scientific reports

Check for updates

OPEN

Enhancing fault detection in new energy vehicles via novel ensemble approach

Iqra Akhtar[1], Mahnoor Nabeel[2], Umair Shahid[1✉], Kashif Munir[2✉], Ali Raza[3,4], Irene Delgado Noya[5,6,7,8], Santos Gracia Villar[5,6,9,10] & Imran Ashraf[11]

New energy vehicles (NEVs) has emerged as a sustainable alternative to conventional vehicles, however have unresolved reliability challenges due to their complex electronic systems and varying operating conditions. Faults in drivetrain and battery systems, occurring at rates up to 12% annually, present significant barriers to the widespread adoption of NEVs. This study proposes a robust fault detection framework that applies multiple machine learning and deep learning models to address these challenges. The research utilizes the benchmark NEV fault diagnosis dataset, which contains real-world sensor data from NEVs. The models tested include logistic regression, passive-aggressive classifier, ridge classifier, perceptron, gated recurrent unit (GRU), convolutional neural network, and artificial neural network. The proposed ensemble GRULogX model stands out among the implemented model, leveraging GRU with logistic regression and other key classifiers, and achieved 99% accuracy, demonstrating high precision and recall. Cross-validation and hyperparameter optimization were adopted to further ensure the model's generalizability and reliability. This research enhances the fault detection capabilities of NEVs, thereby improving their reliability and supporting the wider adoption of clean energy transportation solutions.

Keywords Transportation, New energy vehicles, Fault detection, Deep learning, Sensor data, NEV reliability, Ensemble learning

The transport industry has been under a significant paradigm shift due to the necessity of reducing the carbon footprint and transition to cleaner and more sustainable energy sources[1]. The increasing environmental concerns, combined with the exhaustion of fossil fuels, have fueled the demand for alternative energy solutions in the transportation sector[2]. New energy vehicles (NEVs), such as electric vehicles (EVs), plug-in hybrid electric vehicles (PHEVs), and hydrogen fuel cell vehicles (FCVs), are increasingly becoming prime alternatives to traditional gasoline-powered cars[1,2].

NEVs are regarded as a promising solution to reduce the environmental impact of traditional vehicles, primarily by lowering carbon dioxide (CO_2) emissions. New estimates suggest that NEVs might reduce average car CO_2 emissions by up to 40%, significantly helping urban air and nations reach goals such as climate targets. Cutting emissions really matters, especially in cities where traffic pollution poses big health risks. Despite the promising benefits of NEVs, their adoption is not without challenges. The far-reaching utilisation of NEVs is still hindered by a few specialised and viable barriers[1,2].

Among the most significant challenges are the anxiety of the battery distribution area, the degradation of performance over time, long load times, and incomplete load infrastructure, which exist in many areas, especially in developing countries[3]. Studies have shown that, although NEV battery technology has increased significantly in recent years, the lack of charging infrastructure remains a significant obstacle, especially in rural or underdeveloped areas. Additional issues of high purchase prices of NEV's high, long-term performance

[1]Department of Electrical and Biomedical Engineering, Khwaja Fareed University Of Engineering And Information Technology, Rahim Yar Khan, Pakistan. [2]Institute of Information Technology, Khwaja Fareed University Of Engineering And Information Technology, Rahim Yar Khan 64200, Pakistan. [3]Department of Precision Medicine, Sungkyunkwan University School Of Medicine, Suwon 16419, Republic Of Korea. [4]Department of Software Engineering, University of Lahore, Lahore, Pakistan. [5]Universidad Europea Del Atlantico, Isabel Torres 21, Santander 39011, Spain. [6]Universidad Internacional Iberoamericana, Campeche 24560, Mexico. [7]Fundacion Universitaria Internacional De Colombia, Bogota, Colombia. [8]Universidad De La Romana, La Romana, República Dominicana. [9]Universidad Internacional Iberoamericana Arecibo, Puerto Rico 00613, USA. [10]Universidade Internacional Do Cuanza, Cuito, Bie, Angola. [11]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic Of Korea. ✉email: ushahid@kfueit.edu.pk; kashif.munir@kfueit.edu.pk

requirement of battery, and energy efficiency continue to deter potential buyers. Addressing these challenges requires advanced technologies that can improve NEVs' efficiency, reliability, and user experience[3].

Machine learning (ML) has become an effective tool to solve a wide range of problems in different industries, including transport. By analyzing large amount of sensory data collected from NEVs during operation, ML algorithms can provide valuable information that allows for predictive maintenance, optimization of battery health management, and supports real-time energy consumption[2]. This process can significantly improve the overall performance and reliability of NEVs, solving the main problems such as battery degradation, energy consumption, and fault detection.

ML models are increasingly used to predict the degradation of the battery and detect faults in NEVs which supports manufacturers and operators to take necessary precautions. This can help reduce expensive repairs, minimize the time of death, and improve the reliability of the vehicle[4]. In addition, the use of ML to manage the actual energy is very important to optimize the discharge cycle and battery, thus improving its life expectancy and performance. Algorithms controlling various functions of autonomous systems are also integral to the future of NEVs. Often, such systems utilize ML for object detection, lane recognition, and decision-making tasks in dynamic traffic environments, all of which contribute to the safety and operational efficiency of NEVs.

Fault detection in NEVs has become a critical task and several prominent works have been conduction. The study[5] offered a robust framework for detecting faults in NEVs, utilizing ML and deep learning (DL) models. Many models were assessed, including traditional logistic regression, passive classification, ridge classifier, and perceptron, as well as more advanced models like gated recurrence units (GRU), an architecture that uses gating mechanisms to efficiently capture dependencies in sequential data, convolutional neural networks (CNNs) and artificial neural network (ANN)[4]. Research utilizing the fault diagnostic dataset, providing real sensor data collected from NEVs. This ensures that the results of this study are directly applicable to the real world. In this context, the primary contributions of this research are:

- A comprehensive fault detection framework is proposed, utilizing both traditional ML models and advanced deep learning architectures to address reliability challenges in NEVs.
- Introduced a novel ensemble GRULogX model that integrates GRU and logistic regression, achieving high accuracy on real-world NEV fault diagnosis data.
- Validated the proposed model using real-world sensor data from the Kaggle NEV fault diagnosis dataset, demonstrating its effectiveness in practical environments.
- Improved model robustness and generalizability through cross-validation and hyperparameter optimization, with strong performance on time-series NEV sensor data.

This paper is structured as follows: Section 2 reviews and discusses the related research on the application of ML and DL in NEVs, highlighting the advancements in fault detection and predictive maintenance. Section 3 shows the methodology used in the study, detailing the dataset preparation, model development, and evaluation process. Section 4 discusses the experimental results and provides an analysis of the model's performance. Finally, Section 5 concludes the paper, summarizing the findings and offering suggestions for future research directions to further improve the performance and reliability of NEVs.

Related work

The NEVs sector has drawn much interest recently because of their capacity to somewhat reduce the environmental consequences of traditional transportation[6]. Many studies have investigated different aspects of NEV technology, including battery management, adoption patterns, fault detection, and environmental impacts. Emphasizing each study's methodologies, datasets, and conclusions, the next part condenses significant research that has advanced our knowledge of NEVs. These studies provide insightful suggestions that guide present research and show how to improve NEV acceptance, performance, and safety.

The study[7] dealt with serious problems with battery security (vehicles of electric vehicles) by developing a method for diagnosing data control errors for battery packs. Beijing's surveillance and electric vehicle service centres were monitored and analyzed for real-time operational data from the fleet, focusing on the 91-cell battery pack (all 10 seconds) EV model. To record larger patterns, they also train a three-layer neural network (Architecture 1 10) with aggregated fleet data, combining vehicle diagnostics with big data trends. The most important findings include classifying random errors, such as manufacturing errors and accidents, for two types of failure, as well as the ability to find possible locations within the package. Surprisingly, statistical analysis showed that cells on the front and bottom of the battery tend to break faster than other cells. This indicates that further investigation of various batteries and operating conditions is needed to ensure generalization.

The author examined the impact of China's NEV guidelines through the lens of consumer heterogeneity and addressed the issue of inconsistent political effectiveness in demographic segments[2]. This study used a large dataset for consumer surveys in various cities in China, and used multivariate regression in interaction terms, specifically to examine how individual characteristics (age, income, etc.) moderated the relationship between political incentives and NEV acceptance. The analysis showed that financial subsidies and licensing privileges had a positive effect on acceptance overall, but their effectiveness was significantly different among consumer groups. For example, environmental policies were equally appreciated by both young and older people. On the other hand, older consumers showed greater sensitivity to direct financial incentives. A major contribution of to this study is to highlight the need for tailor-made political design based on consumer profiles. However, the study was limited by its cross-sectional features and reliance on self-reported data that may not be captured over time. The authors recommended the integration of longitudinal studies with behavioral tracking data to improve future analyses.

The critical issue of early error detection in NEVs' lithium-ion batteries was investigated by[8] to improve operational safety and reliability. The author proposed a new diagnostic approach using a combination of statistical process control techniques (SPC) and support vector machines (SVMs) to identify and classify abnormal battery behavior. They used actual data records collected by the battery management system (BMS) that included parameters such as voltage, electricity, temperature, and load/load cycles. The hybrid approach achieved high error detection accuracy and significantly exceeded traditional threshold value methods. One of the most important outcomes was the system's ability to recognize errors early on before escalating to a serious failure, reducing maintenance costs and security risks. However, the authors discover limitations on the treatment of unstructured or missing data, and this may degrade the model's capabilities when exposed to unknown operating conditions. They recommended expanding the data records with adaptive learning mechanisms to improve the generalization of future research.

In Ref.[5], the challenge of accurate rejection diagnosis of new energy vehicles (NEVs) under various operating conditions is tackled by proposing deep neural network models. The main issue mentioned was the poor generalization of traditional diagnostic models when applied to new domains with limited data. To overcome this, the authors used a transfer learning approach that used knowledge from well-marked source domains and adapted to target domains with sparse error names. By integrating CNNs and domain tuning techniques, the frame significantly improved diagnostic performance and achieved greater accuracy compared to traditional deep learning methods. The authors proposed that future work should examine the architecture of light and real-time depreciation strategies.

