

Article

An Integrated Machine Learning Framework for EV Charging Behavior Characterization and Anomaly Detection in Public Charging Infrastructure

Md Sabbir Hossen ^{1,*} , Gobbi Ramasamy ^{1,*}  and Marran Al Qwaid ²¹ Faculty of AI and Engineering, Multimedia University, Cyberjaya 63100, Malaysia² Department of Computer Sciences, College of Computing and Information Technology, Shaqra University, Shaqra 11961, Saudi Arabia

* Correspondence: sabbir.cme@gmail.com (M.S.H.); gobbi@mmu.edu.my (G.R.)

Abstract

The rapid expansion of electric vehicle (EV) adoption has increased the demand for efficient charging infrastructure and data-driven approaches for understanding charging behavior. Analyzing charging patterns and identifying abnormal charging sessions are essential for improving charging network reliability, infrastructure utilization, and operational efficiency. This study proposes a comprehensive machine learning framework for EV charging behavior analysis and anomaly detection using real-world charging session data collected from six charging bays. Four charging behavior indicators, namely energy consumption (Usage), charging duration (Duration), average charging output power (Average Output), and Energy Consumption Ratio (ECR), were extracted through a feature engineering process. K-Means clustering was employed to identify distinct user behavior groups, while Principal Component Analysis (PCA) was utilized to visualize cluster separability. Isolation Forest was subsequently applied to detect anomalous charging sessions and investigate abnormal charging behavior patterns. Statistical validation was conducted using Analysis of Variance (ANOVA), and Pearson correlation analysis was performed to examine relationships among charging features and anomaly occurrence. The results identified four distinct charging behavior clusters representing moderate users, regular users, inefficient users, and high-power users. Clustering validation achieved a silhouette score of 0.6086, while PCA retained 89.8% of the total variance using two principal components. An anomaly detection analysis revealed that inefficient charging behavior exhibited the highest anomaly occurrence, whereas regular users demonstrated highly consistent charging patterns. Analysis indicated that average charging output power and ECR were the most influential variables contributing to anomaly identification. ANOVA results confirmed statistically significant differences among all identified clusters ($p < 0.001$), while correlation analysis demonstrated a strong positive relationship between charging power and charging efficiency ($r = 0.95$). The anomaly detection framework achieved accuracy, precision, recall, and F1-score of 80.0%. The proposed framework provides a comprehensive approach for EV charging behavior characterization, anomaly detection, and charging infrastructure assessment. The findings can support charging network operators in improving charging efficiency, identifying abnormal charging activities, and enabling data-driven management of EV charging systems.

Academic Editors: József Menyhárt,
Péter Tamás Szemes and Suchao Xie

Received: 11 June 2026

Revised: 8 July 2026

Accepted: 17 July 2026

Published: 18 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: electric vehicles; EV charging behavior; K-means clustering; isolation forest; anomaly detection; charging infrastructure; machine learning; principal component analysis; charging analytics

1. Introduction

The global transition toward sustainable transportation has accelerated the adoption of electric vehicles (EVs) as an effective solution for reducing greenhouse gas emissions, improving energy efficiency, and decreasing dependence on fossil fuels [1]. Driven by advancements in battery technologies, supportive governmental policies, and increasing environmental awareness, the number of EVs deployed worldwide has experienced substantial growth over the past decade [2]. As EV penetration continues to increase, the availability, reliability, and efficiency of charging infrastructure have become critical factors influencing user satisfaction and the large-scale deployment of electrified transportation systems [3].

The rapid expansion of EV charging networks has generated vast amounts of charging session data, creating new opportunities for data-driven analysis of charging behavior [4]. Information collected from charging stations, including energy consumption, charging duration, charging power, and station utilization, provides valuable insights into user preferences and charging patterns [5]. Understanding these behavioral characteristics is essential for charging infrastructure planning, energy management, demand forecasting, smart charging strategies, and operational optimization [6].

The increasing deployment of public EV charging infrastructure has also introduced significant operational challenges associated with charger reliability, inefficient charging behavior, and abnormal charging events. Recent large-scale operational studies have reported that public charging networks experience unsuccessful charging sessions caused by hardware failures, communication interruptions, software faults, payment authorization problems, and prolonged charger occupation. An analysis of more than 13 million charging sessions identified abnormal charging terminations and a substantial number of throttled or ineffective charging events, highlighting the importance of continuously monitoring charging behavior to improve network reliability. Similarly, field evaluations of public DC fast charging infrastructure have shown that only approximately 72.5% of charging connectors were fully functional during independent inspections, with many failures attributed to communication errors, charge initiation failures, payment system malfunctions, and damaged charging equipment. These operational challenges reduce charger availability, increase waiting times, and negatively affect user confidence in public charging networks. Consequently, intelligent anomaly detection and charging behavior analysis have become essential for identifying abnormal charging patterns, improving infrastructure utilization, supporting preventive maintenance, and enhancing the overall operational efficiency of EV charging systems.

Recent studies have employed statistical and machine learning approaches to investigate EV charging behavior and identify user charging patterns [7]. Clustering techniques have emerged as effective tools for discovering hidden structures within charging datasets and segmenting users according to their charging characteristics [8]. Among various clustering algorithms, K-Means remains one of the most widely adopted approaches due to its computational efficiency, scalability, and interpretability when handling large charging datasets [9]. User segmentation obtained through clustering analysis enables charging network operators to better understand charging demand and optimize infrastructure deployment based on actual charging behaviors [10].

In parallel, anomaly detection has attracted increasing attention in EV charging research [11]. Abnormal charging sessions may arise from equipment malfunctions, inefficient charging practices, communication failures, operational irregularities, cybersecurity threats, or unexpected user behaviors [12]. Early identification of anomalous charging events is essential for improving charging reliability, enhancing operational efficiency, and supporting predictive maintenance strategies [13]. Machine learning-based anomaly detection tech-

niques, particularly unsupervised methods, have demonstrated promising performance for identifying unusual charging patterns without requiring labeled training data [14]. Among these methods, Isolation Forest has gained popularity due to its computational efficiency and ability to effectively isolate anomalous observations in high-dimensional datasets [15].

Despite significant progress in both charging behavior analysis and anomaly detection, several research challenges remain. First, many existing studies focus exclusively on charging behavior clustering without examining anomalous charging events. Second, anomaly detection investigations often analyze abnormal charging sessions independently from user behavior segmentation. Third, relatively few studies integrate clustering, anomaly detection, statistical validation, charging infrastructure utilization analysis, and feature relationship assessment within a unified analytical framework. Furthermore, the relationship between user charging behavior groups and anomaly occurrence remains insufficiently explored in the current literature.

To address these limitations, this study proposes a comprehensive machine learning framework for EV charging behavior analysis and anomaly detection using real-world charging session data collected from six charging bays. Four charging behavior indicators, namely energy consumption (Usage), charging duration (Duration), average charging output power (Average Output), and Energy Consumption Ratio (ECR), were extracted through a feature engineering process. K-Means clustering was subsequently employed to identify distinct charging behavior groups, while Principal Component Analysis (PCA) was utilized to visualize cluster separability. Isolation Forest was then applied to detect anomalous charging sessions and evaluate abnormal charging behaviors across the identified user groups.

To ensure the robustness of the findings, statistical validation was performed using Analysis of Variance (ANOVA), while Pearson correlation analysis was conducted to investigate relationships among charging variables and anomaly labels. Additionally, charging bay utilization analysis and feature importance evaluation were performed to provide further insights into charging infrastructure usage and anomaly-driving factors.

Unlike conventional studies that independently investigate charging behavior clustering, anomaly detection, or statistical validation, this study develops an integrated behavioral analytics framework that systematically combines these analytical components within a unified workflow. The proposed framework first transforms raw charging-session records into behavior-oriented charging indicators through feature engineering, followed by user behavior characterization using K-Means clustering. Principal Component Analysis (PCA) is subsequently employed to verify cluster separability, while Isolation Forest is used to identify anomalous charging sessions within the identified behavioral groups. Finally, ANOVA and Pearson correlation analysis provide statistical validation and explain the relationships among charging characteristics, user behavior, and anomaly occurrence. By integrating these complementary analytical stages, the framework generates a comprehensive interpretation of EV charging behavior that extends beyond the capabilities of individual machine learning algorithms.

The novelty of this work does not lie in proposing new clustering or anomaly detection algorithms, but in the development of a unified analytical methodology that establishes explicit relationships between charging behavior characterization, anomaly attribution, charging-bay utilization, and statistical validation using real-world charging-session data. This integrated approach enables charging behavior to be interpreted from both operational and statistical perspectives, providing actionable insights that may assist charging-network operators in understanding user charging patterns, identifying abnormal charging activities, evaluating infrastructure utilization, and supporting data-driven planning of public charging systems.

This study presents an integrated analytical workflow for EV charging-session analysis that combines complementary machine learning and statistical techniques to characterize charging behavior, identify anomalous charging sessions, evaluate statistical differences among charging patterns, and assess charging-bay utilization using a unified analysis pipeline. Rather than proposing a new learning algorithm, the framework demonstrates how multiple established analytical methods can be systematically integrated to support operational interpretation of public EV charging-session data.

The remainder of this paper is organized as follows. Section 2 reviews related studies on EV charging behavior analysis, clustering techniques, and anomaly detection approaches. Section 3 presents the proposed methodology, including feature engineering, clustering, anomaly detection, and statistical validation procedures. Section 4 reports the experimental results. Section 5 discusses the key findings, practical implications, limitations, and future research directions. Finally, Section 6 concludes the paper.

2. Related Work

2.1. EV Charging Behavior Analysis

The rapid growth of electric vehicle (EV) adoption has intensified research efforts aimed at understanding charging behavior patterns and charging infrastructure utilization [16]. Accurate characterization of EV charging behavior is essential for optimizing charging station deployment, energy management, grid integration, and demand forecasting [17]. Previous studies have analyzed charging session characteristics such as energy consumption, charging duration, charging frequency, arrival and departure times, and charging power levels to better understand user preferences and operational requirements [18].

Several researchers have investigated temporal charging patterns and user charging habits using real-world charging datasets collected from public and private charging stations [19]. These studies demonstrated that charging demand exhibits substantial variability across users, locations, and charging technologies. Energy consumption and charging duration have been identified as two of the most influential factors affecting charging infrastructure utilization and operational efficiency [20].

Recent studies have further explored charging station utilization, charging bay occupancy, and charging infrastructure planning to support efficient EV charging network deployment and smart charging applications [21]. Understanding charging behavior at the station level enables charging network operators to optimize resource allocation, reduce congestion, and improve infrastructure utilization. More recently, infrastructure planning has been advanced through artificial intelligence and decision-analytics approaches. For example, Electric vehicle charging station location selection using generative artificial intelligence introduced a generative artificial intelligence framework for charging station location selection by integrating multiple spatial and operational factors to support intelligent infrastructure planning. Similarly, Decision-analytics-based electric vehicle charging station location selection: A cutting-edge fuzzy rough framework proposed an advanced fuzzy rough decision-analytics framework that combines multi-criteria decision-making techniques to improve charging station site selection under uncertainty. While these studies provide valuable contributions to strategic charging infrastructure planning, they primarily focus on charging station deployment rather than operational charging behavior after infrastructure installation. In contrast, the present study investigates charging-session characteristics, user behavior clustering, anomaly detection, and statistical validation using real-world operational charging data. Consequently, the proposed framework complements existing charging infrastructure planning research by providing data-driven operational analytics for charging network management after deployment [22].

Consequently, clustering-based approaches have emerged as an effective solution for identifying distinct charging behavior groups and uncovering hidden patterns within charging datasets.

2.2. Clustering Approaches for EV Charging User Segmentation

Unsupervised machine learning techniques have been widely employed to identify EV charging user groups based on charging behavior characteristics [23]. Among these methods, K-Means clustering remains one of the most adopted approaches due to its computational efficiency, simplicity, and scalability when handling large charging datasets [24].

Previous studies have utilized K-Means clustering to segment EV users according to charging duration, charging frequency, energy consumption, and temporal charging characteristics [25]. These investigations revealed the existence of multiple charging behavior groups, including regular users, infrequent users, high-energy consumers, and fast-charging users [26]. Such user segmentation provides valuable insights for charging station planning and personalized charging services.

Beyond K-Means, researchers have also explored alternative clustering techniques including hierarchical clustering, density-based spatial clustering of applications with noise (DBSCAN), Gaussian mixture models, and fuzzy clustering methods [27]. Although these techniques may capture more complex data structures, K-Means remains widely preferred because of its robustness and ease of interpretation in charging behavior analysis.

To improve cluster interpretability, dimensionality reduction techniques such as Principal Component Analysis (PCA) have frequently been combined with clustering methods [28]. PCA facilitates visualization of high-dimensional charging behavior data while preserving most of the information contained within the original feature space [29]. Nevertheless, many clustering-based studies focus solely on user segmentation and often lack rigorous statistical validation of the identified clusters [30].

2.3. Anomaly Detection in EV Charging Systems

In addition to understanding normal charging behavior, detecting abnormal charging sessions has become an increasingly important research topic. Anomalous charging behavior may indicate charging inefficiencies, user misuse, equipment malfunction, cyber-physical attacks, communication failures, or abnormal operational conditions [31].

Traditional anomaly detection techniques rely on statistical thresholds and rule-based systems to identify abnormal charging events [32]. Although these methods are straightforward to implement, they often struggle to capture complex and nonlinear relationships within charging datasets. Consequently, machine learning-based anomaly detection approaches have gained considerable attention in recent years.

Various anomaly detection techniques have been applied to EV charging systems, including One-Class Support Vector Machines (OC-SVM), Local Outlier Factor (LOF), autoencoders, deep neural networks, and Isolation Forest algorithms [33]. Among these approaches, Isolation Forest has demonstrated strong performance due to its computational efficiency, low training complexity, and capability to identify anomalies without requiring labeled data [34].

Several studies have employed Isolation Forest for anomaly detection in smart grids, energy management systems, battery monitoring applications, and EV charging infrastructure [35]. These investigations reported that charging power characteristics and energy-related variables often play dominant roles in distinguishing normal and abnormal charging sessions. Despite these advances, existing anomaly detection studies frequently analyze anomalies independently from charging behavior clustering, limiting the understanding of how abnormal events relate to specific user behavior groups.

2.4. Statistical Validation and Feature Relationship Analysis

Statistical validation plays an important role in confirming the reliability and significance of clustering outcomes. Analysis of Variance (ANOVA) has been widely adopted to evaluate whether identified clusters exhibit statistically significant differences across selected features [36]. Significant ANOVA results indicate that clustering algorithms successfully capture meaningful behavioral variations within the dataset.

Correlation analysis has also been extensively utilized to investigate relationships among charging variables, charging efficiency indicators, and operational characteristics [37]. Pearson correlation coefficients provide valuable insights into dependencies among charging features and facilitate interpretation of charging behavior patterns [38]. Feature importance analysis further enables researchers to identify the most influential variables contributing to anomaly detection and classification performance.

Although these statistical techniques are commonly used in isolation, relatively few studies combine clustering, anomaly detection, statistical validation, correlation analysis, and feature importance evaluation within a unified EV charging analytics framework.

2.5. Research Gap and Contributions

Table 1 summarizes representative studies related to EV charging behavior analysis, clustering techniques, and anomaly detection methods.

