Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

AI-Powered Fault Detection in Gas Turbine Engines: Enhancing Predictive Maintenance in the U.S. Energy Sector

Mohammad Hamid Hasan Amjad[1], Bivash Ranjan Chowdhury[2], Syed Ali Reza[3], Md Shihab Sadik Shovon[4], Mitu Karmakar[5], Md Raisul Islam[6], Mehedi Hasan Ridoy[7], Md Anisur Rahman[8], Sangida Jahan Ripa[9]

Abstract

In the dynamic energy sphere in the U.S., operational dependability of gas turbine engines becomes vital to the continuous power production and sustenance of the national grid. Peak and baseload power production, the mainstay of gas turbines, is exposed to high thermal, mechanical, and chemical stresses that lead to wear out of components over time. Traditionally, maintenance practices have either been reactive or time-based, based upon fixed intervals, which has led to premature replacement of components, or vice versa, undetected degradation leading to catastrophic failure. Such approaches carry high operating costs, lower turbine availability, and jeopardize plant safety. The central objective of this research is to build and test an AI-driven fault detection system to identify early signs of failure in gas turbine engines and apply it specifically to deployment within the energy sector of the U.S. Improving gas turbine performance is essential to raising the cost-effectiveness and sustainability of power production systems. The dataset used in this research entails high-resolution operational parameters gathered from different industrial gas turbines operating within U.S. energy facilities. Engine operational data includes multiple time-dependent measurements that monitor essential parameters like turbine temperature at the inlet and outlet and rotational speed, torque measurements alongside vibration levels and power output, fuel rate and intake pressure, and exhaust temperature and oil temperature. Real-time data acquisition from embedded sensor arrays allowed researchers to track turbine performance throughout changing operational states at sub-minute data points. A centralized time-stamping system maintains channel synchronization, thus allowing analysts to draw accurate conclusions about operational states throughout the recorded period. To support strong and interpretable fault detection, we utilized a variety of machine learning models, each chosen for its specific strength at discriminating operations from fault conditions in gas turbines. We used a multi-metric-based evaluation approach that combined statistical validity with operational applicability to assess each model's fault detection capability. Logistic Regression attained the highest accuracy, followed very closely by Random Forest. XG-Boost attained the lowest accuracy of all three algorithms. The use of AI-driven fault detection under predictive maintenance has the potential to revolutionize U.S. power plants using gas turbines as the primary source of electricity generation. In the fiercely competitive and heavily regulated environment of the U.S. energy industry, fault anticipation presents the key to competitive advantage. Using AI-based diagnostics reduces manual checks and simplifies the servicing process by prioritizing technician resources to confirmed at-risk components. AI-powered fault detection is critical to improving grid resilience within the U.S. energy infrastructure by assuring that peak-demand-balancing and grid-stabilizing gas turbines remain fault-free.

Keywords: *Predictive Maintenance, Gas Turbine Fault Detection, Artificial Intelligence, Machine Learning, U.S. Energy Sector, Condition Monitoring, Data Analytics, Vibration Analysis, Failure Prediction, Energy Infrastructure Resilience.*

Introduction

Gas and steam turbines are instrumental elements in the production of electricity, each functioning on diverse thermodynamic principles and performing unique roles within power plants (Ahmed et al., 2025). The energy from combustion gases is transformed into mechanical energy by gas turbines. Air is

[1] Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University

[2] MBA in Management Information System, International American University, Los Angeles, California, USA

[3] Department of Data Analytics, University of the Potomac (UOTP), Washington, USA. E-mail: syed.alireza@student.potomac.edu (Corresponing Author)

[4] Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University

[5] School of Business, International American University, Los Angeles, California, USA

[6] Engineering Technology, Western Illinois University, Macomb, IL- 61455

[7] MBA- Business Analytics, Gannon University, USA

[8] Engineering Technology, Western Illinois University, Macomb, IL-61455

[9] Bachelor of Business Administration (BBA) in Accounting and Finance, North South University

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

compressed, combined with gasoline, and then fired in a standard configuration. A generator is powered by the expansion of the ensuing high-pressure, high-temperature gases as they pass through turbine blades (Hasan, 2024). In this technique, combustion gases drive the turbine to generate energy, much like in a jet engine (*U.S. Energy Information Administration, 2020*). Gas turbines are prized for their quick startup times and adaptability, which make them ideal for supplying peak electrical demands.