A comprehensive review of eco-efficiency assessments of NEVs using life cycle analysis (LCA) is done in Ref.[9]. The authors compiled and compared several LCA data records, focusing on battery-electric vehicles (BEVs) and plug-in hybrids. The review highlighted key indicators such as greenhouse gas emissions, energy consumption, and the use of materials throughout the vehicle lifecycle. These benefits are partly advantageous due to battery production and the emissions generated at the end of the day. A key limitation was the lack of standardized LCA methods and real-time data integration, which limited the orientation of cross-study comparisons and guidelines. The authors sought a harmonized framework and dynamic LCA model to better inform sustainable transportation guidelines and improvements in NEV designs.

Characteristics and emerging trends of NEVs were examined in Ref.[10] using data from the Web of Science database. The research aims to map developments in this field, identify influential publications, and identify dominant research topics. By using bibliographic technology and visualization instruments such as Citespace, they analyzed over 1,000 articles over several years. This study focuses on battery safety, error diagnosis, and thermal management, with China and the US serving as key participants. The study also highlighted the increasing use of intelligent diagnostic techniques and interdisciplinary approaches. However, a significant limitation was that it relied solely on bibliographic data, which could overlook qualitative knowledge and new, unpublished developments.

Technological gaps in NEV concerning users expectations are investigated in Ref.[3]. The study integrated the results of various empirical articles, political documents, and research foundations in which data records were identified, focused on public opinion, market trends, and demographic behavior. Although the authors did not use direct methods for ML, they categorized the influential factors of economic incentives, infrastructure, environmental perception, and social norms motivation. A key outcome of the review was the identification of political effectiveness that relies heavily on consumer demographics and perceptions. However, the limitation of this study lies in its reliance on existing literature without quantitative validation or predictive modeling. This indicates the need for future ML-controlled mood analysis or predictions based on real-time behavioral data.

The study[11] examined the impact of consumer knowledge on the adoption of EVs in an emerging market context focused on India. The study addresses the issue of low EV records despite increased environmental concerns and state incentives. Using survey data collected by potential EV users, the authors used structural equation modeling (SEM) to analyze the relationship between EV knowledge, perceived benefits, and acceptance intentions. The results showed that consumer awareness significantly improves perceptual value and thus positively influences the likelihood of adoption. However, the study also highlighted important limitations, such as symptoms of knowledge, inadequate charging infrastructure, and the need for clearer communication regarding the benefits of EVs. Although no ML techniques were used, statistical modeling provided robust insight into behavioral predictors. The authors propose political interventions and awareness campaigns to close information gaps and accelerate the adoption of EVs.

The authors[12] provide an in-depth overview of environmental and operation variability (EOV) compensation. The study emphasizes the importance of adaptive modeling, reference-free methods, and robust feature extraction to ensure consistent fault recognition performance across diverse operational conditions. Building on these insights, this work aims to develop and validate an approach that maintains accuracy under variable environmental influences.

By reviewing these recent articles from existing literature, the following key research gaps have been identified:

- Existing fault diagnosis models for NEVs predominantly rely on individual ML and DL models, with limited research exploring ensemble approaches to enhance accuracy and robustness.
- Most current models fail to account for the dynamic nature of fault progression, which could be more effectively captured by advanced deep learning architectures such as GRUs and CNNs.
- There is a lack of integration between domain-specific knowledge and data-driven ML techniques, which is essential for addressing the limited availability of labelled fault data and improving the practical reliability of NEV fault detection systems.
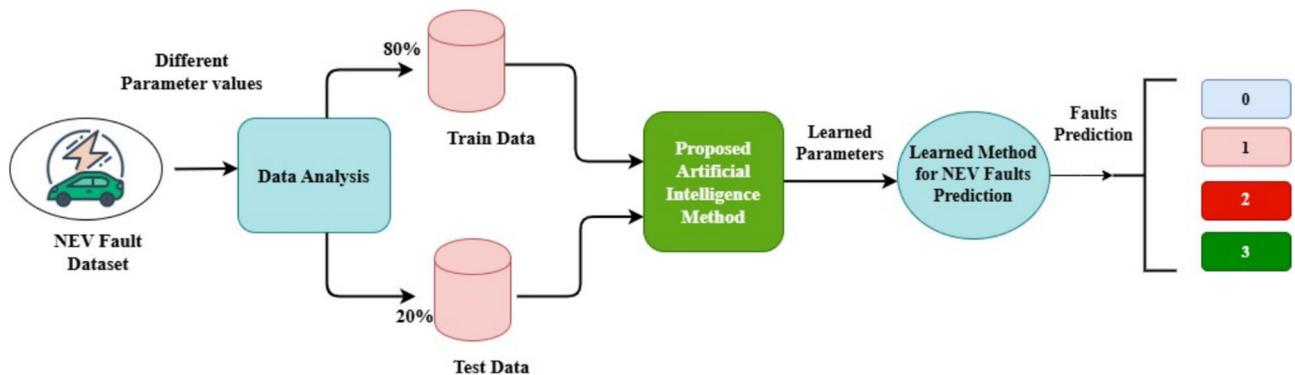
Fig. 1. Abstract workflow analysis for NEVs fault detection.

Label	Fault description
0	Normal operational state
1	Motor fault
2	Inverter fault
3	Battery fault

Table 1. Fault label definitions in NEV dataset.

Proposed methodology

This section describes the methodology we propose for predicting the performance and reliability of NEVs. We provide a comprehensive discussion of the techniques used for analysis and result computation. The abstract workflow analysis for NEVs' fault detection is shown in Fig. 1. Initially, the dataset consisted of numerous high-dimensional features, making it suitable for thorough testing. To tackle this complexity, we have developed a new strategy to select properties that focus on maintaining only attributes that have a significant impact on NEV performance predictions. The selected feature set is split into two subsets: 80% in training and 20% in testing. We then trained and evaluated several advanced ML models. The performance of each model is evaluated based on invisible test data, and models that exhibit the highest accuracy are selected for further prediction and analysis of NEV performance[13].

Dataset collection

For experimental analysis, we utilize the publicly available Kaggle Fault Diagnosis Dataset for NEVs, specifically curated to support fault detection in drivetrain components. The dataset contains approximately 11,000 rows of high-resolution sensor readings. These include parameters such as voltage, current, motor speed, temperature, vibration, ambient temperature, and humidity, each critical for identifying faults in key subsystems such as the motor, inverter, and battery. Each data entry is annotated with a fault label representing the operational condition of the NEV. The classification schema is outlined in Table 1.

Figure 2 shows the distribution of the samples across various classes. It shows that the 'normal' class has the highest number of samples, i.e., 5,000, while the rest of three classes have 2,000 samples each.

Data preprocessing

Data preprocessing ensures that the input data is clean, consistent, and ready for modelling. In this case, all sensor values and performance specifications are systematically checked for missing values to ensure that there are no distortions or calculation errors in the modelling process. Therefore, the dataset is fully proven. There is no need to credit or delete records. Continuous variables are normalized to the standard range, and outliers are recognized and processed accordingly. These steps in preprocessing create a solid foundation to follow and select modelled properties. The number of missing values in each feature is calculated as:

$$\text{Missing Value Count} = \sum_{i=1}^{n} [y_i = \text{null}] \tag{1}$$

where y_i represents value of the feature, and n is total number of instances.

To protect against information leakage, all preprocessing steps were performed strictly within the training folds of data. Particularly, the dataset was split into training (80%, 8,800 samples) and test (20%, 2,200 samples) subsets using stratified sampling to protect the original class proportions. Feature scaling was then used only on the training set using the MinMax normalization techniques, defined as:
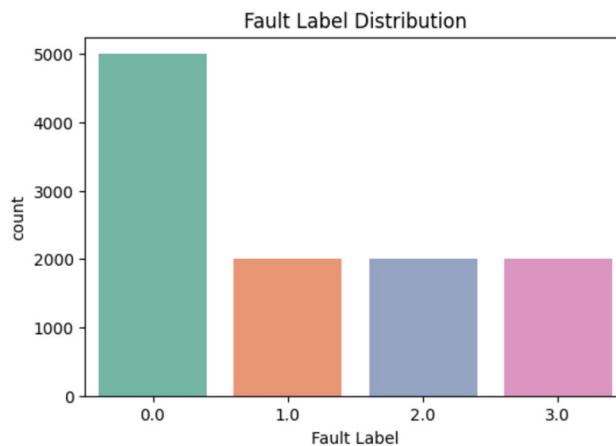
Fig. 2. The NEV faults prediction analysis's target distribution analysis.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{2}$$

where x shows the original feature value, and $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of that feature within the training data. The same fitted transformation was subsequently applied to the test set, ensuring that no information from the test distribution influenced model training or variable estimation.

Feature engineering
Feature engineering involves the correlation analysis of features to determine the best features to predict NEV performance. A correlation heatmap is obtained which depicts the linear relationships among sensor variables and performance labels. The predictive accuracy and efficiency of ML models are thereby enhanced by feature selection, which concentrates on these highly correlated features. This ensures that the models focus on the most informative signals, which is critical for an accurate NEV performance diagnosis[14].

Dataset splitting
For evaluating performance of the models, the dataset was divided into a training set of 80% and an independent test set of 20% using stratified sampling to protect class proportion. The test set was set aside and used only for final evaluation. Within the training set, stratified 10-fold cross-validation was employed to normalize and validate the models. For each fold, 90% of the training subsets were used for training and 10% for validation. Data preprocessing (normalization and feature scaling) was performed sequentially within each fold to prevent target leakage. Model accuracy, confusion matrix, and classification metrics were reported on the independent test set, while the average and standard deviation of cross-validation accuracy were used to assess the robustness of the models. This approach ensured that the test set remained untouched during model development, while cross-validation within the training set was used exclusively for model selection and hyper-parameter training[15].

Models used in study
Data structures are used to reduce and improve model prediction reliability. Recently, ML has been used to predict the performance and reliability of NEVs. This approach applies advanced ML and DL models for the precise and timely assessment of NEV performance indicators. An ML-based system can assist engineers and stakeholders by providing actionable insights for NEV performance optimization and fault detection. Stepwise workflow analysis of applied models in NEVs' fault detection is shown in Figure 3.

Logistic regression
A frequently used method for tasks involving both binary and multiclass categorization is logistic regression (LR). It primarily helps to forecast battery condition, detecting whether a battery is functioning optimally or requires maintenance, within the context of NEVs. Voltage, temperature, and charging history are among the characteristics the model uses. Using the sigmoid function, the model turns these characteristics into probability estimates between 0 and 1. This probability is then applied to provide unambiguous judgments on battery performance[16].