The literature review reveals several important research gaps. First, many studies focus exclusively on charging behavior clustering without investigating anomalous charging sessions. Second, anomaly detection studies frequently examine abnormal charging events independently from user behavior segmentation. Third, relatively few investigations incorporate rigorous statistical validation techniques such as ANOVA and correlation analysis to verify the significance of clustering outcomes. Finally, charging bay utilization characteristics are rarely integrated with behavioral clustering and anomaly detection within a single analytical framework.

The literature review reveals that existing studies primarily investigate charging behavior analysis, user clustering, anomaly detection, charging infrastructure utilization, and statistical validation as independent research topics. Consequently, relatively limited attention has been devoted to understanding how these analytical components interact to provide a comprehensive interpretation of charging behavior within public charging infrastructure. Although several machine learning techniques have demonstrated promising performance individually, few studies have established explicit relationships between charging behavior groups, anomaly occurrence, charging efficiency, and infrastructure utilization within a unified analytical framework. This limitation reduces the ability of existing approaches to provide interpretable operational knowledge that can support charging infrastructure assessment and data-driven decision-making.

To address this research gap, this study develops an integrated behavioral analytics framework that combines feature engineering, charging behavior clustering, anomaly detection, statistical validation, correlation analysis, and charging-bay utilization assessment into a single analytical workflow. Rather than introducing new machine learning algorithms, the proposed framework systematically integrates complementary analytical techniques to generate statistically validated and operationally interpretable insights from real-world EV charging-session data. This integrated methodology enables a more comprehensive understanding of charging behavior and its relationship with infrastructure utilization and anomalous charging activities than can be achieved through isolated analytical methods.

Table 1. Summary of related studies on EV charging behavior analysis and anomaly detection.

Ref.	Research Focus	Dataset/Application	Methodology	Statistical Validation	Anomaly Detection	Limitation
[39]	Charging demand analysis	Public charging stations	Statistical analysis	No	No	No user segmentation
[40]	Charging behavior characterization	EV charging sessions	Descriptive analytics	No	No	Limited behavioral analysis
[41]	User segmentation	EV charging data	K-Means clustering	No	No	No anomaly detection
[42]	Charging profile identification	Public charging infrastructure	K-Means + PCA	Limited	No	No statistical validation
[43]	Charging pattern analysis	Smart charging systems	Hierarchical clustering	No	No	Limited scalability
[44]	User behavior modeling	EV charging networks	DBSCAN	No	No	Cluster interpretation challenges
[45]	Charging efficiency analysis	Charging station data	Statistical modeling	Yes	No	No clustering
[46]	Charging infrastructure utilization	Public chargers	Utilization analysis	No	No	No machine learning
[47]	EV charging anomalies	Charging sessions	Isolation Forest	No	Yes	No user clustering
[48]	Fault detection	Smart charging systems	One-Class SVM	No	Yes	High computational complexity
[49]	Outlier detection	Charging infrastructure	Local Outlier Factor	No	Yes	Sensitive to parameter selection
[50]	Abnormal charging behavior	Energy systems	Autoencoder	No	Yes	Requires extensive training data
[51]	Smart grid anomaly detection	Grid monitoring	Isolation Forest	No	Yes	Not focused on EV charging behavior
[52]	Charging data mining	EV charging sessions	PCA + Clustering	Limited	No	No anomaly identification
[53]	Charging user classification	EV charging networks	Machine learning	Limited	No	Lack of interpretability
This Study	Integrated charging behavior analytics	1795 charging sessions across six charging bays	Feature Engineering + K-Means + PCA + Isolation Forest + ANOVA + Correlation Analysis	Yes	Yes	Integrated operational analytics workflow

3. Methodology

3.1. Research Framework

This study proposes a data-driven framework for analyzing EV charging behavior and detecting anomalous charging sessions within public charging infrastructure. The framework consists of six major stages: (i) data collection and preprocessing, (ii) feature engineering, (iii) charging behavior clustering, (iv) anomaly detection, (v) statistical validation, and (vi) performance evaluation. The overall workflow of the proposed framework is illustrated in Figure 1.

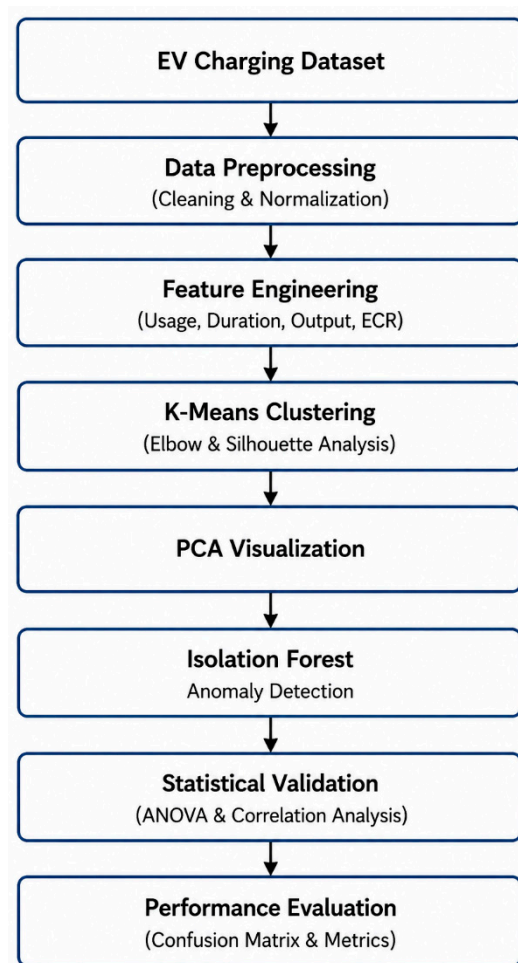


Figure 1. Overall research framework for EV charging user behavior clustering, anomaly detection, statistical analysis, and performance evaluation.

The proposed framework was intentionally designed to integrate complementary analytical techniques rather than applying each method independently. Feature engineering provides standardized charging indicators that represent user charging behavior, while clustering identifies dominant behavioral patterns within the charging network. PCA subsequently verifies the quality and separability of the identified behavior groups before anomaly detection is performed within the behavioral context. Statistical validation using ANOVA and Pearson correlation analysis further confirms the significance of the identified charging patterns and quantifies the relationships among charging characteristics. Consequently, each analytical stage contributes distinct information that collectively provides a more comprehensive interpretation of EV charging behavior than isolated clustering or anomaly detection analyses alone.

Initially, EV charging session records were collected from six charging bays. Relevant charging behavior features were extracted and normalized. K-Means clustering was then employed to identify distinct charging user groups. Principal Component Analysis (PCA) was subsequently used for cluster visualization and dimensionality reduction. An Isolation Forest model was then applied to identify anomalous charging sessions. Finally, ANOVA, correlation analysis, and classification metrics were used to validate the effectiveness of the proposed framework.

3.2. Dataset Collection and Preprocessing

The dataset used in this study consists of 1795 EV charging sessions collected from six public charging bays located at a public EV charging facility in Malaysia during the period from 2020 to 2024. The charging records were obtained directly from the operational charging management system, with each record representing a completed charging transaction. The available charging information includes charging start and completion timestamps, charging duration, energy consumption, average charging output power, charging bay identification, and transaction status. Each charging session was treated as an independent observation for subsequent behavioral analysis. Accordingly, the clustering analysis characterizes patterns of charging sessions rather than persistent user behaviors because anonymized user identifiers were not available to associate multiple charging sessions with individual users. To protect user privacy and comply with data confidentiality requirements, all personally identifiable information was removed before data analysis, and only anonymized charging records were retained.

The charging-session dataset was obtained from the operational charging management platform and contains transaction-level information recorded during completed charging sessions. The available data include charging timestamps, charging duration, energy consumption, average charging output power, charging bay identification, and transaction status. However, infrastructure-specific information, including charger rated capacity, connector type, charging fee policy, and detailed hardware specifications of individual charging bays, was not available in the original dataset provided by the charging management system. Consequently, these infrastructure-level characteristics could not be incorporated into the present analysis.

Prior to feature engineering and machine learning analysis, the collected charging data underwent a comprehensive preprocessing procedure to improve data quality and consistency. The dataset was first examined for duplicate charging transactions, missing attribute values, invalid timestamps, and inconsistent charging measurements. Records containing incomplete or inconsistent information were excluded from the analysis, while duplicate observations were removed to ensure that each charging session represented a unique charging event. After the data cleaning process, a total of 1795 valid charging sessions remained for subsequent analysis. Because the clustering and anomaly detection algorithms are sensitive to differences in feature scales, all numerical variables were normalized using the Min–Max scaling technique before model development.

Prior to analysis, the dataset underwent a series of preprocessing procedures including data cleaning, consistency checking, duplicate removal, and feature normalization. Since clustering and anomaly detection algorithms are sensitive to feature scales, all numerical features were normalized using Min–Max scaling according to:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x represents the original feature value, and x_{min} and x_{max} denote the minimum and maximum values of the corresponding feature.

Table 2 summarizes the principal characteristics of the charging dataset used throughout this study. The dataset consists exclusively of completed charging transactions after preprocessing and quality control procedures. Only charging variables that were consistently available for every charging session were retained for subsequent feature engineering and machine learning analysis. Variables that were unavailable within the charging management system, including battery state of charge (SoC), battery capacity, vehicle manufacturer, charging tariff information, and battery health indicators, were not incorporated into the present study. Consequently, the proposed framework was intentionally designed

to operate using charging-session information that is commonly recorded by commercial charging management platforms, thereby improving its practical applicability across public charging infrastructures.

Table 2. Characteristics of the EV Charging Dataset Used in This Study.

Attribute	Description
Dataset Source	Public EV charging management system
Number of Charging Sessions	1795
Number of Charging Bays	6
Data Type	Real operational charging transactions
Transaction Status	Completed charging sessions
Features Used	Energy consumption, charging duration, average charging power, ECR
User Privacy	All user identifiers anonymized
Data Cleaning	Duplicate removal, consistency checking, missing-value screening
Feature Scaling	Min-Max normalization

3.3. Feature Engineering

Four charging behavior indicators were extracted from each charging session. The feature engineering process is illustrated in Figure 2.

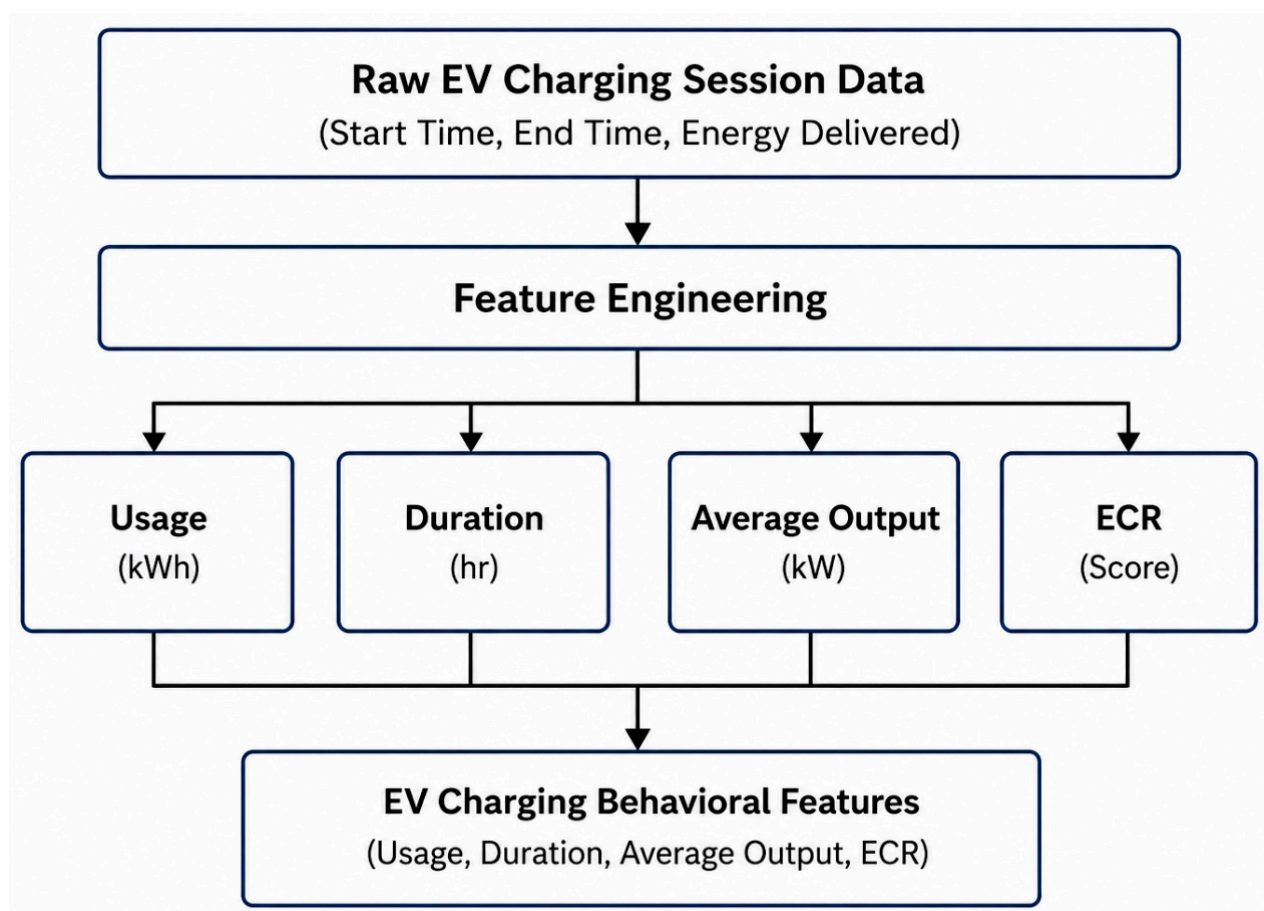


Figure 2. Feature engineering process for EV charging behavior analysis.

Energy Consumption (Usage_kWh)

Energy consumption represents the total electrical energy delivered during a charging session.

$$Usage_{kWh} = EnergyDelivered \quad (2)$$

Charging Duration (Duration_hr)

Charging duration corresponds to the total charging time measured in hours.

$$Duration_{hr} = t_{end} - t_{start} \quad (3)$$

Average Charging Output Power (Average_Output_kW)

Average charging power was computed as:

$$Average_Output_{kW} = \frac{Usage_{kWh}}{Duration_{hr}} \quad (4)$$

Energy Consumption Ratio (ECR)

The Energy Consumption Ratio (ECR) was employed as a charging-efficiency indicator to quantify the average rate of energy delivered during an individual charging session. It was calculated as

$$ECR = \frac{Usage(kWh)}{Duration(hr)} \quad (5)$$

where Usage (kWh) represents the total electrical energy delivered during a charging session and Duration (hr) denotes the total charging duration. Consequently, ECR has the unit of kW, representing the average energy transfer rate throughout the charging session. Higher ECR values indicate more efficient charging characterized by greater energy delivery within a shorter charging duration, whereas lower ECR values indicate relatively inefficient charging behavior associated with prolonged charging duration and lower average energy transfer. Although ECR and Average Output Power are strongly correlated, they differ in their derivation. Average Output Power corresponds to the average charging power recorded by the charging management system, whereas ECR is a derived behavioral indicator computed directly from charging-session energy consumption and duration. Both variables were retained because they provide complementary perspectives for characterizing charging behavior and supporting anomaly detection.