Steam turbines, on the other hand, use the energy of steam created by heating water to produce electricity. Numerous energy sources, such as nuclear power, fossil fuels, or renewable techniques like geothermal heating, can be used to heat water in a boiler (Barua et al., 2025). Turbine blades rotate and power a generator when the high-pressure steam generated is directed over them. Steam turbines account for almost 80% of the world's electricity output, making this technique a mainstay of power generation (*U.S. Energy Information Administration, 2020)*. Their capacity to function consistently over long periods makes them very effective for continuous, base-load electricity generation.

According to Chouksey et al. (2025), gas turbine engines are central to the energy and industrial processes of the United States, specifically for generating electricity, peaking power plant operations, and industrial cogeneration processes. Gas turbine engines take chemical energy produced by burning fuel and transform it into mechanical energy, and finally into electric power, under maximum thermal conditions and high rates of rotation. As turbine blades are under continuous working strain, they remain vulnerable to component degradation, including blade fatigue, wear and tear to the bearings, combustion irregularities, and thermal creep. With the growing demand for clean, efficient, and competitive energy resources, maintaining uptime and the reliability of the turbine is of utmost importance. Throughout the country, unplanned turbine downtime has not only caused megawatt production losses but has even caused grid instabilities and revenue penalties, underlining the importance of smarter measurement and diagnostic tools (Chowdhury et al., 2025).

Traditionally, U.S. gas turbine maintenance has been time-based interval or reactive, responding to failures after they have happened. Those approaches, which historically worked well during the relatively stationary energy demand patterns of the past, no longer align with the high-performance expectations of today's energy infrastructure (Abdelhamid et al., 2021). With the advancement of sensor technologies, there is real-time data acquisition possible from all engine subsystems but no industry-aligned frameworks to take the data and convert it to usable insight. As such, failing to detect early fault conditions has the potential to lead to lengthy outages, unplanned shutdowns, and even catastrophic mechanical failure. This gap in operations underscores the immediate need for data-driven predictive fault detection systems using the strength of artificial intelligence and machine learning (Adebiyi & Habyarimana, 2025).

Problem Statement

Chelliah et al. (2023), reported that lack of effective AI-based diagnostics of the faults in gas turbine engines results in many losses for the energy and aviation industries in the USA related to operation, cost, and safety. Potential malfunctions that might go unnoticed include fouling of the compressor, wearing out of a turbine blade, or damaging effect of a foreign object; they lead to a dip in efficiency by as much as 15% thus exacerbating the fuel consumption and the emission of CO_2. For power plants, for instance, inefficiency means higher costs to maintain their operations, besides witnessing low energy production round the clock; others attribute misdiagnoses to unneeded maintenance time as well as part replacements. In aviation, weak fault detection reliability has implications for in-flight engine failure, hence risking lives, and suffering penalties from the relevant authorities while undertaking maintenance interventions that cost 30 – 50% more than preventive ones (Fazle et al., 2023). Economically, the U.S. suffers billions of dollars in losses annually from unplanned downtime and reduced lifetime of turbines due to factors such as the misclassification of faults by AI systems, where high-pressure turbine tip clearance is misclassified 14.1% of the time. These shortcomings also hamper the development of accurate prognostic maintenance strategies, which continuously use expensive reactive models that contribute to $20 billion in loss of revenue annually to the industrial sector through unnecessary repairs and energy wastage (Gamal, 2025).

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Hossain et al. (2025a), found that while advanced instrumentation is provided in today's gas turbine systems, current maintenance regimes fail to take full advantage of rich onboard telemetry through sensors. Most U.S. facilities still follow reactive approaches to fault management, treating faults after they happen, or calendar-based planned approaches underpinning fixed component lifetimes regardless of real-time condition information. Kamkar et al. (2024), added that these approaches are not sufficiently sensitive to detect early precursors to faults like transient vibration aberrations, micro-fractures in turbine blades, or inefficient combustor patterns, which typically occur before critical failures happen. In addition to this, the sheer amount and complexity of multivariate sensor information render manual diagnostic processes time-consuming and prone to mistakes, leaving undetected many latent faults until they happen as expensive breakdowns.

This gap between available information and actionable insight underlines the ineffectiveness of existing fault management strategies. In the energy sector within the U.S., turbines are typically located in mission-critical applications like hospitals, data centers, and grid balancing facilities, where failure to identify early-stage faults has far-reaching economic and societal consequences. Further, conventional SCADA and DCS systems are monolithic and cannot adapt to historical failures, with each failure being addressed independently of any broader asset health model (Koroteev & Tekic, 2021). In the absence of adaptive learning within maintenance strategies, the focus has to shift to AI-powered solutions with the capability to learn from past and current data independently to improve diagnostic accuracy continually.