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(w \cdot X + b)}} \tag{3}$$

where,

- $P(y = 1 \mid X)$ is the probability of the positive class.
- X is the feature vector.
- w is the weight vector.
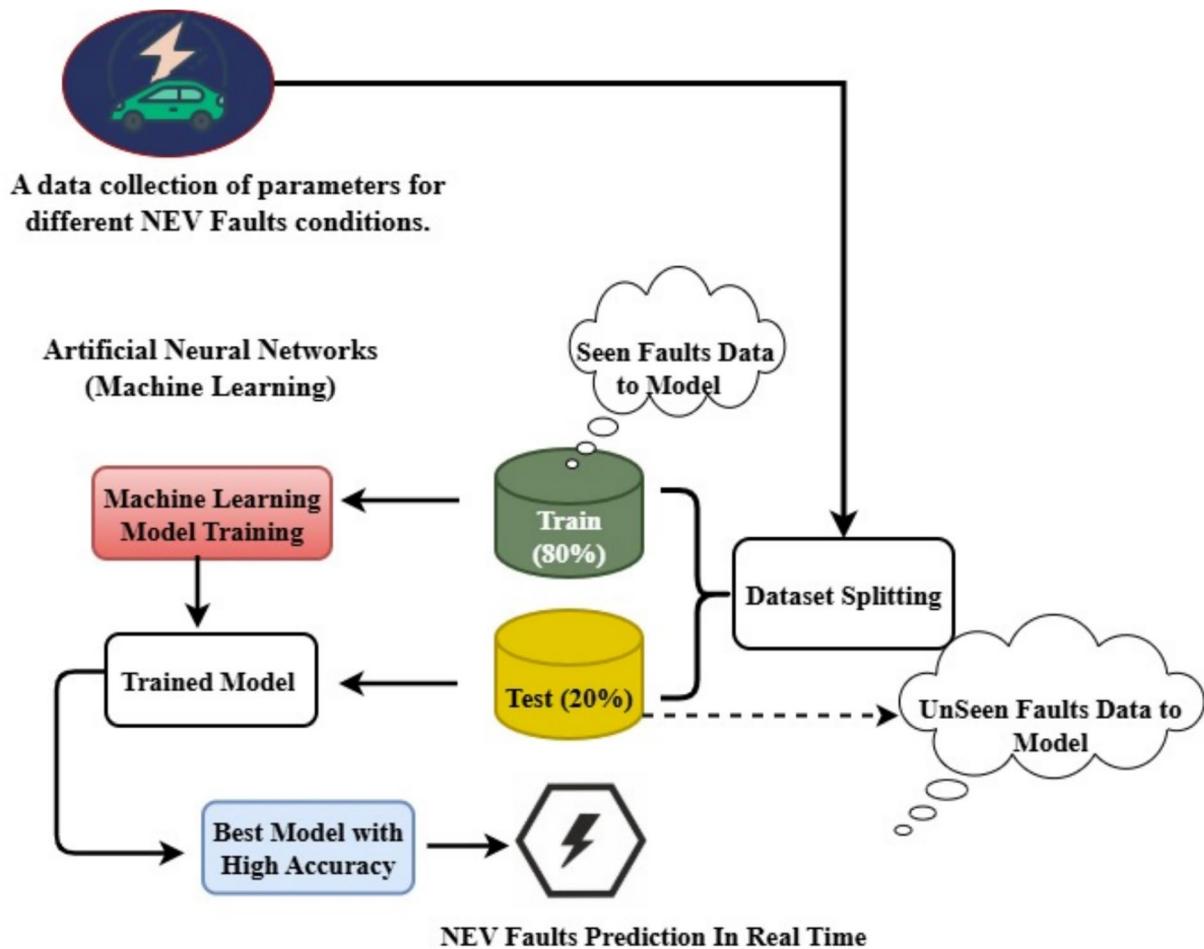
Fig. 3. Stepwise workflow analysis of applied models in NEVs' fault detection.

- b is the bias term.
- e is the base of the natural logarithm.

Ridge classifier
Building on LR, the Ridge classifier uses L2 regularization to reduce overfitting, especially when input features are highly correlated, therefore improving performance and predicting demand for charging stations, where environmental and user behavior data are occasionally connected, benefits especially from this. Penalizing large coefficients, Ridge guarantees more consistent predictions even with overlapping or noisy data. For managing large-dimensional datasets, which are typical in NEV applications, the Ridge classifier is highly successful[17,18].

$$Y_i = w \cdot X_i + b \tag{4}$$

where,

- Y_i is the actual label.
- X_i is the input feature vector.
- w is the weight vector.
- b is the bias term.
- λ is the regularization parameter that controls the penalty on large weights.

Perceptron
Based on sensor data such as voltage, temperature, and electricity, a perceptron can classify NEV fault types. NEV systems are built in real time and benefit from their simplicity and inexpensive computing costs. It is fighting nonlinear data that could limit performance in more sophisticated cases, including multifactorial errors. Despite its drawbacks, the Perceptron remains a vital starting model for benchmarking in error recognition systems[1].

The Perceptron formula is given by:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \qquad (5)$$

where x_i is input features, w_i represents weights, b is bias, f is activation function (commonly the step function or sign function), and y is output prediction (class label).

Passive aggressive classifier
An online learning algorithm is particularly suitable for real-time classification tasks, such as NEV error detection. In contrast to traditional batch learning algorithms, passive aggressive (PA) works gradually and only adapts when the model makes incorrect predictions. The "passive" aspect refers to the model remaining unchanged when the prediction is correct, while "aggressive" signifies rapid weight adjustment when the model is incorrect[15]. Hinge loss is calculated using.

$$L(w; (x_i, y_i)) = \max(0, 1 - y_i \cdot (w \cdot x_i)) \qquad (6)$$

The weight update for the PA classifier is done using.

$$w_{t+1} = w_t + \tau \cdot y_i \cdot x_i \quad \text{where} \quad \tau = \frac{\max(0, 1 - y_i \cdot (w \cdot x_i))}{\|x_i\|^2 + \lambda} \qquad (7)$$

Gated recurrent unit
The GRU is designed to efficiently capture temporal dependencies in sequential data[19]. In this study, the GRU model is applied to structured sequences of sensor readings, where raw features are split into fixed-length windows of 10 times steps with a 50% overlap to preserve continuity. Each input sequence thus comprised multivariate feature channels such as voltage, current, and temperature aligned along the feature dimensions, while the corresponding label is subjected to the final observation within each window to maintain causal consistency. Table 2 presents the architecture of the GRU network employed in this study. This architecture is especially suited for real-time NEV fault prediction, enabling early detection of plug-in failure and battery degradation under continuous monitoring scenarios.

Convolutional neural network
In NEVs, CNNS plays a vital role in thermal anomalies. These models analyze the heat distribution patterns of battery cells by processing thermal images and identifying important issues such as overheating and structural errors. CNN utilizes folding layers to automatically identify local patterns in images, making it highly efficient for detecting anomalies in thermal data. By capturing these anomalies, CNN prevents catastrophic battery failures and allows early intervention to extend battery life, ultimately[20]. The data are first transformed into two-dimensional metrics using a time window of 10 samples with 50% overlap, where the temporal dimension showed a sequence of sensor readings and the positional dimension matched the set of input features. Each segment was marked according to the fault class occurring within that specific time interval. One-dimensional twist layers are then applied along the temporal axis to check localized transition and fault signature, followed by max pooling layers to compress the learned description while preserving key features. Table 3 shows the architecture that begins with one-dimensional convolutional layers. This hierarchical design enables the CNN to automatically identify critical fault patterns from raw input sequences.

Artificial neural network
A versatile model of ANNs is an excellent learner of complex patterns and relationships in data. ANN is often used to optimize energy consumption in NEVs. ANNS can adjust the car's settings to minimise energy loss by predicting energy consumption and analysing driving behaviour, terrain conditions, and battery efficiency. This results in better driving range, which is especially vital for NEVs when resolving range anxiety. By nonlinear transforms, ANNs process data using multiple layers of interconnected nodes, thereby modelling sophisticated interactions among input variables[17,21]. The neural network is represented by

Layer (Type)	Output Shape	Parameters
GRU (gru_12)	(None, 10, 128)	52,608
Dropout (dropout_25)	(None, 10, 128)	0
GRU (gru_13)	(None, 64)	37,248
Dropout (dropout_26)	(None, 64)	0
Dense (dense_44)	(None, 32)	2,080
Dense (dense_45)	(None, 4)	132
Total Parameters		**92,068 (359.64 KB)**
Trainable Parameters		**92,068 (359.64 KB)**
Non-trainable Parameters		**0 (0.00 B)**

Table 2. Analysis of GRU model architecture for fault detection in NEVs.

Layer (Type)	Output Shape	Parameters
Conv1D	(None, 10, 64)	1,408
Batch Normalization	(None, 10, 64)	256
Max Pooling1D	(None, 5, 64)	0
Conv1D	(None, 5, 128)	24,704
Batch Normalization	(None, 5, 128)	512
Max Pooling1D	(None, 2, 128)	0
Flatten	(None, 256)	0
Dense	(None, 128)	32,896
Dropout	(None, 128)	0
Dense (Output Layer)	(None, 4)	516
Total Parameters		**60,292**
Trainable Parameters		**59,908**
Non-trainable Parameters		**384**

Table 3. Analysis of CNN model architecture for fault detection in NEVs.

Layer (Type)	Output Shape	Parameters
Dense (dense_10)	(None, 512)	4,096
Dropout (dropout_4)	(None, 512)	0
Dense (dense_11)	(None, 256)	131,328
Dropout (dropout_5)	(None, 256)	0
Dense (dense_12)	(None, 128)	32,896
Dense (dense_13)	(None, 64)	8,256
Dense (dense_14)	(None, 32)	2,080
Dense (dense_15)	(None, 1)	33
Total Parameters	-	**178,689 (698.00 KB)**
Trainable Parameters	-	178,689
Non-trainable Parameters	-	0

Table 4. Layer-wise architecture of the ANN model.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (8)$$

Where,

- x_i: input features
- w_i: corresponding weights
- b: bias term
- f: activation function
- y: output of the neuron

The ANN model used in this study comprises a stack of fully connected (Dense) layers interleaved with Dropout layers to reduce overfitting, as shown in Table 4. The network begins with an input Dense layer of 512 neurons, followed by successive layers of decreasing size, 256, 128, 64, and 32 neurons, ultimately ending with a single-neuron output layer suitable for binary classification. The total number of trainable parameters is 178,689, and no non-trainable parameters are present. The architecture is lightweight, yet its shallow depth and lack of recurrent or convolutional structure contribute to its lower performance compared to GRU and CNN models.