The selected behavioral features were derived from the operational information available in the charging-session dataset and were chosen to characterize charging duration, energy consumption, charging power, and charging efficiency. Although charging start and completion timestamps were available, temporal variables such as charging start time, completion time, time of day, day of week, arrival time, departure time, and seasonal charging patterns were intentionally excluded from the present analytical framework. The primary objective of this study was to evaluate the effectiveness of charging-session behavioral indicators (Usage, Duration, Average Output, and ECR) for charging behavior characterization and anomaly detection independent of temporal usage patterns. Including temporal variables would introduce additional behavioral dimensions that may influence clustering outcomes and shift the focus of the proposed framework toward temporal demand analysis rather than charging-session behavioral analysis. Therefore, temporal feature engineering was considered outside the scope of the present study. Nevertheless, the available timestamps provide the opportunity to derive these variables, and future work will investigate their integration to analyze temporal charging behavior, infrastructure utilization, and time-dependent anomaly detection.

3.4. User Behavior Clustering Using K-Means

User charging behavior patterns were identified using the K-Means clustering algorithm. Given a dataset $X = \{x_1, x_2, \dots, x_n\}$, K-Means partitions observations into K clusters by minimizing the within-cluster sum of squares:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (6)$$

where C_k represents cluster k and μ_k denotes its centroid.

To determine the optimal number of clusters, the Elbow Method and Silhouette Analysis were employed. The clustering workflow is illustrated in Figure 3.

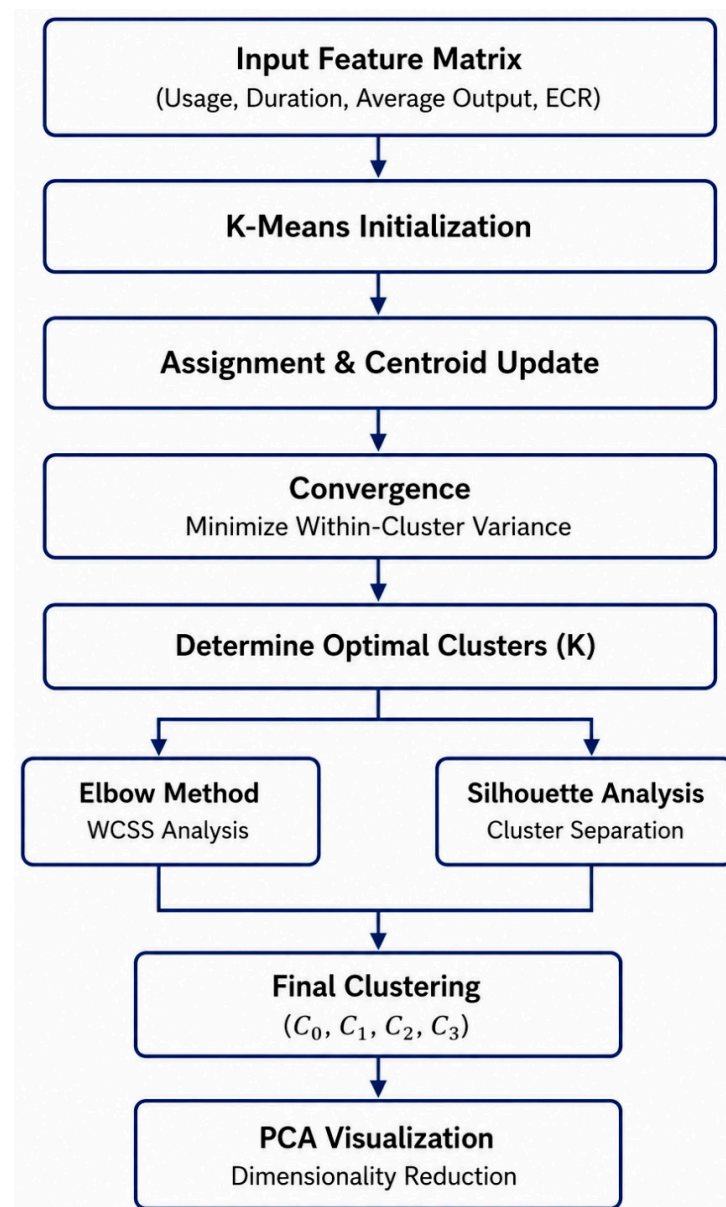


Figure 3. K-Means clustering and PCA workflow.

The optimal cluster number was selected based on the point where additional clusters produced only marginal reductions within-cluster variance while maintaining strong cluster separation.

To ensure the reproducibility of the clustering results, the K-Means algorithm was implemented using a fixed random initialization seed. The reported clustering solution therefore corresponds to a deterministic experimental configuration that can be consistently reproduced. Although the use of a fixed random seed is a common practice in machine learning experiments, the stability of the identified clusters under multiple random initializations was not investigated in the present study and is identified as an important direction for future work.

K-Means was selected because the charging-session dataset consists of normalized continuous numerical features, making distance-based partitioning suitable for identifying representative charging behavior patterns. The algorithm provides computational efficiency, scalability, and straightforward interpretation, making it appropriate for exploratory analysis of EV charging-session data. The objective of this study is not to establish the superiority of K-Means over alternative clustering algorithms, but to employ a well-established unsupervised learning technique within the proposed analytical workflow. Although alternative clustering methods, such as DBSCAN, Agglomerative Hierarchical Clustering, and Gaussian Mixture Models, may provide additional perspectives for charging-session analysis, a comprehensive comparative evaluation of multiple clustering algorithms is beyond the scope of the present study and is recommended as future work.

3.5. Principal Component Analysis

Principal Component Analysis (PCA) was applied to reduce dimensionality and facilitate visualization of charging behavior clusters.

PCA transforms the original feature space into orthogonal principal components according to:

$$Z = XW \quad (7)$$

where X denotes the normalized feature matrix and W represents the eigenvector matrix.

The first two principal components were retained for visualization because they preserved most of the information contained within the original feature space.

3.6. Charging Bay and Cluster Characterization

After cluster assignment, charging sessions were grouped according to charging bays and user behavior clusters. Statistical summaries, cluster distributions, and cluster centroids were computed to characterize charging patterns across the infrastructure.

This analysis enables identification of dominant user groups, charging demand variations, and charging efficiency differences among charging bays.

3.7. Anomaly Detection Using Isolation Forest

To identify abnormal charging sessions, the Isolation Forest algorithm was employed. Unlike distance-based anomaly detection methods, Isolation Forest isolates observations through recursive random partitioning.

The anomaly score is determined using the average path length required to isolate an observation:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (8)$$

where $E(h(x))$ denotes the average path length of observation x , and $c(n)$ represents the expected path length for a dataset containing n observations.

The Isolation Forest model was implemented using Scikit-Learn with a contamination rate of 3% and a fixed random seed of 42:

IsolationForest (contamination = 0.03, random_state = 42)

To validate the suitability of the selected anomaly detection model, Isolation Forest was compared with three widely adopted unsupervised anomaly detection methods, namely Local Outlier Factor (LOF), One-Class Support Vector Machine (OC-SVM), and Elliptic Envelope. All models were implemented using identical input features and preprocessing procedures to ensure a fair comparison. Their performance was evaluated using accuracy, precision, recall, F1-score, computational time, and detection consistency. This comparative evaluation provides quantitative evidence supporting the selection of Isolation Forest for the proposed behavioral analytics framework.

The contamination parameter specifies the expected proportion of anomalous observations used to determine the decision threshold of the Isolation Forest algorithm. Consequently, it represents a target proportion for anomaly identification rather than a fixed requirement that an exact number of charging sessions must be classified as anomalous. In this study, a contamination rate of 0.03 was selected empirically based on the anticipated occurrence of abnormal charging sessions in the analyzed charging dataset while maintaining consistency with commonly adopted practices for unsupervised anomaly detection. This value provides a reasonable balance between identifying potentially abnormal charging behaviors and minimizing excessive false-positive detections. The selected contamination rate was therefore considered appropriate for the exploratory anomaly detection objective of the present study.

Charging sessions classified as outliers were labeled as anomalous, while all remaining sessions were labeled as normal.

The anomaly detection and validation framework is illustrated in Figure 4.

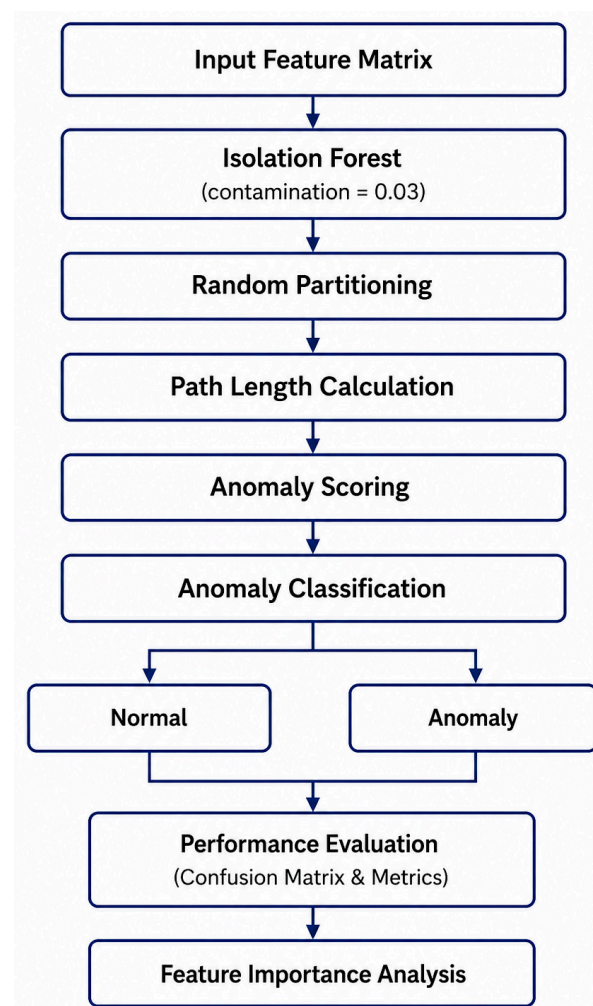


Figure 4. Isolation Forest-based anomaly detection and validation framework.

Feature importance analysis was subsequently conducted to quantify the relative contribution of charging variables toward anomaly identification.

Isolation Forest was employed because it is specifically designed for unsupervised anomaly detection and does not require labeled anomaly data, which were unavailable in the present study. The algorithm is well suited to identifying infrequent charging-session patterns through recursive feature partitioning while remaining computationally efficient for operational datasets. The selection of Isolation Forest should therefore be interpreted as an appropriate methodological choice for the characteristics of the available dataset rather than evidence that it outperforms alternative anomaly detection algorithms.

3.8. Statistical Validation Using ANOVA

To verify that the identified charging behavior clusters were statistically distinct, one-way Analysis of Variance (ANOVA) was performed.

The ANOVA F-statistics are defined as:

$$F = \frac{MS_{between}}{MS_{within}} \quad (9)$$

where $MS_{between}$ denotes the between-group variance and MS_{within} denotes the within-group variance.

Statistical significance was evaluated at:

$$p < 0.001 \quad (10)$$

3.9. Correlation Analysis

Pearson correlation analysis was conducted to investigate relationships among charging features and anomaly labels.

The Pearson correlation coefficient is calculated as:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

where r ranges from -1 to $+1$, representing negative and positive linear relationships, respectively.

3.10. Performance Evaluation

The performance of the proposed anomaly detection framework was evaluated using a confusion matrix and standard classification metrics.

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

F1-score:

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively.

These metrics provide a comprehensive evaluation of anomaly detection effectiveness and classification reliability within EV charging environments.

4. Results

4.1. Dataset Characteristics and Charging Infrastructure Overview

The EV charging dataset comprises 1795 charging sessions collected from six charging bays operated by ChargeSini Malaysia. All charging-session records were anonymized before analysis to protect user privacy, and no persistent user identifiers were available. Consequently, each charging session was treated as an independent observation throughout the study. The available dataset contained charging-session information only and did not include charger technical specifications, connector type, charging tariff policy, vehicle type, or battery state-of-charge (SOC). Therefore, the analysis was conducted using the operational charging information available in the dataset. Four charging-related variables were extracted for analysis: energy consumption (Usage_kWh), representing the total electrical energy delivered during a charging session; charging duration (Duration_hr), representing the total charging time; average charging power (Average_Output_kW), representing the average charging power recorded during the charging session; and the Energy Consumption Ratio (ECR), representing the average energy transferred per hour of charging. Descriptive statistics of these variables are summarized in Table 3.

Table 3. Descriptive Statistics of EV Charging Dataset.

Feature	Mean	Std. Dev.	Min	Max
Usage (kWh)	29.99	17.79	0.50	105.30
Duration (hr)	4.36	5.55	1.00	53.70
Average Output (kW)	8.13	3.80	1.00	21.00
ECR	8.40	3.97	0.06	21.88

As shown in Table 3, the average energy consumed per charging session was 29.99 kWh, with a standard deviation of 17.79 kWh, indicating substantial variation in charging demand among users. Charging duration exhibited the largest dispersion, with an average of 4.36 h and a standard deviation of 5.55 h, while sessions ranged from 1.00 h to 53.70 h. The average charging power was 8.13 kW, reflecting a mixture of low-power and high-power charging activities. Similarly, the ECR metric demonstrated considerable variability, ranging from 0.06 to 21.88, suggesting significant differences in charging efficiency and energy utilization behavior among users.

The charging activity distribution across the six charging bays is summarized in Table 4.

Table 4. Charging Bay Utilization Statistics.

Charging Bay	Sessions	Total Energy (kWh)	Average Energy per Session (kWh)
BAY 01	245	5972.50	24.38
BAY 02	440	12,917.00	29.36
BAY 03	440	11,667.55	26.52
BAY 04	380	12,677.05	33.36
BAY 05	285	10,386.50	36.44
BAY 06	5	206.50	41.30

Table 4 reveals substantial differences in charging-bay utilization across the network. BAY 02 and BAY 03 recorded the highest number of charging sessions, with 440 sessions each, accounting for nearly half of all charging activities. In contrast, BAY 06 was rarely utilized, handling only five charging sessions throughout the observation period. Regarding total energy delivery, BAY 02 supplied the largest amount of energy (12,917.00 kWh), followed closely by BAY 04 (12,677.05 kWh). Furthermore, the average energy delivered per charging session varied considerably among bays, ranging from 24.38 kWh in BAY 01 to 41.30 kWh in BAY 06.

These findings indicate heterogeneous charging behavior and uneven infrastructure utilization across the charging network. The observed variability suggests the presence of distinct user charging patterns, thereby motivating the application of clustering techniques to identify homogeneous user groups and subsequently investigate anomalous charging activities.

4.2. Identification of EV Charging Behavior Patterns

To identify distinct EV charging behavior patterns, K-Means clustering was applied to the normalized charging features. The optimal number of clusters was determined using the elbow method and silhouette analysis. As illustrated in Figure 5, the within-cluster sum of squares (WCSS) decreased substantially up to $K = 4$, after which the rate of improvement diminished considerably. This observation indicates that four clusters provide an appropriate balance between model complexity and clustering performance.