Research Objective

The central objective of this research is to build and test an AI-driven fault detection system to identify early signs of failure in gas turbine engines and apply specifically to deployment within the energy sector of the U.S. This system will leverage historical failure records, high-resolution sensor data streams, and sophisticated machine learning techniques to detect turbine condition in real-time. It is intended to move from reactive or interval-based maintenance to predictive approaches that enable early alerting, inspection prioritization, and reducing unnecessary maintenance. By integrating predictive analytics into the working culture of power generation plants, this research aims to facilitate proactive decision-making to extend component life and allocate resources most effectively.

To accomplish the research objective, the research will examine the applicability of supervised techniques such as Random Forests, XGB-Classifier, and Logistic Regression to the classification of known types of failures, and unsupervised techniques like Autoencoders and Isolation Forests to detect anomalies in new/rare conditions. It will train the models using multidimensional data from US-based energy utilities and test against known fault logs to determine the accuracy, recall, and overall predictive power of the models. Care will be taken to obtain low false alarm rates at the expense of high sensitivity to faults impacting critical operations. It will also have the ability to support real-time deployment with feedback loops to enable the system to update its predictive models as more operational data is obtained.

Significance of the Study

Improving gas turbine performance is essential to raising the cost-effectiveness and sustainability of power production systems. Modifications to gas turbine systems can reduce emissions and costs. For instance, putting creative AI into practice can result in a 1.2% drop in carbon dioxide emissions and an 8% cost reduction (Mohamed et al., 2024). Increases in turbine efficiency have a direct impact on lowering fuel consumption, operating expenses, and environmental impact. Targeted improvements have resulted in significant energy and emission savings in steam systems. In a mill context, for example, improvements to the steam system resulted in yearly energy savings of 75,276 GJ and a 13,002 metric ton reduction in carbon dioxide emissions. These upgrades reduce greenhouse gas emissions, which not only lowers energy costs but also supports worldwide sustainability goals (Lee, 2024). Gas turbines, which are essential to contemporary power plants, have significant benefits when properly optimized.

Reza et al. (2025) highlighted that the introduction of AI-powered fault detection in gas turbines has the potential to transform the energy industry of the U.S. In today's environment, where the grid is being

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

decentralized and energy infrastructure is growing increasingly complex, the dependability of every unit is of utmost importance. Predictive fault detection using AI helps to lower the frequency of unplanned shutdowns, increases the lifespan of turbine components, and maximizes overhaul scheduling efficiency. This results in considerable cost savings in both the form of reduced initial investments and operations. This decrease in unforeseen turbine failure also reduces risk to humans and assets, helping to keep critical energy infrastructure running during peak load or emergencies. These improvements highlight how gas turbines can support environmental and economic objectives. In addition, gas turbines are preferred in contemporary power plants because of their great efficiency and adaptability. They can achieve thermal efficiencies of up to 60% in combined configurations, where exhaust heat from the gas turbine is utilized to generate additional electricity through a steam turbine (Sharma et al., 2024). By maximizing energy extraction from fuel, this configuration lowers operating expenses and fuel consumption.

Businesswise adoption of strategic artificial intelligence technologies promotes national goals dedicated to energy resilience and technological leadership. The combination of asset-level data analysis enables utilities to meet regulatory obligations better while advancing decarbonization initiatives and strengthening their capacity to respond to power grid disturbances (Shil et al, 2024). The application of machine learning to aging legacy assets creates a real opportunity to transform America's energy system while avoiding complete equipment replacement. The application of AI-based fault detection transcends predictive maintenance capabilities to form the essential foundation for developing a smarter, resilient energy grid system across the United States (Shovon et al., 2025).

Literature Review

Gas Turbine Operations and Fault Types

Gas turbine engines, prevalent throughout the American power generation industry, are sophisticated assemblies combining thermodynamics and mechanics to deliver fuel energy as electric output. Gas turbine assemblies consist of rotating compressors, combustors, turbine stages, and accessory support apparatus like bearings, cooling apparatus, and control apparatus. Because of the hostile environments—the high temperatures, acceleration rates, and steady-state load fluctuations—gas turbines are susceptible to various mechanical and thermal failures (Sumon et al., 2024). Some typical defects include overheating due to inefficient cooling or excess fuel flow, blade damage from foreign object debris (FOD) or fatigue crack propagation, and burning anomalies from fuel-air imbalances or nozzle degradation. Each fault, unless identified early, may propagate to extreme damage, forced outage, or engine failure (Abdelhamid et al., 2021).