Proposed ensemble GRULogX approach

The research proposes a new ensemble method, GRULogX, that combines GRU and LR and other prominent classifiers to improve fault detection in NEVs. The GRU itself is best suited to handle sequential data provided by the NEV's sensors, capturing temporal dependencies and patterns within drivetrain and battery subsystems. Logistic Regression is utilized next to improve fault classification through a probabilistic framework for classifying fault occurrences. Through synergistic integration, the strengths of each model, both in terms of fault detection under varying conditions and generalization and reliability, are ensured. This approach not only enhances detection accuracy level, but also shows superior *precision*, *recall*, and attains a remarkable 99% fault detection accuracy. The approach uses cross-validation and hyperparameter techniques in its ensemble

methodology, making it adaptable to test data and ensuring reliability for practical deployment as illustrated in Figure 4.

The GRU model captures sequential patterns in NEV sensor data. It uses the following equations for updating its states:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{9}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{10}$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t] + b_h) \tag{11}$$

$$h_t = (1 - z_t) \cdot \tilde{h}_t + z_t \cdot h_{t-1} \tag{12}$$

Where,

- z_t is the update gate controlling the influence of previous hidden states.
- r_t is the reset gate determining how much of the previous state should be forgotten.
- $\tilde{h}_t$ is the candidate hidden state, and h_t is the final hidden state at each time step.

After the GRU outputs the sequence representation, it is passed to the Logistic Regression classifier for fault prediction:

$$P(y = 1|X) = \frac{1}{1 + e^{-(W \cdot X + b)}} \tag{13}$$

Where,

- $P(y = 1|X)$ is the probability of a fault occurring.
- X is the feature vector obtained from the GRUs output.
- W is the weight vector, and b is the bias term.
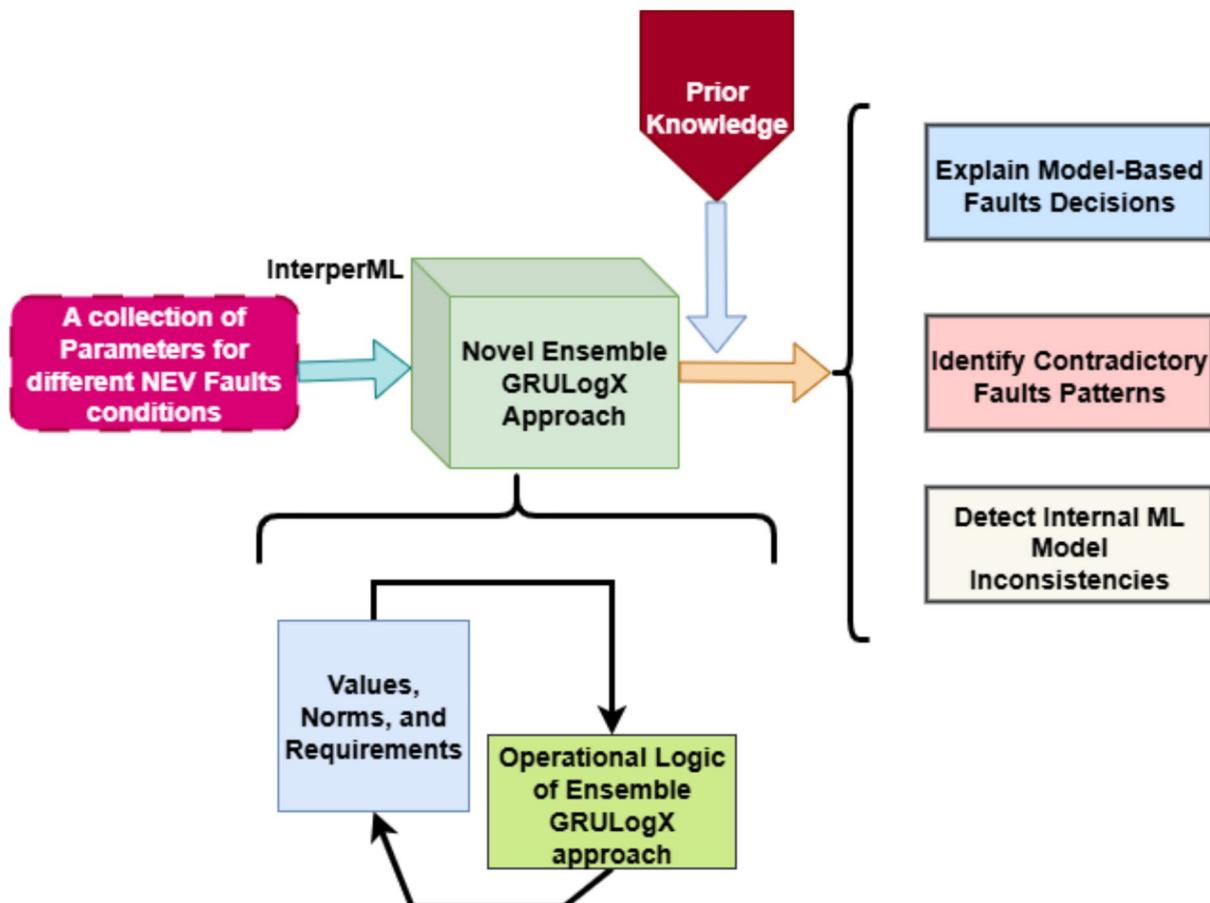
Fig. 4. Proposed architecture of novel ensemble GRULogX.

The final fault prediction is derived from the ensemble of GRU and Logistic Regression, along with other classifiers such as Perceptron and Ridge Classifier. The ensemble combines its outputs using a weighted sum:

$$f_{\text{final}} = \text{Softmax}\left(\sum_{i=1}^{n} \alpha_i f_i\right) \tag{14}$$

Where,

- f_i represents the output of each classifier.
- α_i are the weights assigned to each classifier's prediction.
- The final output f_{final} is passed through a Softmax function to obtain a probability between 0, 1,2, and 3 representing the likelihood of a fault.

The proposed GRULogX ensemble is implemented using a stratified 5-fold cross-validation scheme. Where each fold trained the base learners, GRU, and LR independently on the training subset and evaluated them on a distinct validation subset. The outputs of the two base models were combined through a probabilistic averaging strategy, defined as:

$$P_{\text{ensemble}} = \frac{1}{2}\left(P_{\text{GRU}} + P_{\text{LR}}\right) \tag{15}$$

where P_{GRU} and P_{LR} show the predicted class probabilities obtained from the GRU and LR models.

The GRU model utilized two stacked recurrent layers with 128 and 64 hidden units, and dropout regularization of 0.2 to prevent overfitting. The Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss is used for training. Logistic regression with L2 regularization and a maximum of 1000 iterations is employed as a fast linear classifier for probabilistic learning.

By incorporating these models, the GRULogX ensemble significantly enhances the accuracy and reliability of fault detection systems for NEVs, paving the way for more efficient and widespread adoption of clean energy vehicles.

Model hyperparameter settings

Appropriate hyperparameter selection is very important to obtain the best performance from ML and DL models. Selection of suitable hyperparameters can improve accuracy and recall performance, especially when dealing with complex data records, such as those from NEVs[5,15]. Each model used in this study, including logistic regression, passive-aggressive classifier, ridge classifier, perceptron, GRU, CNN, and ANN, must be carefully configured and tuned. These hyperparameters, such as the number of iterations, control strength, and activation functions, affect the learning dynamics and ultimately the ability of the model to generalize[22,23]. The following Table 5 summarizes the hyperparameter settings used in this study for each model.

Model's evaluation criteria

To evaluate the performance of the models, several key performance indicators are used, including R^2, RMSE, and MSE[22]. The metric root mean squared error (RMSE) measures the inconsistency between predictions and actual values, providing a comprehensive indication of the model's accuracy. The mean squared error (MSE) measures the average difference between the predicted and actual values, providing a detailed representation of the model's performance. The coefficient of determination (R^2) provides insight into the dependency of dependent variables. Accuracy metric represents the ratio of correct predictions to the total number of predictions. Precision measures the number of accurate positive predictions made by the model. Recall refers to the ratio of actual positive cases to all the actual positive cases. F1 score represents the harmonic mean of precision and recall, offering a balanced evaluation of both metrics.

These metrics are especially valuable in classification tasks, such as predicting NEV performance, where both precision and recall are equally crucial for reliable and accurate results[24].

Technique	Hyperparameters
Logistic Regression	random_state: 0, max_iter: 1000
Passive Aggressive Classifier	C: 0.5, max_iter: 1000, early_stopping: True, n_iter_no_change: 5, random_state: 0
Ridge Classifier	alpha: 1.0, solver: 'auto', tol: 1e-3
Perceptron	penalty: 'l2', alpha: 0.0001, max_iter: 1000, tol: 1e-3, random_state: 0
GRU Model	GRU Layer 1: units=128, activation='relu', dropout=0.2, GRU Layer 2: units=64, activation='relu', dropout=0.2, Dense Layer: units=32, Output Layer: units= 4, activation='softmax'
CNN Model	Conv1D Layer 1: filters=256, kernel_size=2, activation='relu', MaxPooling1D: pool_size=2, Conv1D Layer 2: filters=128, kernel_size=2, Dense Layer: units=32, Output Layer: units=4, activation='softmax'
ANN Model	Dense Layer 1: units=512, activation='relu', dropout=0.3, Dense Layer 2: units=256, activation='relu', dropout=0.3, Dense Layer 3: units=128, Dense Layer 4: units=64, Dense Layer 5: units=32, Output Layer: units=4, activation='softmax'

Table 5. Analysis of model hyperparameter settings for fault detection in NEVs.

Results and discussion

This section provides a detailed analysis of the application of ML to predict errors in NEVs. We will thoroughly examine the empirical findings of the research and provide a detailed explanation of their effects. Our results demonstrate the accuracy and efficiency of each method, measured with different performance parameters. This study highlights the potential of ML models to improve error detection, improve vehicle reliability, and introduce clean energy transport solutions.