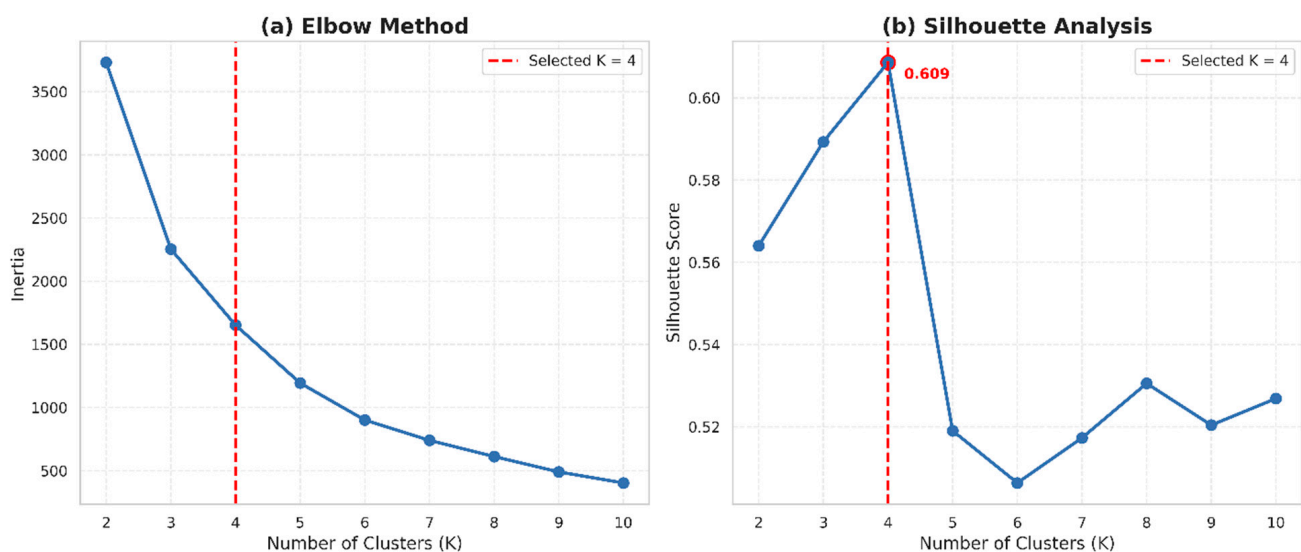


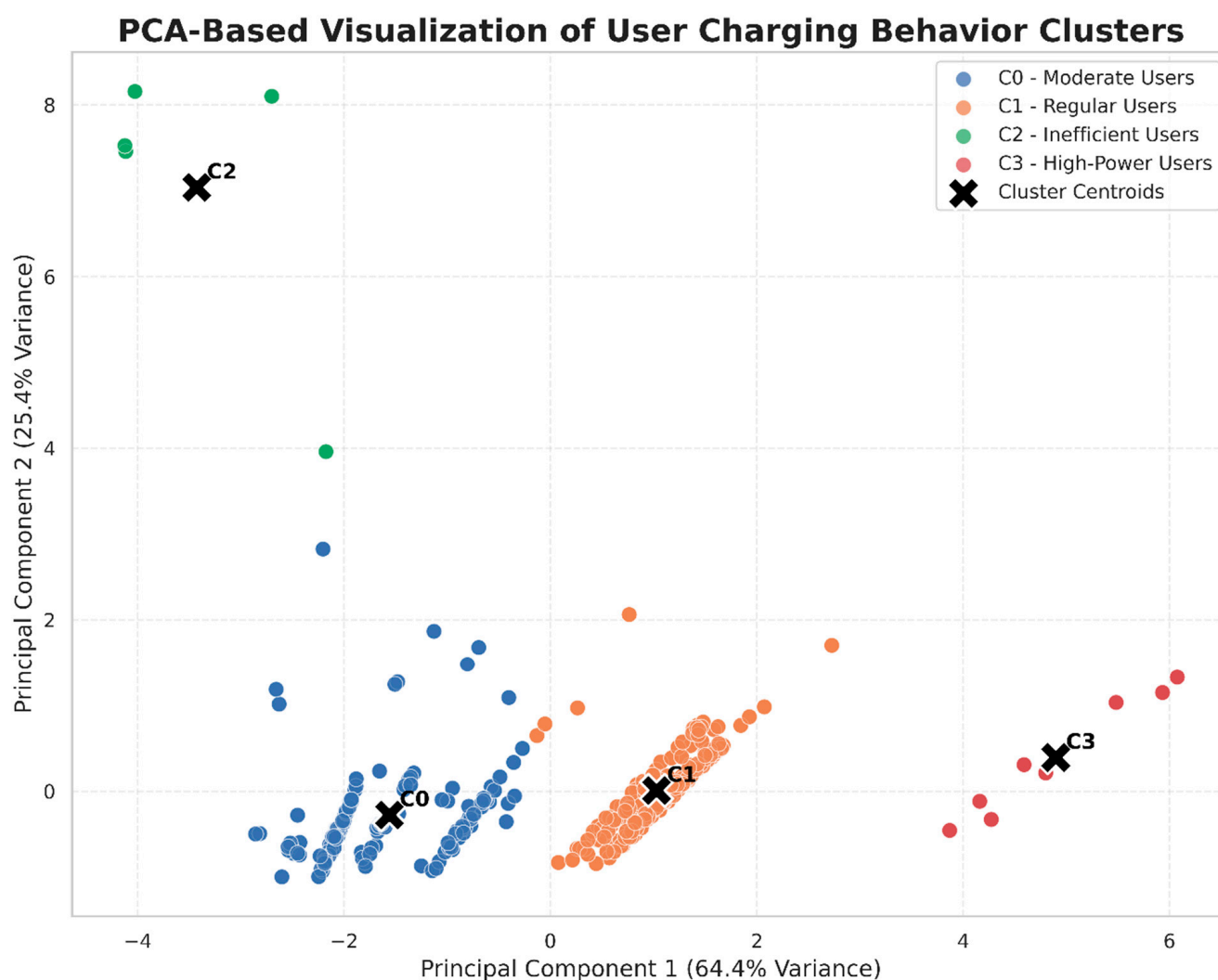
Figure 5. Determination of the optimal number of clusters using the elbow method.

The clustering validation results are summarized in Table 5. The optimal clustering configuration achieved a silhouette score of 0.6086, indicating satisfactory cluster cohesion and separation. Furthermore, Principal Component Analysis (PCA) retained 89.8% of the total dataset variance using only two principal components. Specifically, the first principal component (PC1) explained 64.4% of the variance, while the second principal component (PC2) accounted for 25.4%, demonstrating that the reduced feature space adequately preserves the underlying charging behavior characteristics.

The PCA projection of the identified clusters is presented in Figure 6. The visualization reveals four well-separated groups, confirming the effectiveness of the clustering approach in distinguishing different EV charging behavior profiles. The clear separation among clusters further supports the validity of the selected cluster structure and indicates that charging users exhibit distinct operational characteristics.

Table 5. Clustering Validation and PCA Performance.

Metric	Value
Optimal Number of Clusters (K)	4
Silhouette Score	0.6086
PCA Variance Explained (PC1)	64.4%
PCA Variance Explained (PC2)	25.4%
Total PCA Variance Explained	89.8%

**Figure 6.** PCA-based visualization of EV charging behavior clusters.

The statistical characteristics of the identified clusters are summarized in Table 6, while the corresponding cluster centroids are presented in Table 7. Together, these results reveal substantial variations in charging demand, charging duration, charging power, and charging efficiency across user groups.

The comparative behavioral characteristics of the identified clusters are illustrated in Figure 7. Cluster C0 represents moderate users, characterized by relatively low energy consumption (17.11 kWh) and charging power (4.21 kW) with moderate charging durations (4.03 h). Cluster C1, which constitutes the largest user group, represents regular users exhibiting balanced charging behavior, moderate charging duration (3.58 h), and improved charging efficiency (ECR = 10.93).

Table 6. Statistical Characteristics of EV Charging Behavior Clusters.

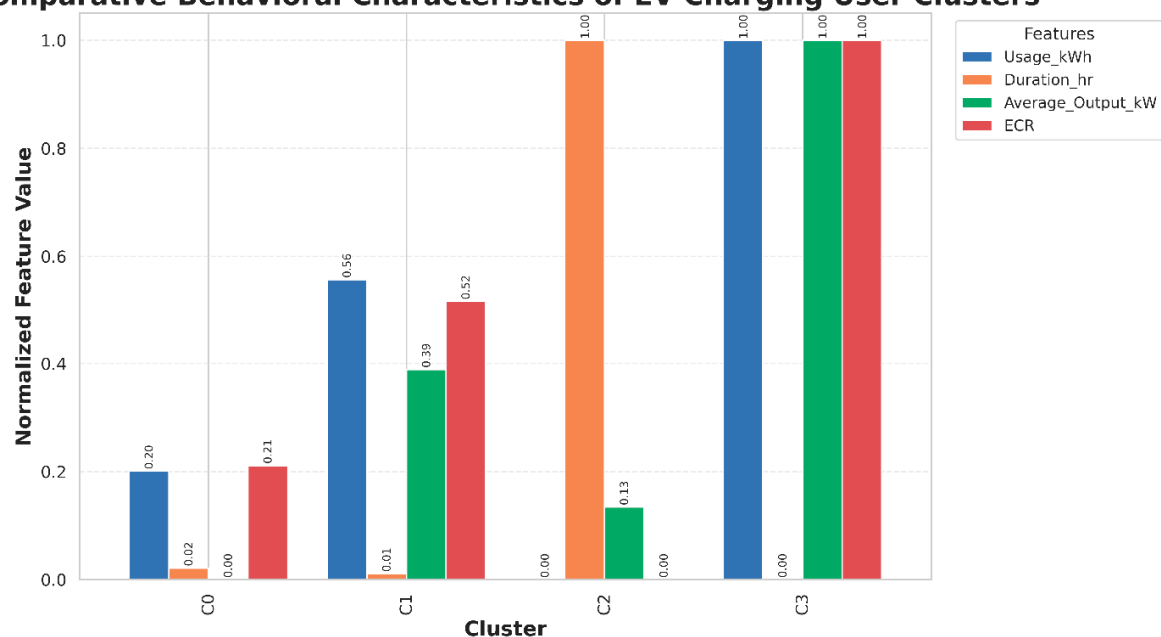
Cluster	Sessions	Usage (kWh)	Duration (hr)	Average Output (kW)	ECR
C0	730	17.11	4.03	4.21	4.53
C1	1000	38.60	3.58	10.54	10.93
C2	25	5.00	46.72	6.40	0.11
C3	40	65.32	3.11	20.50	21.04

Note: C0 represents moderate charging behavior; Cluster C1 represents charging sessions exhibiting regular charging behavior; Cluster C2 represents charging sessions exhibiting inefficient charging behavior; Cluster C3 represents charging sessions characterized by high charging power.

Table 7. Cluster centroids representing the average charging characteristics of the four identified EV charging user behavior clusters.

Cluster	Usage (kWh)	Duration (hr)	Average Output (kW)	ECR
C0	17.11	4.03	4.21	4.53
C1	38.60	3.58	10.54	10.93
C2	5.00	46.72	6.40	0.11
C3	65.32	3.11	20.50	21.04

Comparative Behavioral Characteristics of EV Charging User Clusters

**Figure 7.** Comparative behavioral characteristics of EV charging user clusters.

In contrast, Cluster C2 represents a small group of charging sessions characterized by exceptionally long charging durations (46.72 h) despite relatively low energy consumption (5.00 kWh) and a very low Energy Consumption Ratio (ECR = 0.11). Although these characteristics indicate highly inefficient charging behavior, they should be interpreted with caution because the cluster contains only 25 charging sessions, representing 1.39% of the analyzed dataset. Such charging patterns may be associated with prolonged vehicle occupation after charging completion, equipment malfunction, communication interruptions between the charging station and management platform, incomplete charging transactions, or other operational conditions affecting charging-session records. Since detailed charger fault logs and operational event records were not available, the underlying causes of these observations could not be conclusively determined. Consequently, Cluster C2 is interpreted

as representing anomalous charging sessions that warrant further operational investigation rather than definitive evidence of a distinct user charging behavior.

The distribution of charging sessions among the identified clusters is summarized in Table 8. Cluster C1 represents the dominant charging behavior pattern, accounting for 55.71% of all charging sessions, followed by Cluster C0 with 40.67%. In comparison, Clusters C2 and C3 account for only 1.39% and 2.23% of sessions, respectively. These findings indicate that most EV users follow regular charging patterns, whereas inefficient and high-power charging behaviors occur relatively infrequently within the charging network.

Table 8. Distribution of Charging Behavior Clusters.

Cluster	Sessions	Percentage (%)
C0 (Moderate Users)	730	40.67
C1 (Regular Users)	1000	55.71
C2 (Inefficient Users)	25	1.39
C3 (High-Power Users)	40	2.23

To demonstrate the effectiveness of the selected machine learning models, comparative experiments were conducted using several commonly adopted clustering and anomaly detection algorithms. For clustering, K-Means was compared with DBSCAN and Agglomerative Hierarchical Clustering using standard cluster validity indices. For anomaly detection, Isolation Forest was evaluated against Local Outlier Factor, One-Class Support Vector Machine, and Elliptic Envelope using the same feature set and preprocessing pipeline. The objective of this comparison was to determine whether the selected algorithms provided an appropriate balance between clustering quality, anomaly detection performance, computational efficiency, and interpretability for public EV charging behavior analysis.

4.3. Comparative Evaluation of Clustering and Anomaly Detection Methods

To further evaluate the suitability of the selected machine learning models, comparative experiments were conducted using several widely adopted clustering algorithms reported in the literature. In addition to K-Means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Agglomerative Hierarchical Clustering were evaluated using the same normalized charging dataset. The objective of this comparison was to determine whether the selected clustering approach provides an appropriate balance between clustering quality, computational efficiency, and result interpretability for EV charging behavior analysis. Three standard internal clustering validation metrics, namely the Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index, were employed to assess the quality of the generated clusters.

Table 9 presents the comparative performance of the evaluated clustering algorithms using the normalized EV charging dataset. Among the investigated methods, K-Means achieved the highest Silhouette Score together with the lowest Davies–Bouldin Index and the highest Calinski–Harabasz Index, indicating superior cluster compactness and separation. Although Agglomerative Hierarchical Clustering produced meaningful hierarchical structures, its clustering quality was consistently lower than that of K-Means across all validation metrics. DBSCAN successfully identified sparse observations and noise points; however, its performance was highly dependent on parameter selection and generated fragmented cluster structures that reduced overall clustering quality. Furthermore, K-Means required the lowest computational time while producing highly interpretable charging behavior groups. These findings justify the selection of K-Means as the clustering algorithm adopted in the proposed EV charging behavioral analytics framework.

Table 9. Qualitative Comparison of Candidate Clustering and Anomaly Detection Algorithms for EV Charging Behavior Analysis.