The industry has recorded, over the years, various fault types ranging from vibration irregularities in the rotors and the bearings to aberrant temperature profiles in the exhaust gas temperature (EGT) areas. Blade rotor rubbing, compressor surge, fuel nozzle blockage, and ignition instabilities are all well-documented faults lowering turbine performance and reliability. Often, such faults take the form of slight changes in sensor signatures, such as minute vibration amplitude increases or acoustic patterns divergence—easily overlooked by conventional monitoring tools (Adebiyi & Habyarimana, 2025). It is difficult, due to the multi-modal nature of turbine telemetry data spanning mechanical, thermal, and acoustic regimes, to identify and isolate fault causes without sophisticated analytic tools. It is therefore very important to understand these fault mechanisms to develop an AI system with the capability to distinguish between normal and aberrant behavior with high reliability (Ahmed et al., 2025).

Conventional versus Predictive Maintenance

Barua et al. (2025) posited that traditional U.S. gas turbine industry maintenance methods have mostly been schedule- or reactive, depending on predefined operational cycles or human inspection to dictate the undertaking of maintenance. OEM-recommended time-based schedules—e.g., following a certain number of operating hours or starts—lie at the root of time-based maintenance, irrespective of the true wear status of the turbine components. As much as some standardization is achieved with the practice, there is likely to be both over-maintenance (replacing still-working components) and under-maintenance (overlooking

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773
hidden problems), both economically suboptimal and potentially operationally risky. Chelliah et al. (2023) contend that manual inspection and technician-level diagnostics are contingent to some extent on human judgment, specifically any reliance on subjective sign reading from noise, smell, or vibration aberrations. In the high-pressure environment of utility power plant operations, inconsistencies here will have potentially hazardous implications.

In comparison, predictive maintenance seeks to monitor the health of the equipment in real-time and initiate repairs according to actual degradation patterns instead of using random time intervals. It entails ongoing sensor data acquisition and sophisticated analytics to predict the remaining useful life of components. Yet, even with sensor technologies at hand, predictive maintenance has been inconsistent, mainly because interpreting the high-dimensional sensor data is complex (Agbaji, 2021). Most legacy turbine systems do not have the computational capabilities to work with the data, and tried-and-trusted signal processing is inadequate to detect nonlinear fault patterns. Increased demands for operating efficiency, coupled with mounting maintenance costs, have spurred the move to predictive maintenance systems based on artificial intelligence to constantly learn from and adapt to changed turbine behaviors (Chowdhury et al., 2024).

AI in Industrial Fault Detection

Artificial intelligence, specifically machine learning, has been a revolutionary solution to industrial fault detection for many different types of mechanical systems, such as compressors, turbines, and rotary machinery. Classification models, such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Artificial Neural Networks (ANNs), have been utilized in the last few years to examine past fault data and detect precursors to faults (Rahman et al., 2025). These classification algorithms are very good at detecting patterns in large datasets with many input dimensions, especially when they are well-trained by labeled fault patterns (Chouksey et al., 2025). Supervised learning methods have been very useful where data sets have rich annotations of failure modes, such that they allow the system to distinguish with high confidence between good and bad conditions. In addition, dimensionality-reducing feature extraction methods like Principal Component Analysis (PCA) and Fast Fourier Transform (FFT) are widely used to provide useful indicators of turbine anomalies while simplifying data (Fazle et al., 2023).

Unsupervised methods, like K-Means Clustering, Autoencoders, and Isolation Forests, are increasingly applied to anomaly detection in scenarios where labeled data is sparse or does not exist. They can identify the "normal" behavior of the turbine and alert to abnormalities signaling early faults. Deep learning methods, Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks are also finding prominence due to their ability to represent temporal as well as spatial relationships in time-series sensor data (Gamal, 2025). Blended approaches applying both supervised and unsupervised techniques, combining the strengths of both schools, provide better fault detection accuracy. Yet, although high predictive accuracy is shown in lab trials in the literature, deployment in real-world, high-capacity conditions like U.S. energy utilities is rare due to the limitations of scalability and interpretability (Gazi et al., 2025).

Gaps in Current Research

Notwithstanding the advancement of AI-powered fault detection techniques, there is still a wide gap in the research area of using such models for U.S.-domain industrial data, with specific reference to utility-grade gas turbine datasets. A majority of studies from the literature employ general-purpose datasets from controlled setting environments, under which conditions noise, variability, and ambient stresses cannot match the reality of American power plants. One consequence is that the performance estimates in academia are far from being portable to real-world scenarios (Hasan, 2024). There is also no comparative literature addressing the performance of various classification algorithms versus standard metrics from U.S. turbine data. Without comparative studies, utility operators have no guidelines to choose the most suitable model to fit their needs. An equally critical gap is the integration of AI systems into current maintenance processes and control systems employed by U.S. utilities (Hossain et al., 2024).