Data exploration

This study examined the relationships between several variables in the dataset using an in-depth correlation matrix analysis. Understanding how components work together benefits from this approach. By analyzing the correlations, we can recognize patterns that are very important for predicting NEVs. Correlation matrices provide information on the orientation and intensity of these interactions, which can lead to the selection and determination of characteristics for modeling. The correlation matrix displays linear relationships between various characteristics within a data record, as illustrated in Fig. 5. Each cell in the matrix reflects the connection between the two properties. Red colours refer to positive correlation. Blue suggests negative. Engine speed (rpm) and error labels show a robust positive correlation of 0.55. This means that the probability of error increases as engine speed increases. Conversely, the voltage (v) shows a significant negative correlation of -0.84 with the error label. This indicates that a higher probability of error is associated with a drop in voltage levels. Understanding the characteristic dependencies of NEV error detection and improvements in predictive models depends on these observations from the correlation matrix[24].

Experimental settings

Using Python 3.0, we conducted our studies on forecasting NEV performance utilizing sophisticated ML techniques. Google Colab, an open-source platform with a GPU backend, 13 GB of RAM, and 90 GB of storage, enabled us to undertake these investigations. Runtime computation, accuracy, precision, recall, and F1 score were used to evaluate the performance of our NEV prediction model[24].

Performance analysis of applied machine learning models

This section describes the evaluation of services to predict errors in NEVs with all characteristics. All ML models are tested based on entire dataset. The performance of the error detection model is assessed by accuracy, accuracy, recall, and F1 score. These metrics are important to recognize how each model can reliably categorize faulty and normal observations. The results show that ML methods for classification accuracy and other performance parameters are superior to deep learning methods[24]. The models being evaluated are logistic regression, passive-aggressive classifier, ridge classifier, and perceptron, as shown in Table 6. In all categories, Logistic Regression provided the best performance, achieving a precision, recall, F1 score, and accuracy of 0.98. Passive Aggressive Classifier performed immediately after with steady results of 0.96 precision, recall, F1 score, and 0.96 accuracy. Ridge Classifier slightly lagged with 0.94 accuracy, while the perceptron model significantly underperformed,
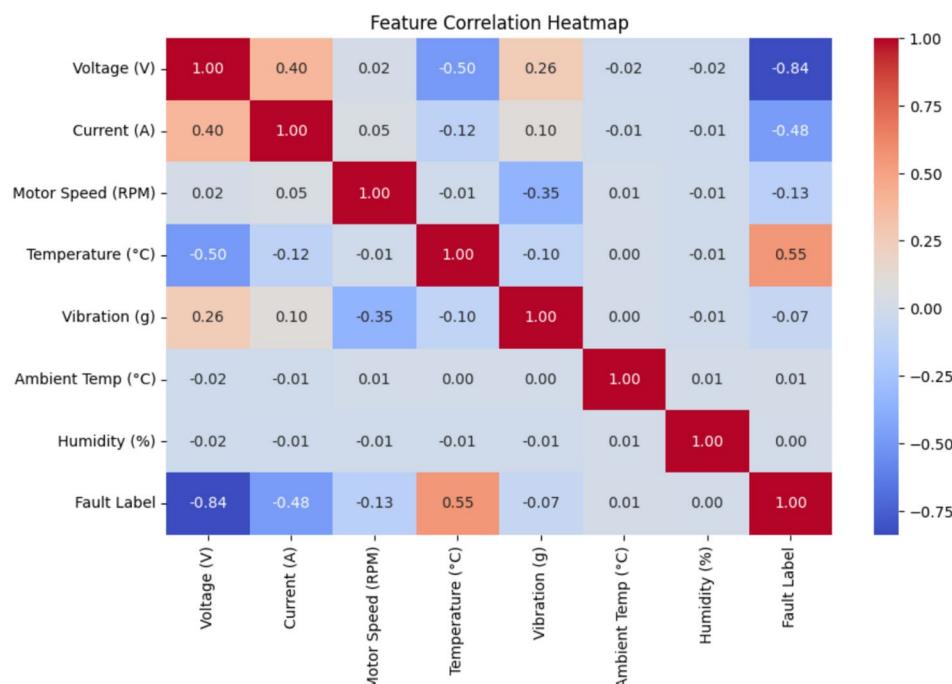

Fig. 5. Feature correlation heatmap showing the correlation between different sensor variables and the fault label.

Method	Accuracy	Target Class	Precision	Recall	F1 Score
Logistic Regression	0.98	0	0.99	1.00	0.99
		1	1.00	0.99	1.00
		2	0.93	0.95	0.94
		3	0.96	0.93	0.95
		Average	**0.98**	**0.98**	**0.98**
Passive Aggressive	0.96	0	0.97	0.98	0.98
		1	0.99	0.99	0.99
		2	0.88	0.95	0.91
		3	0.98	0.89	0.93
		Average	**0.96**	**0.96**	**0.96**
Ridge Classifier	0.93	0	0.93	0.99	0.96
		1	0.97	0.97	0.97
		2	0.97	0.77	0.86
		3	0.90	0.92	0.91
		Average	**0.94**	**0.94**	**0.93**
Perceptron	0.84	0	0.78	1.00	0.88
		1	0.93	0.88	0.90
		2	0.97	0.40	0.56
		3	0.94	0.87	0.91
		Average	**0.87**	**0.85**	**0.83**

Table 6. Performance comparison of applied ML models.

receiving metrics of 0.87, 0.85, 0.83, and 0.85 in precision, recall, F1 score, and accuracy, respectively. These results emphasize the superior performance of traditional ML models over DL methods, especially for this dataset.

The confusion matrices for four classification models, LR, PA classifier, Ridge classifier, and perceptron, demonstrate varying levels of performance across four target classes as shown in Figure 6. LR achieved the highest accuracy with minimal misclassifications, particularly strong in predicting class 0 and class 1. The PA classifier model showed comparable performance but exhibited more confusion between class 0 and class 2. Ridge classifier performed well overall but struggled with class 2, showing frequent misclassifications into class 3. The perceptron model exhibited the weakest performance, particularly for classes 1 and 3, with significant prediction errors. Overall, LR proved to be the most reliable model for multi-class classification in this study.

Performance of applied deep learning models

Evaluation of the applied deep learning models, GRU, CNN, and ANN, revealed that the GRU model outperformed the others with an accuracy of 98%, and high precision, recall, and F1-score values as shown in Table 7. The CNN model also demonstrated strong performance with consistent metrics across all evaluation parameters, achieving 97% accuracy. In contrast, the ANN model exhibited significantly lower performance, with an accuracy of only 28% and notably weak precision, recall, and F1 score. These results indicate that recurrent architectures, such as GRU, are highly effective for sequence-based classification tasks in this context. At the same time, ANN lacks the capacity to model complex patterns in the data.

Figure 7 presents the confusion matrices for three deep learning models, GRU, CNN, and ANN, applied to predict faults in NEVs. Each matrix compares the true class labels with the predicted labels for a multi-class classification problem involving four classes. The GRU model exhibits strong performance, with minimal misclassifications, particularly for classes 0 and 1, and very few errors for classes 2 and 3. The CNN model performs similarly well, with some misclassifications occurring in class 0 and class 3, but it still successfully predicts the majority of instances. The ANN model demonstrates solid accuracy for classes 0 and 1, although some class 2 instances are misclassified as class 3, indicating areas where the model can be improved. These confusion matrices provide insights into each model's ability to distinguish between classes and highlight areas that could benefit from further refinement.

Result of proposed GRULogX approach

The performance of the proposed GRULogX approach was evaluated across multiple fault detection categories. The results, which include precision, recall, and F1 scores, are summarised in Table 8. These metrics provide an indication of the model's effectiveness in identifying different fault classes and its overall performance.

The confusion matrix shown in Fig. 8 illustrates the excellent classification performance of the proposed GRULogX model. The model achieved perfect predictions for Classes 0 and 1, with zero misclassifications. Class 2 was accurately predicted with 390 correct classifications, and only nine samples were misclassified as class 3. Similarly, class 3 had 384 correct predictions with minimal confusion 14 samples predicted as class 2 and one as class 0. These results reflect high precision, recall, and F1 score, confirming the model's effectiveness in multiclass prediction tasks.
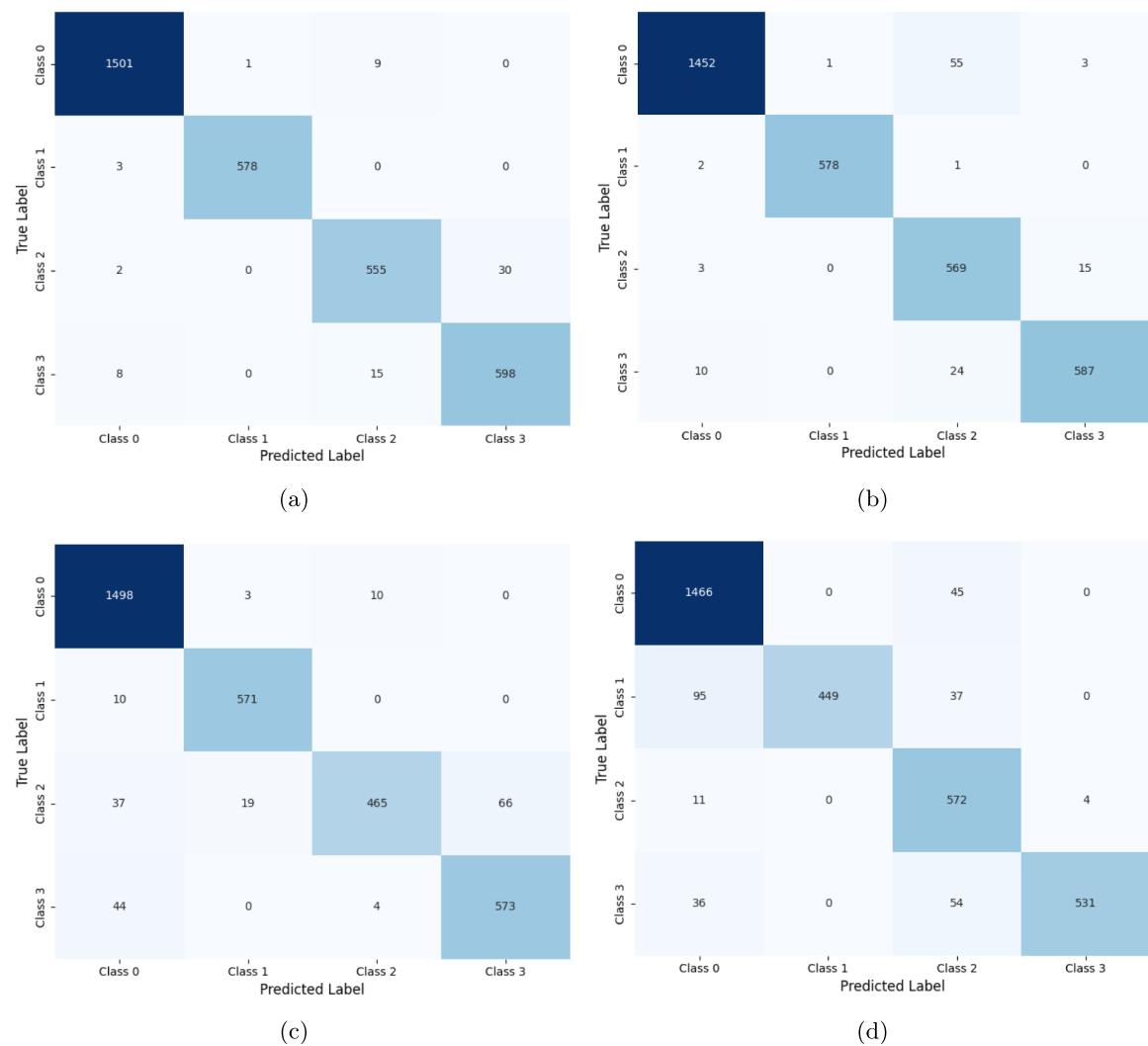
Fig. 6. Confusion matrices of applied ML models, (**a**) Logistic regression, (**b**) Passive aggressive, (**c**) Ridge classifier, and (**d**) Perceptron.