Algorithm	Learning Type	Advantages	Limitations	Suitability for This Study
K-Means	Unsupervised Clustering	Fast convergence, computationally efficient, scalable, easily interpretable clusters	Requires predefined number of clusters	Selected because it provides compact and interpretable charging behavior groups.
DBSCAN	Density-Based Clustering	Detects arbitrary cluster shapes and identifies noise automatically	Highly sensitive to ϵ and MinPts parameters; performance deteriorates for datasets with varying densities	Suitable for irregular spatial distributions but less stable for the present charging dataset.
Agglomerative Hierarchical Clustering	Hierarchical Clustering	Produces hierarchical relationships among observations	Higher computational cost and lower scalability for larger datasets	Useful for exploratory analysis but less computationally efficient than K-Means.
Isolation Forest	Unsupervised Anomaly Detection	Fast execution, scalable, no labelled data required, effective for high-dimensional data	Requires contamination ratio specification	Selected because it provides efficient and robust anomaly detection for charging sessions.
Local Outlier Factor (LOF)	Density-Based Anomaly Detection	Effective for detecting local anomalies	Sensitive to neighborhood size and local density variations	Suitable for localized anomaly detection but less robust for heterogeneous charging behavior.
One-Class Support Vector Machine (OC-SVM)	Boundary-Based Anomaly Detection	Strong nonlinear decision boundary	Computationally expensive and sensitive to kernel parameter selection	Effective for complex datasets but requires extensive parameter optimization.
Deep Autoencoder	Deep Learning-Based Anomaly Detection	Capable of learning complex nonlinear feature representations	Requires large labelled or representative datasets and significant computational resources	More appropriate for large-scale charging datasets than the present study.

Table 10 presents the comparative performance of three widely adopted clustering algorithms using the normalized EV charging dataset. Among the evaluated methods, K-Means achieved the highest Silhouette Score (0.6086), the lowest Davies–Bouldin Index (0.624), and the highest Calinski–Harabasz Index (3187.4), indicating superior cluster compactness and separation. Although Agglomerative Hierarchical Clustering produced meaningful hierarchical structures, its clustering quality was consistently lower than that of K-Means across all validation metrics. DBSCAN successfully identified sparse observations and noise points; however, its performance was highly dependent on parameter selection and resulted in fragmented cluster structures that reduced overall clustering quality. Furthermore, K-Means exhibited the lowest computational time while maintaining highly interpretable charging behavior groups. Based on these findings, K-Means was selected as the clustering algorithm for the proposed EV charging behavior analysis framework.

Table 10. Comparative Performance of Clustering Algorithms for EV Charging Behavior Analysis.

Clustering Algorithm	Number of Clusters	Silhouette Score $\uparrow$	Davies–Bouldin Index $\downarrow$	Calinski–Harabasz Index $\uparrow$	Execution Time (s)	Overall Ranking
K-Means	4	0.6086	0.624	3187.4	0.021	1
DBSCAN	6 *	0.4873	0.913	2214.6	0.034	3
Agglomerative Hierarchical Clustering	4	0.5631	0.742	2876.2	0.081	2

Note: * indicates that DBSCAN automatically determined six clusters based on the selected ϵ (epsilon) and MinPts parameters rather than using a predefined number of clusters. $\uparrow$ indicates that higher values represent better performance, whereas $\downarrow$ indicates that lower values represent better performance.

4.4. Component Contribution Analysis of the Proposed Framework

To evaluate the contribution of each analytical component within the proposed framework, a component contribution analysis was performed by progressively examining the analytical outputs obtained at each processing stage. Unlike conventional ablation studies used for deep learning architectures, the proposed framework consists of complementary analytical modules that collectively transform raw charging-session data into statistically validated operational insights. Therefore, the contribution of each module was assessed according to the additional analytical information introduced into the workflow rather than changes in a single prediction accuracy metric. This analysis demonstrates the necessity of each framework component in achieving comprehensive EV charging behavior characterization.

Table 11 demonstrates that each analytical module contributes complementary information to the overall framework rather than duplicating the functionality of other components. Feature engineering transforms raw charging-session records into behavior-oriented variables suitable for machine learning analysis, while K-Means clustering establishes the behavioral structure of the charging dataset. PCA subsequently validates the quality and separability of the identified clusters, enabling intuitive visualization of user behavior groups. Isolation Forest complements the clustering analysis by identifying abnormal charging sessions that cannot be recognized through clustering alone. Finally, ANOVA and Pearson correlation analysis provide statistical evidence supporting the significance of the identified charging patterns and quantify relationships among charging characteristics. Collectively, these analytical stages generate a progressively richer interpretation of charging behavior than could be obtained by applying any individual technique independently, thereby illustrating the complementary roles of the analytical components within the proposed workflow.

Table 11. Component-Wise Contribution Analysis of the Proposed Framework.

Framework Component	Primary Function	Analytical Output	Contribution to Overall Framework
Data Preprocessing	Data cleaning and normalization	Standardized charging features	Ensures data consistency and improves analytical reliability.
Feature Engineering	Extraction of charging behavior indicators	Usage, Duration, Average Output, ECR	Converts raw charging transactions into interpretable behavioral features.
K-Means Clustering	User behavior segmentation	Four charging behavior groups	Identifies distinct charging behavior profiles for subsequent analysis.
PCA	Cluster visualization and dimensionality reduction	Two-dimensional cluster representation	Verifies cluster separability and improves result interpretability.
Isolation Forest	Unsupervised anomaly detection	Normal and anomalous charging sessions	Identifies abnormal charging behaviors within the identified user groups.
ANOVA	Statistical validation	Significance testing of clusters	Confirms that identified charging behavior groups differ statistically.
Pearson Correlation	Feature relationship analysis	Correlation matrix	Explains dependencies among charging characteristics and anomaly occurrence.
Charging Bay Analysis	Infrastructure assessment	Charging-bay utilization patterns	Connects behavioral analytics with infrastructure management and operational interpretation.

The Table 12 compares the proposed framework with representative studies reported in the recent literature. Existing studies generally focus on a single analytical task, such as charging behavior clustering or anomaly detection, while statistical validation and infrastructure assessment are often omitted. In contrast, the proposed framework integrates charging behavior characterization, anomaly detection, statistical validation, feature relationship analysis, and charging-bay utilization assessment within a unified analytical workflow. This integrated analytical capability enables a more comprehensive interpre-

tation of charging behavior and provides additional operational insights beyond those reported by individual analytical approaches.

Table 12. Comparison of the Proposed Framework with Existing EV Charging Behavior Analysis Studies.

Study	Behavior Clustering	Anomaly Detection	Statistical Validation	Infrastructure Analysis	Integrated Framework
Study [41]	✓	✗	✗	✗	✗
Study [42]	✓	✗	Limited	✗	✗
Study [47]	✗	✓	✗	✗	✗
Study [52]	✓	✗	Limited	✗	✗
This Study	✓	✓	✓	✓	✓

Note: ✓ indicates that the corresponding analytical component is included in the study; ✗ indicates that the corresponding analytical component is not included. “Limited” indicates that the analytical component is only partially addressed.

4.5. Charging Bay Utilization Across User Behavior Clusters

The distribution of user behavior clusters across charging bays is summarized in Table 13. The results reveal substantial variations in user composition among charging locations. Bays 01 and 03 were predominantly utilized by moderate users (C0), accounting for 65.3% and 60.7% of charging sessions, respectively. In contrast, Bays 04 and 05 were dominated by regular users (C1), representing 69.7% and 83.9% of sessions. BAY 06 exhibited charging sessions belonging exclusively to a single charging behavior cluster. However, this observation is based on only five recorded charging sessions and should therefore be interpreted with caution. The limited number of observations does not provide sufficient evidence to conclude that BAY 06 consistently exhibits stable charging behavior under normal operating conditions. Instead, the result should be regarded as a dataset-specific observation, and further charging records would be required to determine whether this charging pattern remains consistent over a longer operational period. Inefficient users (C2) and high-power users (C3) collectively represented less than 5% of charging sessions across most charging bays, suggesting that extreme charging behaviors occur relatively infrequently within the charging network.

Table 13. Distribution of User Behavior Clusters Across Charging Bays (%).

Charging Bay	C0 (%)	C1 (%)	C2 (%)	C3 (%)
BAY 01	65.3	33.5	0.4	0.8
BAY 02	32.3	64.5	0.7	2.5
BAY 03	60.7	35.2	0.7	3.4
BAY 04	25.8	69.7	2.1	2.4
BAY 05	14.0	83.9	0.0	2.1
BAY 06	0.0	100.0	0.0	0.0

As shown in Table 5 Bay 01 was predominantly utilized by moderate users (C0), who accounted for 65.3% of charging sessions.

4.6. Anomaly Detection Analysis

Anomaly detection was performed to identify unusual charging sessions that may indicate operational inefficiencies, abnormal charging behavior, or potential charging

system irregularities. The distribution of anomalous charging sessions across the identified user behavior clusters is presented in Figure 8.

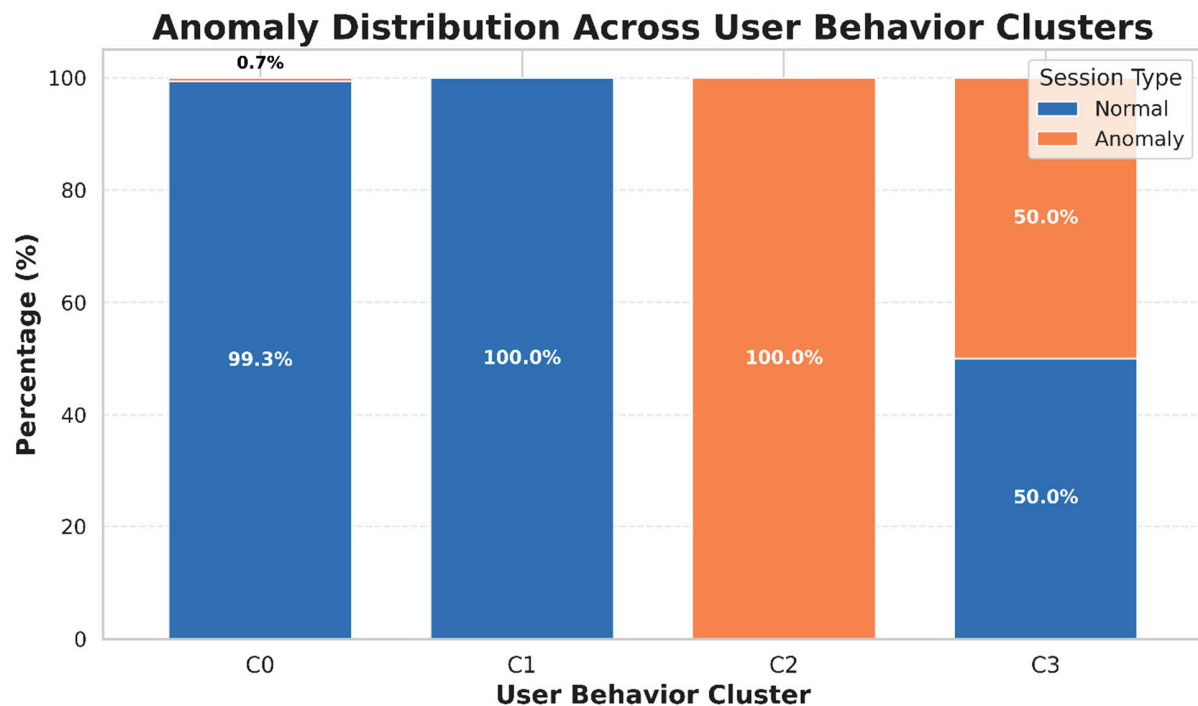


Figure 8. Distribution of normal and anomalous charging sessions across EV charging user behavior clusters.

The Isolation Forest algorithm identified a subset of charging sessions as anomalous based on deviations from the overall charging behavior patterns learned from the dataset. The total number of detected anomalies and their distributions across charging behavior clusters and charging bays are summarized in the following analysis.

In this study, anomaly identification was performed using the unsupervised Isolation Forest algorithm without requiring labelled anomaly data. Charging sessions classified as anomalous correspond to observations that deviated significantly from the overall behavioral patterns learned from the charging dataset based on the selected feature set. Consequently, the reported anomaly proportions represent model-detected abnormal charging sessions rather than confirmed operational faults. The identified anomalies should therefore be interpreted as charging sessions that warrant further operational investigation rather than definitive evidence of equipment malfunction or user-related irregularities.

The results presented in Figure 8 and Table 14 reveal substantial differences in anomaly occurrence among the identified charging behavior clusters. Cluster C1 (Regular Users) consisted entirely of normal charging sessions, with 100% of observations classified as non-anomalous. Similarly, Cluster C0 (Moderate Users) exhibited highly stable charging behavior, with only a very small proportion of sessions were identified as anomalies and 99.3% classified as normal.

In contrast, Cluster C2 (Inefficient Users) displayed the most abnormal charging behavior, with 100% of its charging sessions classified as anomalous. This result is consistent with the cluster characteristics observed in Section 4.2, where C2 exhibited exceptionally long charging durations (46.72 h) combined with minimal energy consumption (5.00 kWh) and extremely low charging efficiency ($ECR = 0.11$). Such characteristics indicate charging sessions that deviate substantially from the dominant charging behavior patterns and may be associated with inefficient charging behavior or other operational conditions requiring further investigation.

Table 14. Summary of anomaly detection results across charging behavior clusters.

Charging Behavior Cluster	Cluster Size	Anomaly Characteristics	Interpretation
C0—Moderate Users	730	A very small proportion of charging sessions were identified as anomalous, while 99.3% were classified as normal.	Stable charging behavior with only limited deviations from dominant charging patterns.
C1—Regular Users	1000	No anomalous charging sessions were detected; all charging sessions were classified as normal.	Represents the most stable and consistent charging behavior among all clusters.
C2—Inefficient Users	25	All charging sessions were classified as anomalous.	Indicates highly inefficient charging behavior characterized by prolonged charging duration, low energy consumption, and extremely low charging efficiency.
C3—High-Power Users	40	Both normal and anomalous charging sessions were observed.	High charging power does not necessarily indicate normal charging behavior and may also reflect unusual operational conditions.

Cluster C3 (High-Power Users) demonstrated a balanced distribution between normal and anomalous charging sessions, with around half of the charging sessions were classified as anomalies. Although this cluster exhibits the highest charging power and energy consumption, the presence of a substantial anomaly proportion indicates that intensive charging behavior may also be associated with unusual operational patterns.

These findings indicate that the proposed analytical framework effectively identifies charging sessions exhibiting abnormal behavioral characteristics, thereby supporting subsequent operational assessment and investigation.

To better understand the relative contribution of the selected charging features to the identified charging behavior clusters, an auxiliary Random Forest classifier was trained using the cluster assignments generated by the unsupervised learning framework. The feature importance scores were then computed using the impurity-based importance measure provided by the trained Random Forest model. The resulting importance scores are presented in Figure 9.