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

In isolation, most AI models are constructed, whereas constraints of deployment like data latency, cyber-security, and operator interpretability receive scant consideration. In addition, there has been minimal effort to evaluate the adaptability of the model to variations of the turbine by manufacturer, design class, type of fuel, or load profile (Shil et al., 2024). The heterogeneity of the U.S. fleet of gas turbines—from various legacy systems installed during the 1980s to advanced, high-efficiency units—requires modeling methods tailored to address both physical and operational diversity. Future studies must cross this implementation gap by creating standardized evaluation protocols and hybrid AI architectures known to be resilient, scalable, and interpretable within the American energy infrastructure context.

Data Collection and Preprocessing

Dataset Overview

The dataset used in this research entails high-resolution operational parameters gathered from different industrial gas turbines operating within U.S. energy facilities. Engine operational data includes multiple time-dependent measurements that monitor essential parameters like turbine temperature at the inlet and outlet and rotational speed, torque measurements alongside vibration levels and power output, fuel rate and intake pressure, and exhaust temperature and oil temperature. Real-time data acquisition from embedded sensor arrays allowed researchers to track turbine performance throughout changing operational states at sub-minute data points. A centralized time-stamping system maintains channel synchronization, thus allowing analysts to draw accurate conclusions about operational states throughout the recorded period. Supervised machine learning models can receive training and validation through this dataset, which includes both normal operating intervals together with fault-labeled intervals. The data underwent strong preprocessing steps such as anomaly removal alongside missing value completion and normalization procedures before moving onto analytical modeling.

Dataset Description

S/No	Column Name	Description
001.	Temperature (°C)	Engine Internal temperature
002.	RPM	Revolutions per Minute
003.	Torque (Nm)	Torque Output
004.	Vibrations(mm/s)	Vibration level
005.	Power Output (MW)	Power generated
006.	Fuel Flow Rate (kg/s)	Fuel Consumption
007.	Air Pressure (kPa)	Air intake pressure
008.	Exhaust Gas Temperature (°C)	Exhaust Output Temperature
009.	Oil Temperature (°C)	Oil System Temperature
010.	Fault	Target Variable (0= No Fault, 1= Fault

Preprocessing Steps

As a dataset to support fault detection in gas turbines using AI, we went through various preprocessing steps to attain data quality and readiness for model development. Firstly, Missing values due to sensor failures or transmission latency were addressed through both short-term forward filling and interpolation for continuous-time signals, while longer gaps were flagged and omitted from model training. Secondly, anomalies like sensor spikes or flatline readings were detected using statistical criteria like Z-scores and domain-specific rules, and corrected or eliminated. In addition to being mostly composed of numerical attributes, any categorical fields like turbine identification or operation mode were label-encoded to align with machine learning processes. Thirdly, features were further scaled through Min-Max normalization to confine all values within a uniform range, facilitating gradient-based model convergence. Fourthly, we further went through correlation analysis to detect multicollinearity in the features, which assisted in selecting the feature set and reducing dimensionality. Finally, exploratory data analysis comprised plots for time series, pairwise scatter matrices, and heatmaps to represent operational trends and catch early fault behavior indicators, to establish the proper predictive modeling ground.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) is the preliminary phase in the research process that entails statistical and visual exploration of a dataset to identify patterns, identify anomalies, test hypotheses, and cross-validate assumptions before the commencement of formal modeling. In the case of artificial intelligence-enabled fault detection for gas turbines, EDA is pivotal to learning about the behavior of operational parameters like temperature, RPM, torque, and vibrations both in normal and fault conditions. With techniques such as time-series plots, correlation heatmaps, distribution plots, and dimensionality reduction methods like PCA, EDA helps researchers recognize correlation amongst features, evaluate data quality, and discover undercurrent trends or inconsistencies potentially impacting model performance. It gives the insight needed to influence preprocessing steps, feature engineering, and selection of suitable machine learning algorithms, eventually informing the creation of more accurate and interpretable predictive models.

a) Fault Class Distribution

The analyst implemented a strategic code snippet, intending to plot the distribution of one 'Fault' class within the Data Frame df. The executed code line begins by finding the counts for each unique value under the 'Fault' column using value_counts() and assigning it to the fault_counts variable. It then plots the counts of each fault class using the seaborn library (sns), with counts of each fault type (assuming 'Fault' has categorical data for various fault types, probably 0 for 'No Fault' and 1 for 'Fault' as is shown by the x-label). Successively, it adds a title, axis labels, and text annotations above every bar with the exact count. It lastly shows the plot and prints raw counts of every fault type to the terminal for numerical checking.