Method	Accuracy	Target Class	Precision	Recall	F1 score
GRU	0.98	0	1.00	1.00	1.00
		1	1.00	1.00	1.00
		2	0.95	0.97	1.00
		3	0.98	0.95	0.96
		Average	**0.98**	**0.98**	**0.98**
CNN	0.97	0	0.99	1.00	0.99
		1	0.99	0.98	0.99
		2	0.93	0.99	0.96
		3	0.98	0.91	0.94
		Average	**0.97**	**0.97**	**0.97**
ANN	0.28	0	0.96	0.22	0.36
		1	0.20	0.98	0.34
		2	0.00	0.00	0.00
		3	0.00	0.00	0.00
		Average	**0.47**	**0.28**	**0.22**

Table 7. Performance comparison of deep learning models.

(a)

(b)

(c)
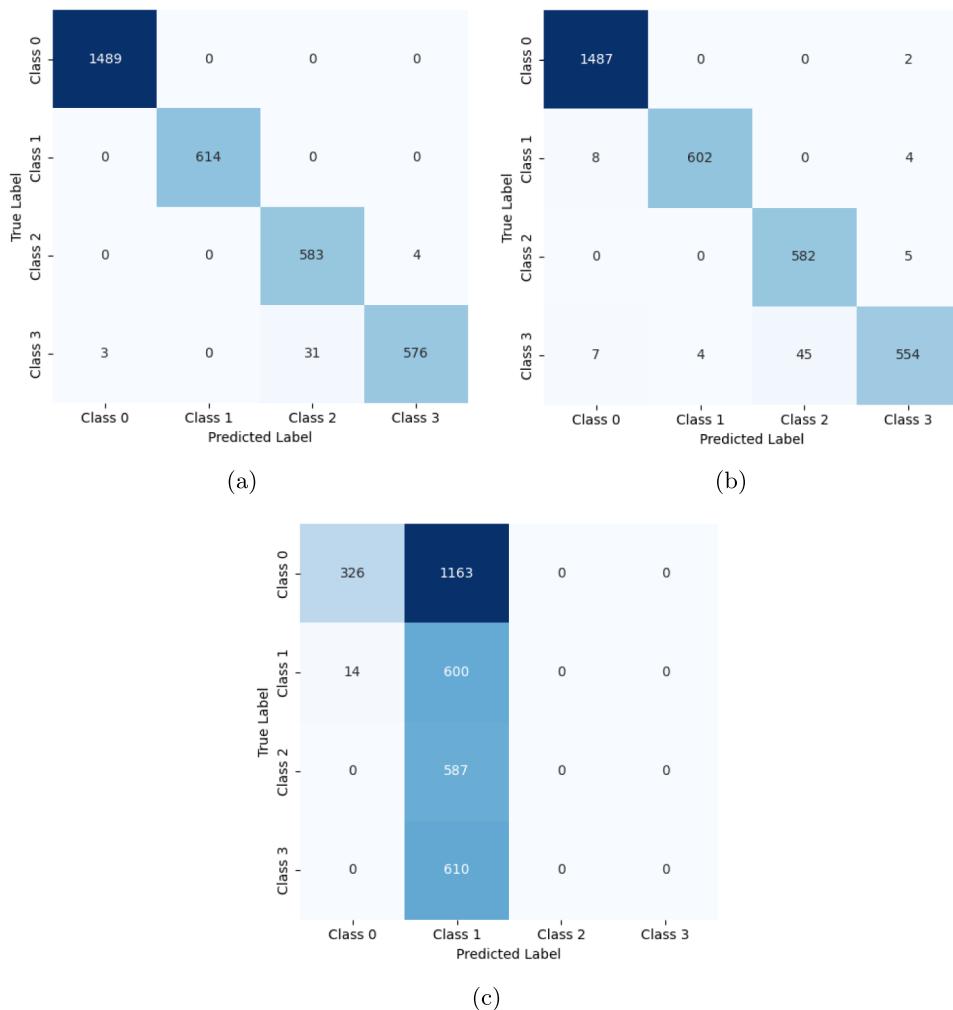
Fig. 7. Confusion matrices of DL models, (**a**) Gated recurrent unit, (**b**) Convolutional neural network and (**c**) Artificial neural network.

Model	Accuracy	Target Class	Precision	Recall	F1 score
GRULogX	0.99	0	1.00	1.00	1.00
		1	1.00	1.00	1.00
		2	0.98	0.98	0.98
		3	0.98	0.97	0.98
		Average	**0.99**	**0.99**	**0.99**

Table 8. Performance metrics for GRULogX fault detection.

Figure 9 presents a histogram comparison of various ML models based on four key performance metrics: accuracy, precision, recall, and F1 score. Notably, the GRULogX ensemble model demonstrates superior performance across all metrics, achieving the highest accuracy, precision, recall, and F1 score, thereby outperforming the other evaluated models.

Figure 10 presents the Receiver Operating Characteristic (ROC) curves for the applied ML and DL models, consisting of LR, Perceptron, PA, and the proposed GRULogX ensemble model. To ensure a fair evaluation across the multi-class fault detection problem, the ROC curves were determined using macro-averaging, where the true positive rate (TPR) and false positive rate (FPR) are independently averaged over all fault categories. This technique treats each class equally, regardless of its sample size, giving an extensive reflection of model generalization across different types of faults.

As illustrated, all models showed strong classification performance. However, the GRULogX ensemble achieves a nearly perfect ROC curve with an AUC of 0.999, highlighting its superior ability to differentiate between different fault types in new energy vehicles. This consequence validates the model's robustness in
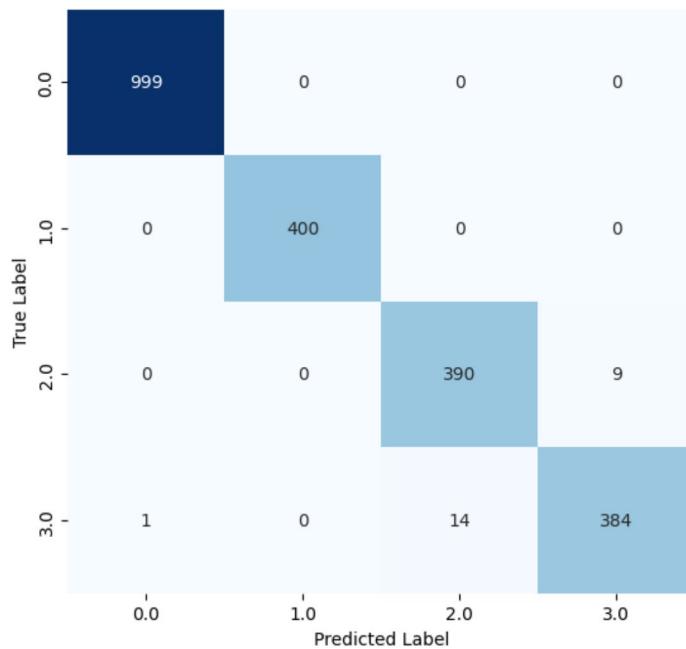
Fig. 8. Confusion matrix of the proposed GRULogX model for multiclass classification.
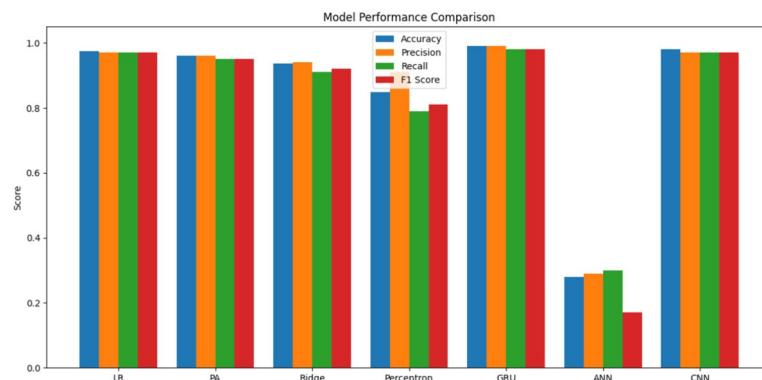
**Fig. 9**. Model performance comparison histogram for Accuracy, Precision, Recall, and F1 score.

handling complex temporal dependencies and sequential patterns of sensor data. The macro-averaged ROC indicates that this performance is consistent across all fault categories rather than being dominated by any single class.

Figure 11 presents a radar chart that compares the performance of different ML models. From the chart, it is evident that the GRULogX ensemble model demonstrates superior performance, achieving consistently high scores across all evaluation metrics: accuracy, precision, recall, and F1 score. LR and PA classifiers also perform competitively, with curves closely aligned and exhibiting strong results. This radar chart provides a comprehensive visual comparison, effectively highlighting the models' strengths and areas needing improvement, with GRULogX clearly standing out as the most robust model.