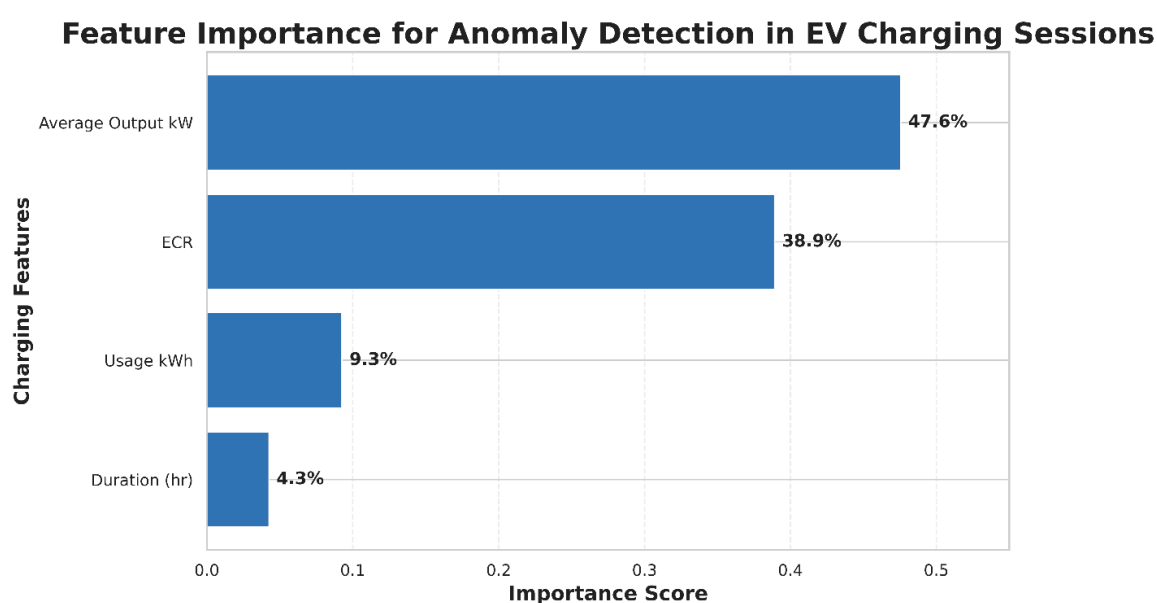


Figure 9. Feature importance of charging features derived from the auxiliary Random Forest classifier trained using the identified charging behavior clusters.

The Random Forest feature importance analysis indicates that Average Output Power was the most influential factor in anomaly detection, contributing 47.6% of the overall importance score. Energy Consumption Ratio (ECR) ranked second with an importance contribution of 38.9%. In comparison, Usage (kWh) and Duration (hr) contributed only 9.3% and 4.3%, respectively.

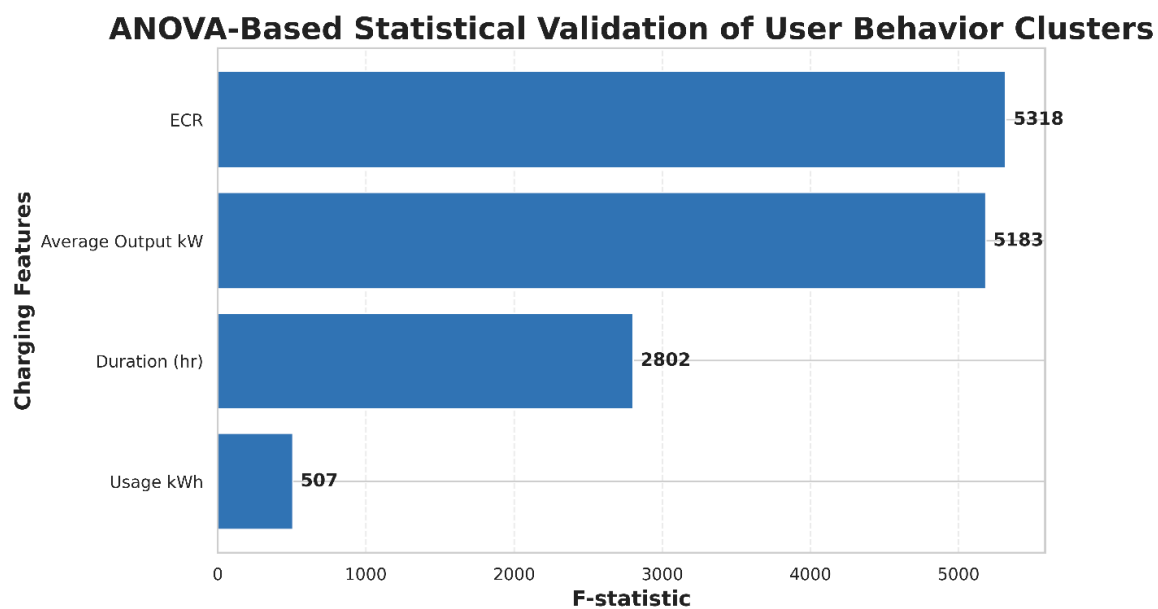
The dominance of Average Output Power and ECR suggests that anomalies are primarily associated with charging efficiency and power delivery characteristics rather than overall energy consumption or charging duration alone. This observation is consistent with the characteristics of Cluster C2, which exhibited abnormal charging efficiency despite relatively low energy consumption. Consequently, monitoring charging power and efficiency metrics may assist charging-network operators in identifying charging sessions that exhibit unusual operational characteristics and therefore merit further inspection.

4.7. Statistical Validation of User Behavior Clusters

To evaluate the statistical significance of the identified EV charging user behavior clusters, a one-way Analysis of Variance (ANOVA) was performed on all charging features. The ANOVA results are summarized in Table 15 and graphically illustrated in Figure 10.

Table 15. ANOVA-Based Statistical Validation of EV Charging User Behavior Clusters.

Feature	F-Statistic	<i>p</i> -Value	Significance
Usage (kWh)	506.58	<0.001	Significant
Duration (hr)	2801.68	<0.001	Significant
Average Output (kW)	5182.77	<0.001	Significant
ECR	5318.14	<0.001	Significant



All features show statistically significant differences across clusters ($p < 0.001$)

Figure 10. ANOVA-based statistical validation of EV charging user behavior clusters.

The ANOVA results demonstrate statistically significant differences among the identified clusters for all investigated charging features ($p < 0.001$), confirming that the clustering framework successfully partitioned the dataset into distinct behavioral groups. As shown in Figure 10, ECR achieved the highest F-statistic (5318.14), followed closely by Average

Output Power (5182.77), indicating that charging efficiency and power delivery are the most discriminative characteristics separating user behaviors.

Charging duration also exhibited substantial between-cluster variation, producing an F-statistic of 2801.68, while energy consumption remained highly significant with an F-statistic of 506.58. Although the F-statistics associated with energy consumption was comparatively lower than those of the other features, its value still indicates strong statistical separation among the identified user groups.

The exceptionally large F-statistics observed for ECR and Average Output Power are consistent with the cluster centroid analysis presented in Table 6. Cluster C2 exhibited extremely low charging efficiency (ECR = 0.11) and prolonged charging duration (46.72 h), whereas Cluster C3 demonstrated the highest charging efficiency (ECR = 21.04) and charging power (20.50 kW). These pronounced differences contribute significantly to the observed between-cluster variance and explain the strong statistical separation identified by the ANOVA analysis.

Overall, The ANOVA results quantitatively describe the statistical differences among the identified charging behavior clusters with respect to the selected charging features.

4.8. Correlation Analysis of Charging Features

To further investigate the relationships among charging characteristics and anomaly occurrence, Pearson correlation analysis was conducted. The resulting correlation coefficients are summarized in Table 16 and visualized in Figure 11.

Correlation Analysis of EV Charging Features and Anomaly Labels

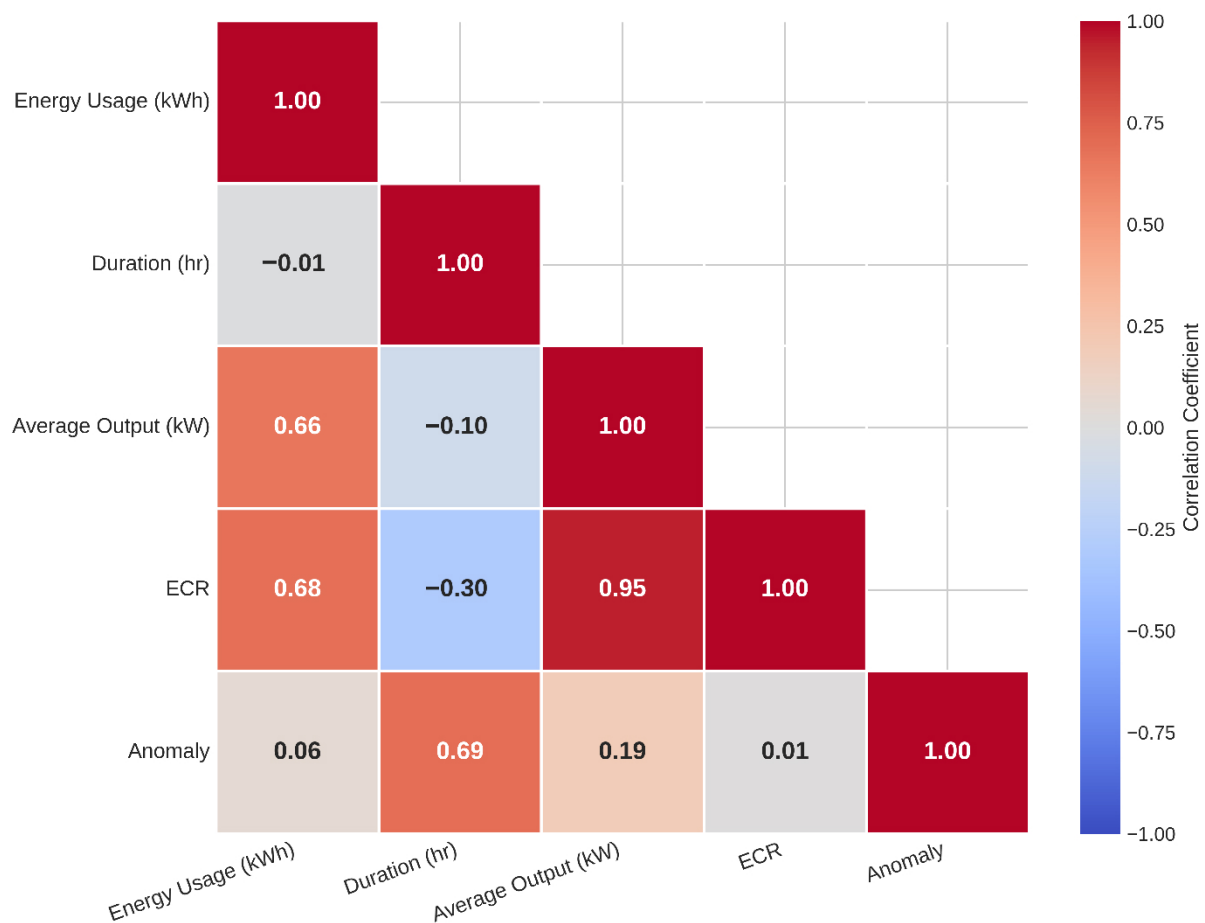


Figure 11. Correlation analysis of EV charging features and anomaly labels.

Table 16. Pearson correlation coefficients among EV charging features and anomaly labels.

Variable	Usage (kWh)	Duration (hr)	Average Output (kW)	ECR	Anomaly
Usage (kWh)	1.00	−0.01	0.66	0.68	0.06
Duration (hr)	−0.01	1.00	−0.10	−0.30	0.69
Average Output (kW)	0.66	−0.10	1.00	0.95	0.19
ECR	0.68	−0.30	0.95	1.00	0.01
Anomaly	0.06	0.69	0.19	0.01	1.00

The correlation analysis reveals several important relationships among charging characteristics. The strongest positive correlation was observed between Average Output Power and ECR ($r = 0.95$), indicating that charging efficiency is highly dependent on the charging power delivered during a session. This finding is consistent with the feature importance analysis presented in Figure 9, where both variables emerged as the most influential factors for anomaly detection.

Energy consumption also demonstrated moderate positive correlations with Average Output Power ($r = 0.66$) and ECR ($r = 0.68$), suggesting that sessions with higher energy demand are generally associated with higher charging efficiency and power delivery. In contrast, charging duration exhibited weak negative correlations with both Average Output Power ($r = -0.10$) and ECR ($r = -0.30$), indicating that prolonged charging sessions tend to be associated with lower charging efficiency.

Most notably, charging duration displayed the strongest relationship with anomaly occurrence ($r = 0.69$), suggesting that unusually long charging sessions are a primary indicator of abnormal charging behavior. This observation is consistent with the characteristics of Cluster C2, which exhibited extremely long charging durations and was entirely composed of anomalous charging sessions. Conversely, energy consumption ($r = 0.06$) and ECR ($r = 0.01$) showed only weak direct correlations with anomaly labels, indicating that anomalous behavior is more strongly associated with charging duration than with overall energy demand.

The strong positive correlation observed between Average Output and Energy Consumption Ratio (ECR) indicates that these variables describe closely related aspects of charging performance. Nevertheless, they were intentionally retained because each provides a distinct operational interpretation within the proposed analytical framework. Average Output quantifies the average charging power delivered during a charging session, whereas ECR characterizes the efficiency of energy transfer relative to charging duration. Since these variables were employed as descriptive features for unsupervised clustering and anomaly detection rather than as explanatory variables in a regression model, the observed correlation does not directly affect model parameter estimation. Instead, retaining both variables facilitates a more comprehensive characterization of charging behavior from complementary operational perspectives. Future research may further investigate feature redundancy using quantitative feature selection techniques, such as Variance Inflation Factor (VIF) analysis, to optimize the feature set for large-scale EV charging applications.

Overall, the correlation analysis confirms that charging power, charging efficiency, and charging duration are the key factors governing EV charging behavior and anomaly occurrence. These findings further support the clustering and anomaly detection results presented in Sections 4.2 and 4.4.

4.9. Key Findings and Practical Implications

The results demonstrate that EV charging behavior can be effectively characterized using a combination of clustering and anomaly detection techniques. The proposed K-Means clustering framework successfully identified four statistically distinct user behavior groups, namely Moderate Users (C0), Regular Users (C1), Inefficient Users (C2), and High-Power Users (C3). The clustering structure was validated through a silhouette score of 0.6086 and further supported by ANOVA results, which confirmed significant differences among all charging features ($p < 0.001$).

Among the identified clusters, Regular Users (C1) represented the dominant charging behavior pattern, accounting for 55.71% of all charging sessions. In contrast, Inefficient Users (C2) constituted only 1.39% of the dataset but exhibited exceptionally long charging durations (46.72 h), minimal energy consumption (5.00 kWh), and the lowest charging efficiency (ECR = 0.11). These characteristics indicate potentially abnormal charging practices and operational inefficiencies that warrant further investigation.

The anomaly detection analysis further highlighted the effectiveness of the proposed framework in identifying unusual charging behaviors. Cluster C2 consisted entirely of anomalous charging sessions, whereas Cluster C1 exhibited no detected anomalies. Average Output Power (47.6%) and Energy Consumption Ratio (38.9%) were the most influential variables contributing to anomaly identification, emphasizing the importance of charging efficiency and power delivery characteristics in distinguishing normal and abnormal charging behavior.

Correlation analysis provided additional insights into the relationships among charging features. A strong positive correlation was observed between Average Output Power and ECR ($r = 0.95$), indicating that charging efficiency is closely associated with power delivery performance. Furthermore, charging duration demonstrated the strongest association with anomaly occurrence ($r = 0.69$), suggesting that prolonged charging sessions represent a key indicator of abnormal charging behavior.