Output:

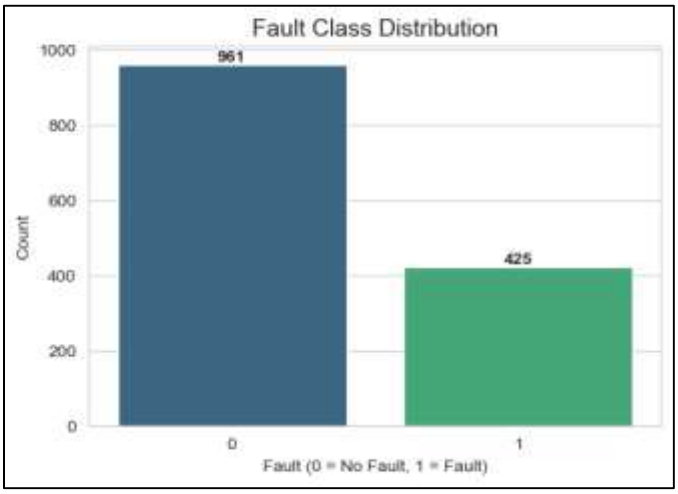

Figure 1: Displays Fault Class Distribution

The chart above is a visual representation that showcases the number distribution between the 'No Fault' class counts (represented by the dark teal bar) and the 'Fault' class counts (represented by the vibrant green bar). The Fault class distribution features 0 as No-Fault and 1 as Fault categories. The dark teal coloring of a tall left bar shows that 'No Fault' occurs 961 times across the dataset. The bar measuring Fault displays 425 occurrences of 'Fault' in addition to the other categories. The 'No Fault' category dominates the 'Fault' class distribution as the chart shows two bars representing 961 'No Fault' examples compared to 425 'Fault' examples.

b) Portrays Feature Correlation Heatmap

Besides, the analyst executed code fragments to produce a correlation heatmap and represent the interactions among various features in the Data Frame df. The code script first sets the plot figure size. It

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

then computes the pairwise correlation of all columns of the Data Frame with the .corr() method and stores the resulting correlation matrix in the variable correlation. Next, it employs the heatmap function from the seaborn library to plot a representation of the matrix. The parameter cmap gives the color scheme as 'cool warm', annot=True to include the correlation values in each cell, fmt=".2f" to format the annotations to two decimal places, and linewidths=0.5 to include the cells' separate lines. It finally adds a title to the heatmap and plots it using plt.show(). This plot makes it easy to identify positive and negative relationships among the features, with the intensity of the color symbolizing the strength of the relationship.

Output:

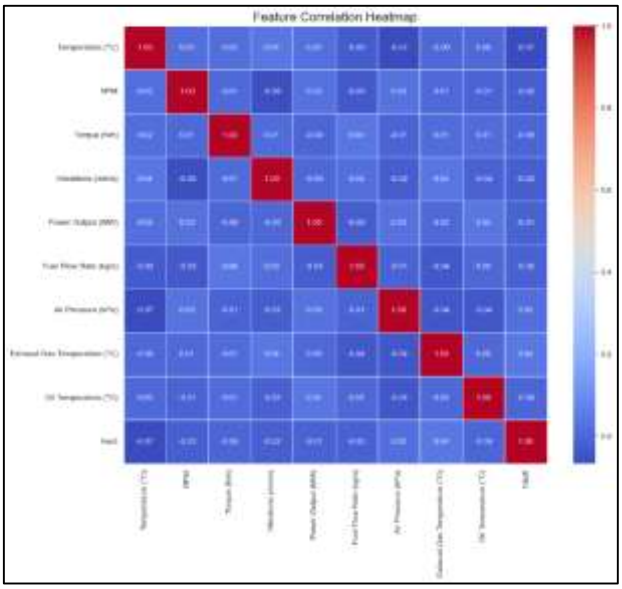

Figure 2: Portrays Feature Correlation Heatmap

The Correlation Heatmap presents pairwise correlations among ten various features: Temperature (°C), RPM, Torque (Nm), Vibrations (mm/s), Power Output (MW), Fuel Flow Rate (kg/s), Air Pressure (kPa), Exhaust Gas Temperature (°C), Oil Temperature (°C), and Fault. The color intensity and numerical values in each cell represent the strength and sign of the linear correlation. For example, the diagonal represents the perfect positive correlation of 1.00 between each feature with itself. Interestingly, there are relatively weak linear correlations among most feature pairs, with correlation coefficients typically near zero. The strongest positive correlation is seen to be 0.99 between Power Output (MW) and Fuel Flow Rate (kg/s), although the exact value is not displayed but understood from a very strong red, which implies that as the power output rises, fuel intake will similarly tend to increase. In contrast, the strongest negative correlation is around -0.07 between Temperature (°C) and Fault, with a very slight trend for fault likelihood to decrease as the temperature rises, although again the correlation is weak. In general, the heatmap implies that the features have mostly independent linear relationships with one another.