K-fold cross validation analysis

K-fold cross-validation to assess the performance of each applied method. Cross-validation analysis is based on K-fold accuracy and standard deviation. The data records for the selected function were split into 10 wrinkles for evaluation, providing a robust level of model output by testing its ability to generalize different subgroups of the model as shown in Table 9. The analysis revealed that the deep learning model exhibited poor performance and high standard deviation during validation, indicating greater variability and sometimes exceeding that of the training data. Logistics regression showed that the highest mean accuracy was 0.9756 with a low standard deviation of 0.0070.

The results show both high performance and stability. The passive-aggressive classifier demonstrated strong performance, achieving an accuracy of 0.9536, although its standard deviation of 0.0112 was slightly higher. The Ridge classifier and the perceptron achieved relatively lower accuracy, with medium accuracy of 0.9327 or
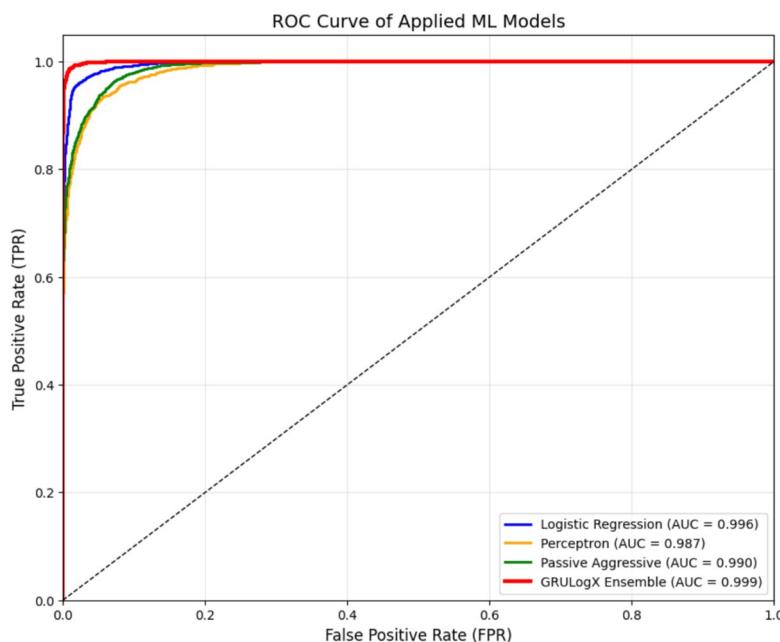
**Fig. 10.** ROC curves of applied models.
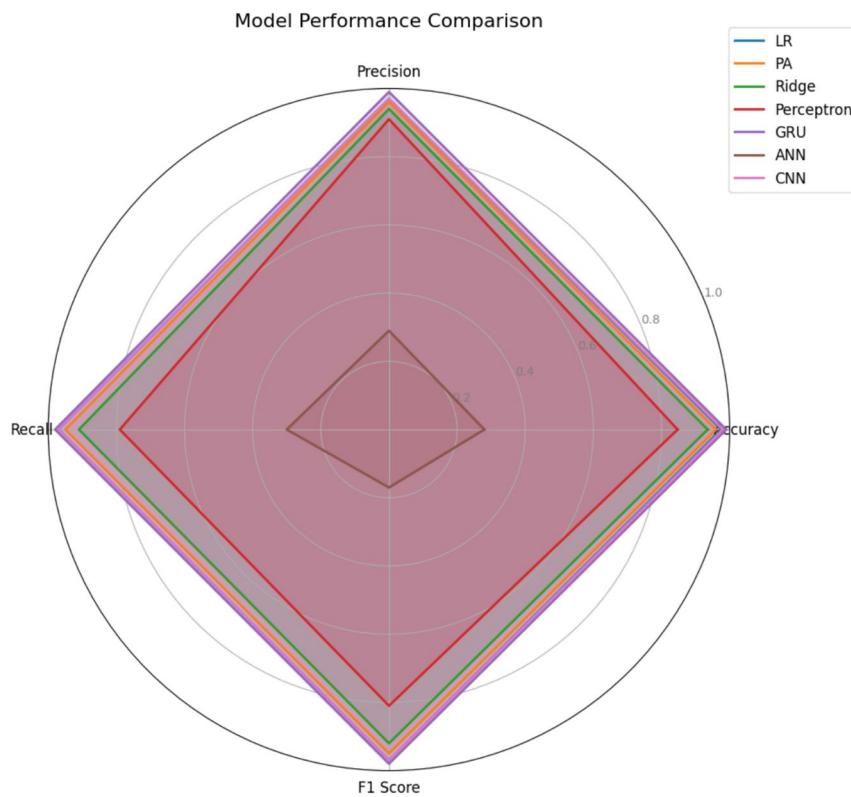
Fig. 11. Model performance comparison radar chart for Accuracy, Precision, Recall, and F1 score.

0.9153 and higher standard deviations of 0.0083 or 0.0189, which indicates that the performance increased the variability of the power in different folds. The ensemble GRULogX approach achieved a high mean accuracy of 0.99 with the lowest standard deviation of 0.0022. The bar diagram below shows the performance of various ML models related to medium accuracy and standard deviation during 10-fold cross-validation, as shown in Fig. 12.

Model	Mean accuracy	Standard deviation
Logistic Regression	0.9756	0.0070
Passive Aggressive	0.9536	0.0112
Ridge Classifier	0.9327	0.0083
Perceptron	0.9153	0.0189
Ensemble GRULogX	**0.9900**	**0.0022**

Table 9. K-Fold cross validation results for machine learning models.
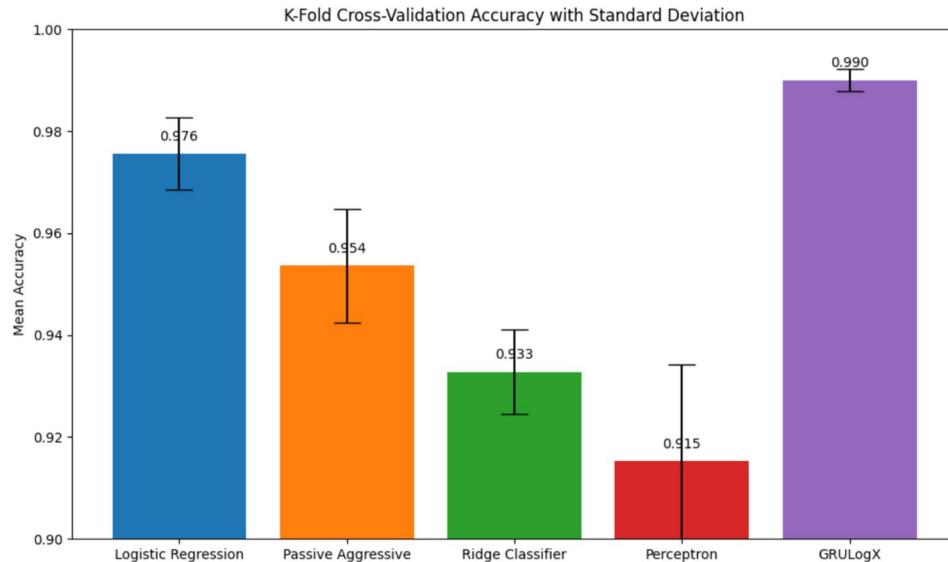
Fig. 12. K Fold cross-validation accuracy with standard deviation for various models.

Model	Runtime (s)
Logistic Regression	2.8899
Passive Aggressive	0.7963
Ridge Classifier	0.2140
Perceptron	0.4852
GRU	156.6400
ANN	145.8228
CNN	112.5900
Ensemble GRULogX	330.6200

Table 10. Computation runtime for various models.

Time complexity analysis

The proposed approach, GRULogX, which is an ensemble of GRU and LR, achieves a balance between high accuracy and efficient computation time, as shown in Table 10. Compared to individual models, GRULogX demonstrates an impressive computational efficiency, with a total runtime of 330.62 seconds. In comparison, traditional ML models such as LR require 2.8899 seconds, PA classifier takes 0.7963 seconds, and Ridge classifier completes in 0.2140 seconds, while perceptron requires 0.4852 seconds. When considering DL models, GRU takes 156.6400 seconds, ANN training time is 145.8228 seconds, and CNN training requires 112.59 seconds. Despite the longer computation time, the ensemble approach of GRULogX, which leverages both the GRU model's learning capability and Logistic Regression's efficiency, results in significantly higher accuracy. Therefore, GRULogX strikes an optimal trade-off, offering both high accuracy and a manageable computation runtime for complex tasks.

Performance error analysis

In this section, the results of the evaluation of ML models are presented based on the most important performance metrics: MSE, RMSE, R^2. These metrics were used to assess the accuracy and predictive performance of each

Model	MSE	RMSE	R² (R-squared)
Passive Aggressive	0.1215	0.3486	0.9106
Ridge Classifier	0.2285	0.4780	0.8319
Perceptron	0.1173	0.3425	0.9137
Logistic Regression	0.0567	0.2380	0.9583

Table 11. Model evaluation criteria for NEV fault detection.
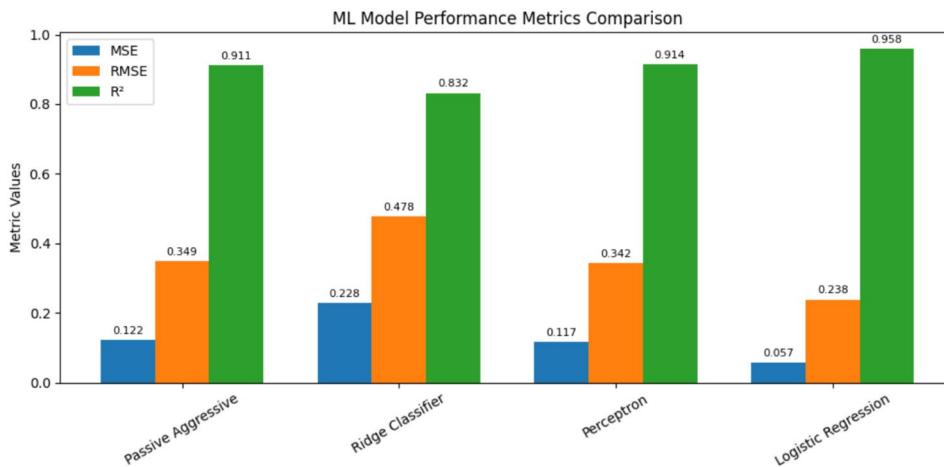
Fig. 13. MSE, RMSE, and R^2 for applied ML models.