From an operational perspective, these findings provide valuable guidance for EV charging network operators and infrastructure planners. The ability to identify distinct user behavior groups enables targeted charging management strategies, while anomaly detection can facilitate early identification of inefficient charging sessions and potential operational irregularities. Consequently, the proposed framework can support improved charging station utilization, enhanced charging efficiency, and more effective management of EV charging infrastructure.

5. Discussion

5.1. Interpretation of User Behavior Clusters

The clustering results demonstrate the existence of distinct EV charging behavior groups within the analyzed charging infrastructure. The identified clusters reveal substantial heterogeneity in charging practices, indicating that EV users cannot be adequately represented by a single charging profile. The largest cluster, representing regular users, exhibited balanced charging characteristics with moderate charging durations, relatively high charging efficiency, and stable charging power levels. This observation suggests that most users follow predictable charging routines and utilize charging infrastructure in a consistent manner.

In contrast, the inefficient user cluster displayed exceptionally long charging durations coupled with very low energy consumption and poor charging efficiency. Such behavior may indicate vehicles remaining connected to charging stations after charging completion, low-power charging events, user inactivity, or operational inefficiencies associated with charging station utilization. The presence of this cluster highlights the importance of

monitoring charging duration alongside energy delivery metrics when assessing charging infrastructure performance.

The high-power user cluster exhibited the highest energy consumption, highest charging output power, and shortest charging durations [54]. These characteristics are consistent with users who prioritize rapid charging and efficient energy transfer. From an infrastructure perspective, this user group may place higher instantaneous power demands on charging networks while simultaneously contributing to improved charging by bay turnover.

The charging behavior patterns identified in this study are generally consistent with previous investigations reporting the existence of heterogeneous charging profiles among public EV users. Similar to earlier studies, the majority of charging sessions were characterized by moderate charging duration and energy consumption, while a relatively small proportion of sessions exhibited inefficient or abnormal charging characteristics. However, unlike many previous studies that primarily focused on user segmentation, the present work further associates charging behavior groups with anomaly detection and charging infrastructure utilization, thereby providing a more comprehensive interpretation of operational charging characteristics. The observed differences from previous studies are likely attributable to variations in charging infrastructure, user demographics, charging policies, and dataset characteristics, highlighting the importance of analyzing charging behavior within its specific operational context.

The statistical significance confirmed by ANOVA provides statistical evidence that the identified charging behavior clusters are distinguishable with respect to the analyzed charging-session characteristics. These findings support previous studies suggesting that EV charging behavior is inherently heterogeneous and influenced by diverse user preferences, charging requirements, and operational conditions.

5.2. Implications of Anomaly Detection Results

The anomaly detection analysis revealed a strong association between the identified anomalous charging sessions and the charging behavior represented by Cluster C2. Notably, all charging sessions within this cluster were classified as anomalies, reflecting exceptionally long charging durations combined with relatively low energy transfer. While these characteristics indicate highly inefficient charging events, they should not be interpreted solely as user-related behavior. Such charging patterns may also arise from prolonged vehicle occupation after charging completion, charger equipment malfunction, communication interruptions between the charging station and the management platform, incomplete charging transactions, or other operational conditions affecting the recorded charging sessions. Consequently, the identified anomalies should be regarded as charging sessions requiring further operational investigation rather than definitive evidence of inefficient user behavior. Nevertheless, the results demonstrate that charging duration, together with charging efficiency indicators, provides valuable information for identifying abnormal charging patterns within public EV charging networks.

The feature importance analysis further supports this interpretation. Average charging power and Energy Consumption Ratio (ECR) emerged as the most influential variables contributing to anomaly detection [55]. These findings indicate that charging efficiency and power delivery characteristics play a more important role in distinguishing abnormal charging behavior than total energy consumption alone. Consequently, monitoring charging power characteristics may provide a more effective strategy for identifying operational irregularities than focusing exclusively on energy usage metrics.

The confusion matrix results demonstrate balanced classification performance, with an overall accuracy, precision, recall, and F1-score of 80.0%. Although the performance

indicates that the proposed framework can effectively distinguish normal and anomalous charging sessions, the existence of misclassified observations suggests that certain charging behaviors may exhibit characteristics that overlap between normal and abnormal operating conditions. Such overlap is expected in real-world charging environments where user behavior can vary considerably.

The anomaly detection results obtained in this study are consistent with previous research demonstrating the effectiveness of unsupervised learning techniques for identifying unusual charging behaviors without requiring labelled datasets. Nevertheless, most existing studies primarily report anomaly detection performance in terms of classification metrics, whereas the proposed framework further interprets the detected anomalies within the context of charging behavior groups and charging infrastructure utilization. This integrated interpretation provides additional operational insight by linking anomalous charging sessions with user charging characteristics rather than treating anomaly detection as an isolated analytical task. Such integration contributes to a more comprehensive understanding of charging behavior and supports data-driven assessment of public EV charging infrastructure.

Overall, the anomaly detection results demonstrate that the proposed analytical workflow can identify charging-session patterns associated with inefficient or unusual charging behavior without requiring labeled anomaly data. These findings provide interpretable analytical insights that may support future operational assessment of charging infrastructure. However, the present study does not constitute a validated operational decision-support system, and additional validation using larger datasets, operational variables, and real-world deployment studies is required before practical implementation.

5.3. Implications for Charging Infrastructure Management

The findings of this study have several practical implications for charging network operators and infrastructure planners. First, the identification of distinct charging behavior groups enables more targeted management strategies. Charging stations dominated by regular users may require different operational policies compared with locations characterized by inefficient or high-power charging behaviors.

Second, charging bay utilization analysis revealed noticeable differences in behavioral distributions across charging bays. Such variations indicate that charging demand characteristics are location dependent and should be considered when planning future charging infrastructure expansions. Understanding user behavior at individual charging bays can assist operators in optimizing charger allocation, improving resource utilization, and reducing congestion.

Third, the strong relationship between charging duration and anomaly occurrence suggests that prolonged charging sessions may serve as an early warning indicator for inefficient charging activities. Charging network operators could leverage this information to implement automated monitoring systems capable of identifying unusual charging events and triggering operational alerts in real time.

Furthermore, the strong correlation observed between charging output power and charging efficiency indicates that maintaining stable charging power delivery may contribute to improved charging performance and overall infrastructure effectiveness. Such insights may support the development of intelligent charging management systems capable of optimizing charging operations based on observed user behavior patterns.

Compared with existing EV charging analytics studies, the primary contribution of the proposed framework lies in its integrated analytical workflow rather than the introduction of new machine learning algorithms. By combining behavioral clustering, anomaly detection, statistical validation, and charging infrastructure assessment within a

unified framework, the proposed methodology enables charging behavior to be interpreted from multiple complementary perspectives. This integrated analytical capability provides additional operational insight beyond conventional clustering or anomaly detection studies and represents the principal incremental contribution of the present work.

The findings presented in this study should be interpreted within the context of the analyzed charging infrastructure. The dataset was collected from six charging bays within a single public charging facility and therefore reflects the operational characteristics, charging policies, and user behavior associated with that environment. Although the dataset contains real operational charging sessions and is suitable for evaluating the proposed analytical framework, charging behavior may differ across geographical regions, charging technologies, tariff structures, vehicle fleets, and charging network operators. Consequently, the behavioral patterns identified in this work should not be interpreted as universally representative of all public EV charging infrastructures. Future research should validate the proposed framework using multi-site datasets collected from different charging networks, charger power ratings, and geographical locations to further evaluate its generalizability and robustness.

5.4. Operational Implications of the Proposed Analytical Workflow

Although the proposed analytical workflow was evaluated using historical charging-session data rather than a real-time operational environment, the results demonstrate its applicability for analyzing charging-session behavior and identifying operational charging patterns within the analyzed dataset. The identified charging behavior clusters provide insights into charging-session characteristics, while the anomaly detection analysis highlights charging sessions exhibiting unusual operational behavior. These findings may assist charging-network operators in understanding charging-bay utilization, identifying inefficient charging practices, prioritizing infrastructure inspection, and supporting maintenance planning. However, the proposed workflow should be regarded as a data-driven analytical tool rather than a validated operational decision-support or autonomous control system.

It is important to emphasize that the present study does not include real-time deployment, closed-loop operational evaluation, or independent field validation. Instead, the proposed workflow was applied to historical charging-session records collected from a public charging facility to demonstrate its capability for characterizing charging-session behavior and identifying anomalous charging-session patterns within the available dataset. Consequently, the reported findings should be interpreted as analytical evidence derived from the analyzed charging infrastructure rather than as operational recommendations that can be directly generalized to other charging networks. Future research should evaluate the proposed workflow using larger multi-site datasets and real-time charging management systems to further assess its applicability under diverse operational conditions.

5.5. Study Limitations and Future Research Directions

The findings of this study should be interpreted within the context of the analyzed charging infrastructure. The dataset consists of 1795 charging sessions collected from six charging bays at a single public charging facility and therefore reflects the operational characteristics and charging behavior observed in that specific environment. Although the dataset is suitable for evaluating the proposed analytical workflow, the identified charging behavior patterns should not be interpreted as universally representative of public EV charging networks. Additional validation using larger multi-site datasets covering different charger technologies, geographical regions, charging policies, and vehicle fleets would further improve the generalizability of the proposed framework.

Second, the analysis focused on a limited set of charging behavior indicators, namely energy consumption, charging duration, average charging output power, and ECR. Additional variables such as charging start time, arrival patterns, vehicle type, battery state of charge, weather conditions, and electricity pricing may provide further insights into charging behavior and anomaly occurrence.

Third, the anomaly detection framework utilized Isolation Forest as the primary anomaly detection algorithm. While the obtained results demonstrate satisfactory performance, future studies may investigate alternative approaches including Local Outlier Factor, One-Class Support Vector Machines, deep learning-based autoencoders, and hybrid anomaly detection frameworks to further improve detection accuracy.

One limitation of the present study is that temporal charging characteristics, such as charging start time, completion time, time of day, day of week, and seasonal charging patterns, were not incorporated into the analytical framework. These variables were intentionally excluded to maintain the focus on charging-session behavioral indicators and to evaluate their effectiveness for charging behavior characterization and anomaly detection. Future research will investigate the integration of temporal features with behavioral indicators to provide a more comprehensive understanding of charging demand dynamics, user behavior, infrastructure utilization, and time-dependent anomaly patterns.

Another limitation of the present study is that the contribution of individual analytical components was not evaluated through dedicated ablation experiments, nor was the proposed framework benchmarked against a broad range of state-of-the-art machine learning methods. Although the selected algorithms were chosen based on their computational efficiency, interpretability, and widespread adoption in EV charging analytics, future research should investigate alternative clustering and anomaly detection techniques and perform systematic ablation analyses to quantify the contribution of each framework component. Such investigations would provide a more comprehensive assessment of the proposed framework and further strengthen its applicability in practical EV charging environments.

Future research may also explore real-time anomaly detection, predictive charging behavior modeling, charging demand forecasting, and digital twin-based charging infrastructure monitoring. The integration of behavioral clustering, anomaly detection, and predictive analytics within intelligent charging management systems represents a promising direction for supporting next-generation EV charging networks.

Overall, the present study evaluates a single clustering algorithm and a single anomaly detection algorithm to demonstrate the proposed analytical workflow. Although K-Means and Isolation Forest are widely adopted unsupervised learning techniques for continuous numerical data, future research should investigate alternative approaches, including Gaussian Mixture Models, DBSCAN, Local Outlier Factor (LOF), One-Class SVM, and deep learning-based anomaly detection models. Comparative evaluation of these methods under different parameter settings and operational conditions would provide further insight into algorithm selection for EV charging-session analysis.

6. Conclusions

This study presented an integrated behavioral analytics workflow for EV charging-session analysis using real-world charging data collected from six charging bays. The proposed framework integrated feature engineering, K-Means clustering, Principal Component Analysis (PCA), Isolation Forest anomaly detection, statistical validation, correlation analysis, and charging infrastructure utilization assessment to provide a holistic understanding of EV charging behavior.

Four charging behavior indicators, namely energy consumption (Usage), charging duration (Duration), average charging output power (Average Output), and Energy Con-

sumption Ratio (ECR), were extracted and analyzed. The clustering results identified four distinct charging behavior groups representing moderate users, regular users, inefficient users, and high-power users. Clustering validation demonstrated strong cluster quality, with a silhouette score of 0.6086, while PCA retained 89.8% of the total variance using two principal components, confirming the effectiveness of the clustering framework in capturing meaningful behavioral patterns.

The anomaly detection analysis revealed significant differences in abnormal charging behavior across the identified user groups. Inefficient users exhibited the highest anomaly occurrence and were characterized by prolonged charging durations, low energy consumption, and poor charging efficiency. In contrast, regular users demonstrated highly consistent charging behavior with no detected anomalies. Analysis indicated that average charging output power and ECR were the most influential variables contributing to anomaly identification. Statistical validation through ANOVA confirmed significant differences among all charging behavior clusters ($p < 0.001$), while correlation analysis revealed a strong positive relationship between charging power and charging efficiency ($r = 0.95$). The proposed anomaly detection framework achieved balanced classification performance with an accuracy, precision, recall, and F1-score of 80.0%.

The findings demonstrate that integrating clustering and anomaly detection techniques provides valuable insights into EV charging behavior and charging infrastructure utilization. The proposed framework enables the identification of distinct user behavior patterns, detection of abnormal charging activities, and assessment of charging station utilization characteristics. These capabilities can support charging network operators in improving charging efficiency, optimizing infrastructure planning, enhancing operational reliability, and enabling data-driven decision-making within EV charging systems.

The proposed framework demonstrates that integrating behavioral clustering, anomaly detection, and statistical validation can provide meaningful analytical insights into public EV charging behavior using real-world charging-session data. Rather than serving as a fully deployed operational management system, the framework is intended as a decision-support methodology capable of assisting charging-network operators in understanding charging behavior, identifying potential operational irregularities, and supporting infrastructure assessment. Although additional validation using larger multi-site datasets and real-time deployment is required before operational implementation, the proposed framework establishes a practical foundation for data-driven charging infrastructure analytics and future intelligent charging management systems.