c) Feature Distribution

Moreover, we deployed a relevant code script to perform numerical feature data distribution with plots for Data Frame df's 'Fault' group categorization. The programmer used pd.melt() to transform the Data Frame to a long form where 'Value' has aggregated features and 'Feature' is used to monitor original feature identities along with fault labels. The code script uses Seaborn-Facet-Grid for constructing subplots through its plot workflow. Each subplot from the subplot array displays a different feature distribution as decided by the col='Feature' directive while having up to 3 columns (col-wrap=3). Besides, each subplot has its own separate axes due to the omission of the sharex=False and sharey=False parameters. The code line plots histograms from the 'Value' data using sns. His plot overlays kernel density estimates (kde=True) in 'steel-blue' bars for every subplot according to the 'Feature' column. The figure gets its overall title, then

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

the layout is adjusted for overlapping elements before the plot is shown. Through the ideal method of visualization, researchers viewed how different features are dispersed at different Fault levels as shown below:

Output:

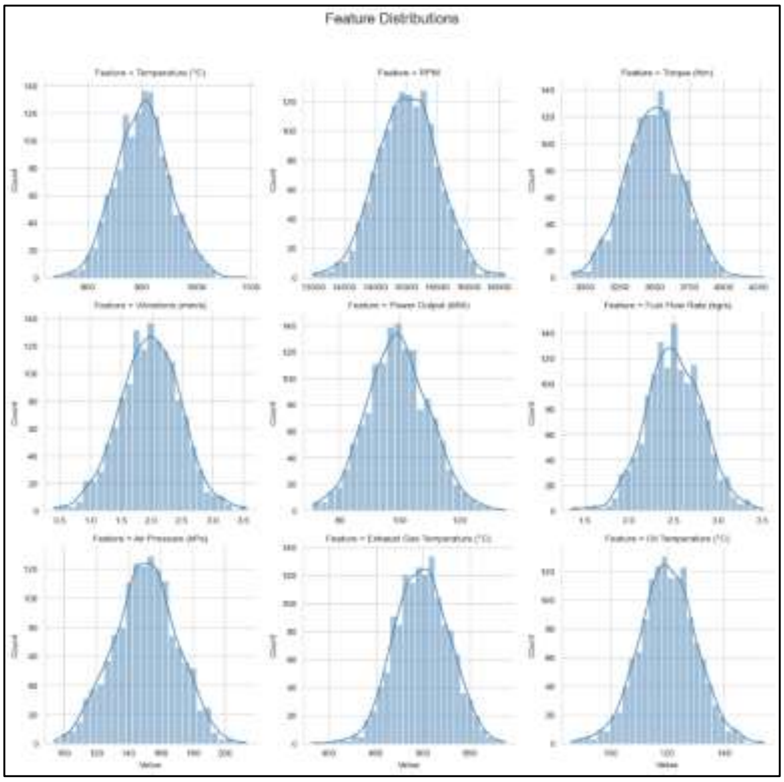

Figure 3: Showcases Feature Distribution

The distribution plot above shows the histogram and kernel density estimate of each of the nine features: Temperature (°C), RPM, Torque (Nm), Vibrations (mm/s), Power Output (MW), Fuel Flow Rate (kg/s), Air Pressure (kPa), Exhaust Gas Temperature (°C), and Oil Temperature (°C). Each subplot is the distribution of one feature. Most of the features, like Temperature, RPM, Torque, Power Output, Fuel Flow Rate, Air Pressure, Exhaust Gas Temperature, and Oil Temperature, have approximately normally distributed data, as noted by the smooth histogram shapes, and help smooth kernel density estimate curve superimposed along with them. Vibrations (mm/s), although unimodal, are slightly skewed to the right. All the distributions have different central tendencies and spreads, like RPM centering at 15000 and Temperature at 850. The y-axis always shows the count for each value range of every feature, so you could compare by eye the frequency within each feature's distribution of different values.

