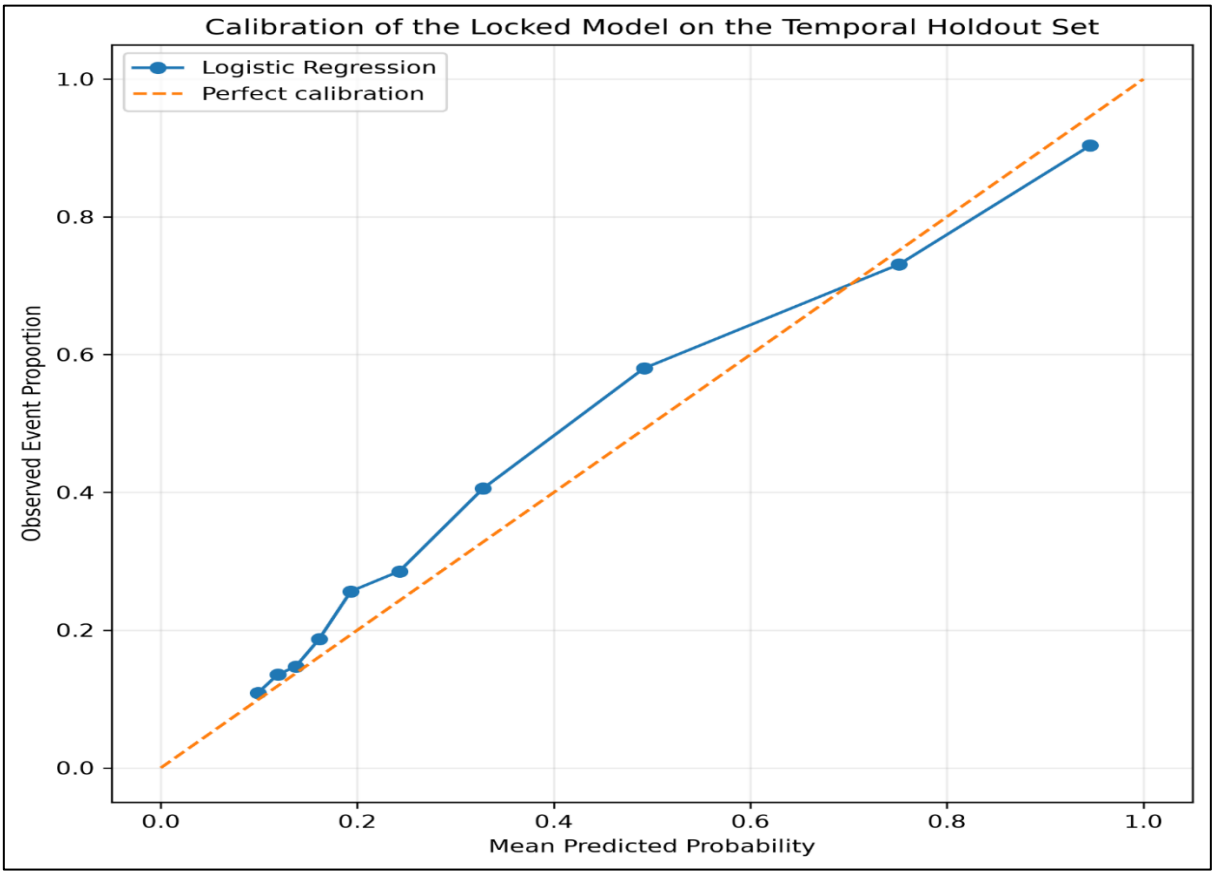
[28]. Wang, H., Wang, N., Li, Y., & Tang, X. (2025). A fault prediction model for electric vehicle charging equipment based on adaptive dynamic thresholds. *SAE Technical Paper 2025-01-8117*. SAE International. https://doi.org/10.4271/2025-01-8117

[29]. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

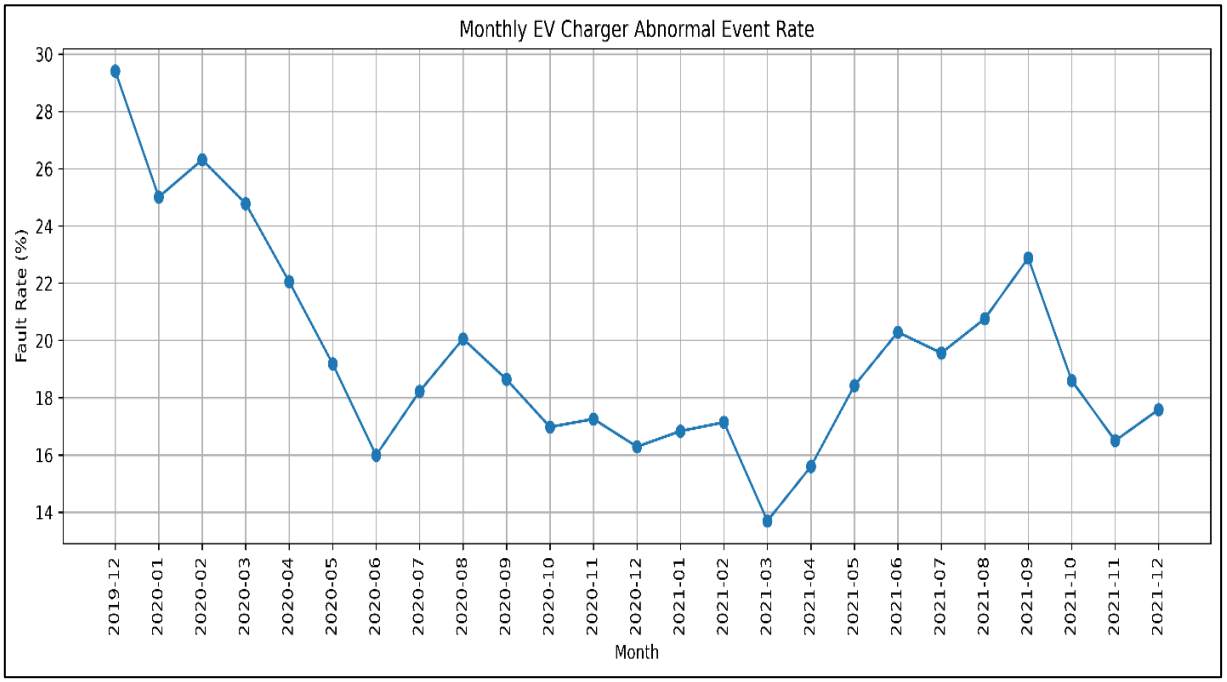
[30]. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering, 150*, 106889. https://doi.org/10.1016/j.cie.2020.106889.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26sep157

APPENDIX



Appendix 1 Calibration of the Locked Model on the Temporal Holdout Set



Appendix 2 Monthly EV Charger Abnormal Event Rate