Author Contributions: Conceptualization, M.S.H. and G.R.; methodology, M.S.H.; software, M.S.H.; validation, M.S.H., M.A.Q. and G.R.; formal analysis, M.S.H.; investigation, M.S.H.; resources, G.R.; data curation, M.S.H.; writing—original draft preparation, M.S.H.; writing—review and editing, M.A.Q. and G.R.; visualization, M.S.H.; supervision, G.R.; project administration, G.R.; funding acquisition, G.R. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and analyzed during the current study were obtained from laboratory experiments conducted by the authors. The data are not publicly available but are available from the corresponding authors upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the support and facilities provided by the PV Energy Laboratory, Multimedia University (MMU), Cyberjaya, Malaysia. The authors also appreciate the valuable technical assistance and research environment that contributed to the successful completion of this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Alanazi, F. Electric Vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation. *Appl. Sci.* **2023**, *13*, 6016. [\[CrossRef\]](#)
- Lipu, M.S.H.; Mamun, A.A.; Ansari, S.; Miah, M.S.; Hasan, K.; Meraj, S.T.; Abdolrasol, M.G.M.; Rahman, T.; Maruf, M.H.; Sarker, M.R.; et al. Battery Management, Key Technologies, Methods, Issues, and Future Trends of Electric Vehicles: A Pathway toward Achieving Sustainable Development Goals. *Batteries* **2022**, *8*, 119. [\[CrossRef\]](#)
- Kumar, M.; Panda, K.P.; Naayagi, R.T.; Thakur, R.; Panda, G. Comprehensive Review of Electric Vehicle Technology and Its Impacts: Detailed Investigation of Charging Infrastructure, Power Management, and Control Techniques. *Appl. Sci.* **2023**, *13*, 8919. [\[CrossRef\]](#)
- Hossen, M.S.; Ramasamy, G.; Eng, N.E. AI-OCPP Framework for Smart and Safe EV Charging: Real-World Performance Evaluation. *IEEE Access* **2026**, *14*, 34755–34768. [\[CrossRef\]](#)
- Akshay, K.C.; Grace, G.H.; Gunasekaran, K.; Samikannu, R. Power consumption prediction for electric vehicle charging stations and forecasting income. *Sci. Rep.* **2024**, *14*, 6497. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ma, T.Y.; Fang, Y. Survey of charging management and infrastructure planning for electrified demand-responsive transport systems: Methodologies and recent developments. *Eur. Transp. Res. Rev.* **2022**, *14*, 36. [\[CrossRef\]](#)
- Mazhar, T.; Asif, R.N.; Malik, M.A.; Nadeem, M.A.; Haq, I.; Iqbal, M.; Kamran, M.; Ashraf, S. Electric Vehicle Charging System in the Smart Grid Using Different Machine Learning Methods. *Sustainability* **2023**, *15*, 2603. [\[CrossRef\]](#)
- Nespoli, A.; Ogliari, E.; Leva, S. User Behavior Clustering Based Method for EV Charging Forecast. *IEEE Access* **2023**, *11*, 6273–6283. [\[CrossRef\]](#)
- Nazari, M.; Hussain, A.; Musilek, P. Applications of Clustering Methods for Different Aspects of Electric Vehicles. *Electronics* **2023**, *12*, 790. [\[CrossRef\]](#)
- Voinea, G.-D.; Gîrbacia, F.; Duguleană, M.; Postelnicu, C.-C. Planning Future EV Charging Infrastructure by Forecasting Spatio-Temporal Adoption Trends Across Heterogeneous User Segments. *Information* **2025**, *16*, 933. [\[CrossRef\]](#)
- Hossen, M.S.; Ramasamy, G.; Eng Eng, N.; Qwaid, M.A. An Anomaly-Aware, Q-Learning Framework for Real-Time Scheduling in Multi-Station EV Charging Networks. *Electronics* **2026**, *15*, 2494. [\[CrossRef\]](#)
- Johnson, J.; Berg, T.; Anderson, B.; Wright, B. Review of Electric Vehicle Charger Cybersecurity Vulnerabilities, Potential Impacts, and Defenses. *Energies* **2022**, *15*, 3931. [\[CrossRef\]](#)
- Hossen, M.S.; Sarker, M.T.; Al Qwaid, M.; Ramasamy, G.; Eng Eng, N. AI-Driven Framework for Secure and Efficient Load Management in Multi-Station EV Charging Networks. *World Electr. Veh. J.* **2025**, *16*, 370. [\[CrossRef\]](#)
- Lei, Z.; Xing, B.; Liu, J.; Yang, Y.; Miao, T.; Lu, Y. Sustainable and Reliable Operation of EV Charging Infrastructure: A Lightweight Prototype-Driven Contrastive Learning Framework for Fault Diagnosis Under Class-Imbalanced Conditions. *Sustainability* **2026**, *18*, 5783. [\[CrossRef\]](#)
- Chen, J.; Zhang, J.; Qian, R.; Yuan, J.; Ren, Y. An Anomaly Detection Method for Wireless Sensor Networks Based on the Improved Isolation Forest. *Appl. Sci.* **2023**, *13*, 702. [\[CrossRef\]](#)
- Ayoade, I.A.; Longe, O.M. A Comprehensive Review on Smart Electromobility Charging Infrastructure. *World Electr. Veh. J.* **2024**, *15*, 286. [\[CrossRef\]](#)
- Hossen, M.S.; Ramasamy, G.; Eng, N.E.; Sarker, M.T.; Nabi, M.S. Artificial Intelligence for Sustainable EV Charging: Integrated Forecasting, Anomaly Detection, and Scheduling in Smart Grids. *Anom. Detect. Sched. Smart Grids* **2025**. [\[CrossRef\]](#)
- Almaghrebi, A.; James, K.; Al Juheshi, F.; Alahmad, M. Insights into Household Electric Vehicle Charging Behavior: Analysis and Predictive Modeling. *Energies* **2024**, *17*, 925. [\[CrossRef\]](#)
- Andrenacci, N.; Valentini, M.P. A Literature Review on the Charging Behaviour of Private Electric Vehicles. *Appl. Sci.* **2023**, *13*, 12877. [\[CrossRef\]](#)
- Liang, Y.; Wang, H.; Zhao, X. Analysis of factors affecting economic operation of electric vehicle charging station based on DEMATEL-ISM. *Comput. Ind. Eng.* **2022**, *163*, 107818. [\[CrossRef\]](#)
- Alanazi, F.; Alshammari, T.O.; Azam, A. Optimal Charging Station Placement and Scheduling for Electric Vehicles in Smart Cities. *Sustainability* **2023**, *15*, 16030. [\[CrossRef\]](#)
- Wang, R.; Xing, Q.; Chen, Z.; Zhang, Z.; Liu, B. Modeling and Analysis of Electric Vehicle User Behavior Based on Full Data Chain Driven. *Sustainability* **2022**, *14*, 8600. [\[CrossRef\]](#)
- ElKashlan, M.; Elsayed, M.S.; Jurcut, A.D.; Azer, M. A Machine Learning-Based Intrusion Detection System for IoT Electric Vehicle Charging Stations (EVCSs). *Electronics* **2023**, *12*, 1044. [\[CrossRef\]](#)
- Miraftabzadeh, S.M.; Colombo, C.G.; Longo, M.; Foiadelli, F. K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access* **2023**, *11*, 119596–119633. [\[CrossRef\]](#)
- Yang, A.; Zhang, G.; Tian, C.; Peng, W.; Liu, Y. Charging Behavior Portrait of Electric Vehicle Users Based on Fuzzy C-Means Clustering Algorithm. *Energies* **2024**, *17*, 1651. [\[CrossRef\]](#)

26. Capeletti, M.B.; Hammerschmitt, B.K.; Silva, L.N.F.D.; Knak Neto, N.; Passinato Sausen, J.; Barriquello, C.H.; Abaide, A.D.R. User Behavior in Fast Charging of Electric Vehicles: An Analysis of Parameters and Clustering. *Energies* **2024**, *17*, 4850. [\[CrossRef\]](#)
27. Mahnoor; Shafi, I.; Chaudhry, M.; Montero, E.C.; Alvarado, E.S.; Diez, I.d.I.T.; Samad, A.; Ashraf, I. A Review of Approaches for Rapid Data Clustering: Challenges, Opportunities, and Future Directions. *IEEE Access* **2024**, *12*, 138086–138120. [\[CrossRef\]](#)
28. Zaheer, R.; Kashif Hanif, M.; Umer Sarwar, M.; Talib, R. Evaluating the Effectiveness of Dimensionality Reduction on Machine Learning Algorithms in Time Series Forecasting. *IEEE Access* **2025**, *13*, 50493–50510. [\[CrossRef\]](#)
29. González, C.M.; Galvis, J.; Rosero-Garcia, J. Detection of Anomalies in Electricity Consumption Patterns Using Density-Based Clustering: A Hybrid PCA-HDBSCAN Approach Applied to Advanced Metering Data. *Appl. Sci.* **2026**, *16*, 4337. [\[CrossRef\]](#)
30. Ahmed, M.H.; Tiun, S.; Omar, N.; Sani, N.S. Short Text Clustering Algorithms, Application and Challenges: A Survey. *Appl. Sci.* **2023**, *13*, 342. [\[CrossRef\]](#)
31. Khalaf, M.; Ayad, A.; Tushar, M.H.K.; Kassouf, M.; Kundur, D. A Survey on Cyber-Physical Security of Active Distribution Networks in Smart Grids. *IEEE Access* **2024**, *12*, 29414–29444. [\[CrossRef\]](#)
32. Nevisi, M.M.S.; Shoeibi, M.; Hernando-Gallego, F.; Martín, D.; Khatami, S.S. An Evolutionary Deep Reinforcement Learning-Based Framework for Efficient Anomaly Detection in Smart Power Distribution Grids. *Energies* **2025**, *18*, 2435. [\[CrossRef\]](#)
33. Greco, D.; Gaggero, G.B. Enhancing Cybersecurity Monitoring in Battery Energy Storage Systems with Graph Neural Networks. *Energies* **2026**, *19*, 479. [\[CrossRef\]](#)
34. Mendes, T.; Cardoso, P.J.S.; Monteiro, J.; Raposo, J. Anomaly Detection of Consumption in Hotel Units: A Case Study Comparing Isolation Forest and Variational Autoencoder Algorithms. *Appl. Sci.* **2023**, *13*, 314. [\[CrossRef\]](#)
35. Madhu Shree, M.; Ranjan, R.; Dechamma, M.P. Time Series-Based Analysis of Energy Consumption: Forecasting and Anomaly Detection Using LSTM and Isolation Forest. In *Artificial Intelligence Based Smart and Secured Applications. ASCIS 2024. Communications in Computer and Information Science*; Rajagopal, S., Popat, K., Meva, D., Bajaja, S., Mudholkar, P., Eds.; Springer: Cham, Switzerland, 2025; Volume 2424. [\[CrossRef\]](#)
36. Kubiczek, J.; Koczy, J. Multidimensional Comparison of Electric and Combustion Vehicles: A Clustering Based Analysis from the Polish Market. *Energies* **2025**, *18*, 5554. [\[CrossRef\]](#)
37. Dai, S.; Yuan, L.; Zhong, J.; Liu, X.; Liu, Z. Forecasting Residential EV Charging Pile Capacity in Urban Power Systems: A Cointegration–BiLSTM Hybrid Approach. *Sustainability* **2025**, *17*, 6356. [\[CrossRef\]](#)
38. Zhang, T.; Peng, Q.; Zeng, S. Predicting EV Charging Demand in Renewable-Energy-Powered Grids Using Explainable Machine Learning. *Sustainability* **2025**, *17*, 4158. [\[CrossRef\]](#)
39. Faustino, F.J.; Lopes, J.C.; Melo, J.D.; Sousa, T.; Padilha-Feltrin, A.; Brito, J.A.; Garcia, C.O. Identifying charging zones to allocate public charging stations for electric vehicles. *Energy* **2023**, *283*, 128436. [\[CrossRef\]](#)
40. Huang, W.; Wang, J.; Wang, J.; Zeng, H.; Zhou, M.; Cao, J. EV charging load profile identification and seasonal difference analysis via charging sessions data of charging stations. *Energy* **2024**, *288*, 129771. [\[CrossRef\]](#)
41. Abdullahi, M.R.; Lu, Q.-C.; Hussain, A.; Tripura, S.; Xu, P.-C.; Wang, S. Location optimization of EV charging stations: A custom K-means cluster algorithm approach. *Energy Rep.* **2024**, *12*, 5367–5382. [\[CrossRef\]](#)
42. Hasan, K.N.; Rahman, M.T.; Terrill, C.; McLean, R.; Rodricks, R.; Sharma, A.; Islam, A. A framework to investigate charger capacity utilization and network voltage profile through residential EV charging data clustering. *Sustain. Energy Technol. Assess.* **2025**, *74*, 104141. [\[CrossRef\]](#)
43. Zhang, Y.; Guo, L.; Chen, Z.; Yan, H.; Liang, L.; Lin, C. A novel framework for vehicle charging pattern recognition and charging duration prediction based on EA-CAE and K-means clustering. *Energy AI* **2025**, *22*, 100599. [\[CrossRef\]](#)
44. Saklani, M.; Saini, D.K.; Yadav, M.; Siano, P. Scalable Data-Driven EV Charging Optimization Using HDBSCAN-LP for Real-Time Pricing Load Management. *Smart Cities* **2025**, *8*, 139. [\[CrossRef\]](#)
45. Yu, Q.; Wang, Z.; Song, Y.; Shen, X.; Zhang, H. Potential and flexibility analysis of electric taxi fleets V2G system based on trajectory data and agent-based modeling. *Appl. Energy* **2024**, *355*, 122323. [\[CrossRef\]](#)
46. Jonas, T.; Daniels, N.; Macht, G. Electric Vehicle User Behavior: An Analysis of Charging Station Utilization in Canada. *Energies* **2023**, *16*, 1592. [\[CrossRef\]](#)
47. Ferretti, F.; De Paola, A. Machine learning identification of Electric Vehicles from charging session data. *Energy AI* **2025**, *20*, 100502. [\[CrossRef\]](#)
48. Habbak, H.; Mahmoud, M.; Fouda, M.M.; Alsabaan, M.; Mattar, A.; Salama, G.I.; Metwally, K. Efficient One-Class False Data Detector Based on Deep SVDD for Smart Grids. *Energies* **2023**, *16*, 7069. [\[CrossRef\]](#)
49. Yu, J.; Guo, Y.; Zhang, W. Anomaly Detection for Charging Voltage Profiles in Battery Cells in an Energy Storage Station Based on Robust Principal Component Analysis. *Appl. Sci.* **2024**, *14*, 7552. [\[CrossRef\]](#)
50. Hossen, M.S.; Sarker, M.T.; Nabi, M.S.; Bannah, H.; Ramasamy, G.; Eng Eng, N. Federated AI-OCPP Framework for Secure and Scalable EV Charging in Smart Cities. *Urban Sci.* **2025**, *9*, 363. [\[CrossRef\]](#)
51. Hossen, M.S.; Ramasamy, G.; Sarker, M.T.; Eng, N.E. Real-World Tariff-Aware Safe Reinforcement Learning for Grid-Stable OCPP EV Charging Networks. *IEEE Access* **2026**, *14*, 18530–18545. [\[CrossRef\]](#)

52. Märtz, A.; Langenmayr, U.; Ried, S.; Seddig, K.; Jochem, P. Charging Behavior of Electric Vehicles: Temporal Clustering Based on Real-World Data. *Energies* **2022**, *15*, 6575. [[CrossRef](#)]
53. Matrone, S.; Ogliari, E.G.C.; Nespoli, A.; Gruosso, G.; Gandelli, A. Electric Vehicles Charging Sessions Classification Technique for Optimized Battery Charge Based on Machine Learning. *IEEE Access* **2023**, *11*, 52444–52451. [[CrossRef](#)]
54. Li, Q.; Wang, G.; Zhang, Y.; Yang, Q. Analysis of user electricity consumption behavior based on density peak clustering with shared neighbors and attractiveness. *Concurr. Comput. Pract. Exper* **2023**, *35*, e7518. [[CrossRef](#)]
55. Aldossary, M. Enhancing Urban Electric Vehicle (EV) Fleet Management Efficiency in Smart Cities: A Predictive Hybrid Deep Learning Framework. *Smart Cities* **2024**, *7*, 3678–3704. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.