Reference	Technique	Accuracy (%)
[7]	Statistical Methods	~85%
[8]	ML-based Classification	~88%
[5]	Transfer-based Deep Neural Network	~91%
Proposed	Ensemble GRULogX Approach of GRU and LR	**99%**

Table 12. Comparison of state-of-the-art approaches for NEV fault diagnosis.

model in error detection for NEVs. Table 11 presents performance metrics for each ML model evaluated in this study. The logistic regression model performed the best among all models, achieving the lowest MSE of 0.056667, the lowest RMSE of 0.238048, and the highest R^2 value of 0.958309. This shows that logistic regression provides the most accurate prediction with minimal error. On the other hand, the Ridge Classifier showed the best MSE and RMSE, indicating slightly poor performance compared to other models. However, since all models showed strong R^2 values, this suggests that everything was well explained to account for the variance of the target variables[22].

Figure 13 illustrates the differences in MSE, RMSE, and R^2 for each model, with Logistic Regression demonstrating the best performance in terms of the lowest MSE and RMSE, as well as the highest R^2.

State-of-the-art comparison

The comparison of state-of-the-art NEV fault diagnosis techniques is analyzed in Table 12. In the existing literature, several approaches have investigated the NEV's fault detection using the sensor data, being utilized in this study. For example Ref.[7] utilized statistical methods to achieve up to 85% accuracy for fault detection. Another work is by Li et al.[8], which uses ML models for the same task and reported an accuracy of up to 88%. The work from Wang et al.[5] is an important one and involves using transfer learning for NEVs fault detection; it reports a 91% accuracy. Compared to these works, the proposed approaches obtain 99% accuracy and outperform these works.

Discussions

The proposed GRULogX collaborative model demonstrates significant improvements in accuracy, sturdiness, and generalization for fault detection in new energy vehicles (NEVs). The main strength of this approach lies in its ability to effectively combine sequential feature extraction from the GRU with the probabilistic interpretability of Logistic Deterioration, providing balanced and consistent fault prediction outcomes. Compared with prior studies, this work presents a more comprehensive and adaptive solution to NEV fault detection.

The study[7] aimed at detecting control-related data faults in battery packs using a neural network trained on combined fleet data. While their method achieved successful classification of certain random faults, its dependency on large-scale, high-frequency data limited its adaptability to dynamic environmental or operational conditions. In the same way,[8] proposed a hybrid SPC–SVM-based diagnostic framework for early fault detection in lithium-ion batteries. Although their approach successfully identified abnormal behaviours at an early stage, it was less effective when dealing with shapeless or missing data and lacked real-time adaptability. In contrast,[5] employed a transfer learning-based deep neural network to improve fault diagnosis under varying operating conditions. While transfer learning enhanced generalization, the model was computationally intensive and required domain-specific tuning for new fault scenarios.

In contrast to the existing method, the GRULogX model addresses key limitations by integrating temporal learning and probabilistic decision making in a combined ensemble structure. Through probabilistic averaging of the GRU and Logistic Regression outputs, the model achieves improved stability, interpretability, and adaptability across varying EOV conditions. Additionally, the inclusion of stratified cross-validation ensures robust performance without overfitting to specific datasets. This hybrid ensemble framework offers a generalised, data-efficient, and computationally practical approach to NEV fault detection. The findings not only align with previous research but also advance it by presenting a validated ensemble mechanism that balances performance accuracy and real-time feasibility, advancing the state of the art in intelligent analytic systems for NEVs.

Conclusion and future work

This study successfully developed a robust, data-driven fault detection framework for new energy vehicles (NEVs) by integrating multiple machine learning and deep learning models. The proposed ensemble GRULogX model demonstrated good performance and achieved an accuracy of 99%, indicating high reliability in detecting faults within the complex drivetrain and battery systems of NEVs. By leveraging real-world sensor data and employing advanced cross-validation techniques, this research significantly improves fault diagnosis capabilities, which are crucial for enhancing the performance and safety of NEVs. The results emphasise the potential of this approach to reduce system failures and minimise downtime, thereby supporting the broader adoption of clean energy vehicles.

Future research could focus on real-time implementation and the integration of these fault detection models into onboard diagnostic systems within NEVs. This would facilitate continuous monitoring and enable immediate interventions to prevent failures. Additionally, exploring the use of hybrid models combining both machine learning and traditional engineering methods could further enhance the system's ability to handle a wider range of fault conditions. Further work could also look into expanding the dataset to include more diverse operating conditions, which would increase the model's generalization and accuracy across different types of NEVs.

Despite the strong performance of the proposed GRULogX ensemble model, certain limitations continue to exist. The dataset utilized in this study originates from a specific source, which may limit the generalizability of the results to other NEV types or fluctuating environmental and operational conditions. Moreover, the present work focuses on offline fault detection, without real-time implementation or hardware integration into onboard diagnostic systems. As a result, the system's efficiency under real-world dynamic driving conditions has not yet been fully validated. Addressing these limitations in future work through dataset expansion, real-time deployment, and the use of hybrid models combining both machine learning and traditional engineering methods will further enhance the model's adaptability for large-scale NEVs fault detection.

Data availability

The dataset used in this study is publicly available at the following link: https://www.kaggle.com/datasets/willianoliveiragibin/new-energy-vehicles.

References

1. Khaneghah, M. Z., Alzayed, M. & Chaoui, H. Fault detection and diagnosis of the electric motor drive and battery system of electric vehicles. *Machines* **11**(7), 713 (2023).
2. Zhao, X., Li, X., Wu, Y., Qiao, L., & Zhang, C. Assessment of the effects of china's new energy vehicle industry policies: from the perspective of moderating effect of consumer characteristics. Environment, Development and Sustainability, 1–22 (2023).
3. Jiang, Y., Wu, Q., Li, M., Gu, Y. & Yang, J. What is affecting the popularity of new energy vehicles? a systematic review based on the public perspective. *Sustainability* **15**(18), 13471 (2023).
4. Fan, Y. et al. Fault detection for li-ion batteries of electric vehicles with feature-augmented attentional autoencoder. *Sci. Rep.* **15**(1), 1–17 (2025).
5. Wang, Y. & Li, W. Transfer-based deep neural network for fault diagnosis of new energy vehicles. *Front. Energy Res.* **9**, 796528 (2021).
6. Liu, Z., Hao, H., Cheng, X. & Zhao, F. Critical issues of energy efficient and new energy vehicles development in china. *Energy Policy* **115**, 92–97 (2018).
7. Zhao, Y., Liu, P., Wang, Z. & Hong, J. Electric vehicle battery fault diagnosis based on statistical method. *Energy Procedia* **105**, 2366–2371 (2017).
8. Li, X. & Wang, Z. A novel fault diagnosis method for lithium-ion battery packs of electric vehicles. *Measurement* **116**, 402–411 (2018).
9. Wang, N. & Tang, G. A review on environmental efficiency evaluation of new energy vehicles using life cycle analysis. *Sustainability* **14**(6), 3371 (2022).
10. Zhao, X., Wang, S. & Wang, X. Characteristics and trends of research on new energy vehicle reliability based on the web of science. *Sustainability* **10**(10), 3560 (2018).

11. Jaiswal, D., Kant, R., Singh, P. K. & Yadav, R. Investigating the role of electric vehicle knowledge in consumer adoption: evidence from an emerging market. *Benchmark. Int. J.* **29**(3), 1027–1045 (2022).
12. Rezazadeh, N., De Luca, A., Perfetto, D., Salami, M.R., Lamanna, G.: Systematic critical review of structural health monitoring under environmental and operational variability: approaches for baseline compensation, adaptation, and reference-free techniques. Smart Materials and Structures (2025).
13. Dettinger, F. et al. *Machine-Learning-Based Fault Detection in Electric Vehicle Powertrains Using a Digital Twin* (Technical report, SAE Technical Paper, 2023).
14. Li, X., Peng, Y., He, Q., He, H. & Xue, S. Development of new-energy vehicles under the carbon peaking and carbon neutrality strategy in china. *Sustainability* **15**(9), 7725 (2023).
15. Wang, J. & Chen, J. Rapid diagnosis of power battery faults in new energy vehicles based on improved boosting algorithm and big data. *Energy Inform.* **7**(1), 139 (2024).
16. Ju, Y., Zhu, Y., Zhou, H., Ge, S. & Xie, H. Microwave pyrolysis and its applications to the in situ recovery and conversion of oil from tar-rich coal: An overview on fundamentals, methods, and challenges. *Energy Rep.* **7**, 523–536 (2021).
17. Zhang, H. et al. Large language model based intelligent fault information retrieval system for new energy vehicles. *Appl. Sci.* **15**(7), 4034 (2025).
18. Li, Z. et al. Research on new energy vehicle charging prediction based on monte carlo algorithm and its impact on distribution network. *Front. Energy Res.* **11**, 1269041 (2023).
19. Shen, Y., & Zhang, X. The impact of artificial intelligence on employment: the role of virtual agglomeration. Humanities and Social Sciences Communications **11**(1) (2024).
20. Zhao, X. et al. Progress in active infrared imaging for defect detection in the renewable and electronic industries. *Sensors* **23**(21), 8780 (2023).
21. Lin, B. & Shi, L. Identify and bridge the intention-behavior gap in new energy vehicles consumption: Based on a new measurement method. *Sustain. Prod. Consump.* **31**, 432–447 (2022).
22. Li, J., Zhang, Y., Chen, W. & Liu, Z. Performance and reliability prediction of new energy vehicles using regression-based machine learning techniques. *Energy Rep.* **9**, 462–475 (2023).
23. Liu, M. Fed-bev: A federated learning framework for modelling energy consumption of battery electric vehicles. In: 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), pp. 1–7 (2021). IEEE.
24. Liu, H., Song, X. & Zhang, F. Fault diagnosis of new energy vehicles based on improved machine learning. *Soft. Comput.* **25**(18), 12091–12106 (2021).

Author contributions

IA conceptualization, data curation, writing manuscript. MN formal analysis, conceptualization, writing manuscript. US methodology, formal analysis, data curation. KM software, project administration, methodology. AR visualization, software, investigation. IDN funding acquisition, visualization, investigation. SGV visualization, formal analysis, validation. IA supervision, validation, writing and editing the manuscript. All authors reviewed the manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to U.S. or K.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.