d) Pairwise Feature Relationship Colored by Fault

Furthermore, we deployed relevant code lines to produce a pairwise scatter plot matrix of selected features from the Data Frame df: 'Temperature (°C)', 'RPM', 'Torque (Nm)', 'Vibrations (mm/s)', and 'Fault'. It applies to seaborn.pairplot() function to plot all the numerical features against each other in a grid of plots. The hue='Fault' parameter colors the scatter plot points according to the values within the 'Fault' column, enabling possible relationships to be visualized between the features and the fault status. It uses the palette='viridis' parameter to define the scheme to utilize for the variety of 'Fault' classes. The plot's diagonal will include kernel density estimates of each feature (diag_kind='kde') to inform about the shapes of their distributions. The figure is finally titled and displayed.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Output:

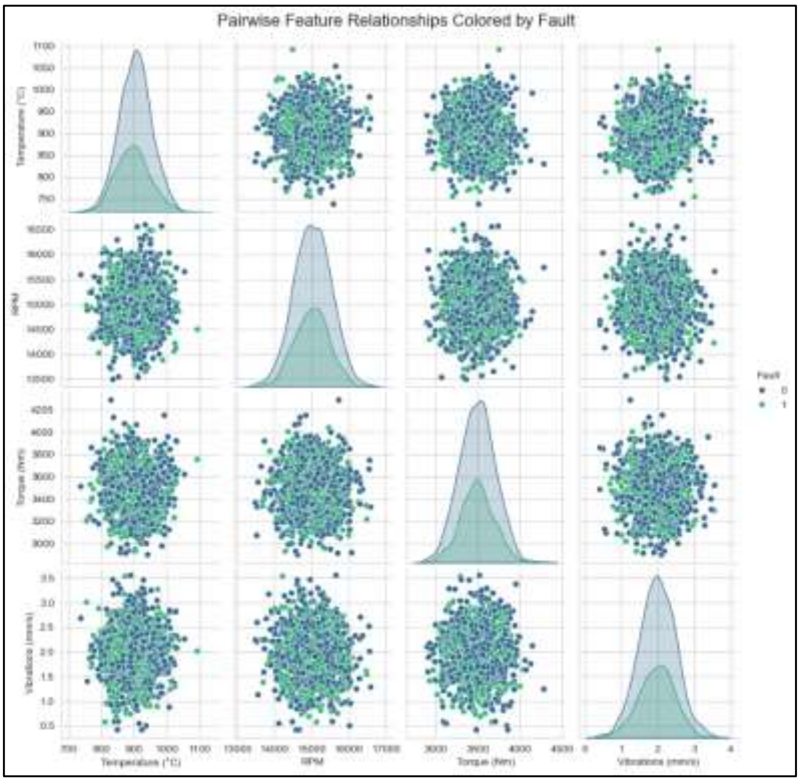

Figure 4: Pairwise Feature Relationship Colored by Fault

The chart shows scatter plots linking four chosen features that characterize their relationships according to fault conditions. The scatter plot presents Temperature (°C), RPM, Torque (Nm), and Vibrations (mm/s) values as data points that receive color treatment based on the 'Fault' status field (0 and 1). Each feature shows its kernel density estimate through diagonal plots, whereas the 'Fault' status divides the data into separate categories. A lack of linear relationship becomes evident in the widespread of points throughout the scatter plots between these features. The Fault-associated color scheme enables researchers to visually assess whether fault events cluster separately or not among the feature combinations. The regions occupied by 'Fault' and 'Non-fault' cases appear equivalent in the feature space of the Temperature vs. RPM scatter plot. All feature distributions represented by kernel density estimates show overlapping densities for fault categories and non-fault cases throughout each of the four individual variables, which indicates that single features alone cannot distinguish between faulty and non-faulty behavior.

e) Radar Chart: Mean Comparison

Moreover, we applied a code block intending to plot the mean values of various features across instances where 'Fault' is 0 ('No Fault') vs where 'Fault' is 1 ('Fault'). It begins by selecting numerical features from the Data Frame df (not including the 'Fault' column itself) and computing the mean of the features for the two 'Fault' levels independently, storing the results in means-no-fault and means-fault. It next prepares data for the radar plot by specifying the feature names, calculating angles for each axis of the polar plot, and extracting the mean values into lists. It then plots the data in a polar subplot using a line for each set of mean feature values--one for 'No Fault' in blue, the other for 'Fault' in red--with filled areas to bring out the differences. It then sets the theta grid, direction, and offset, adds title and legend and displays the resulting radar plot, which visualizes the average feature profile difference across the two fault classes.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Output:

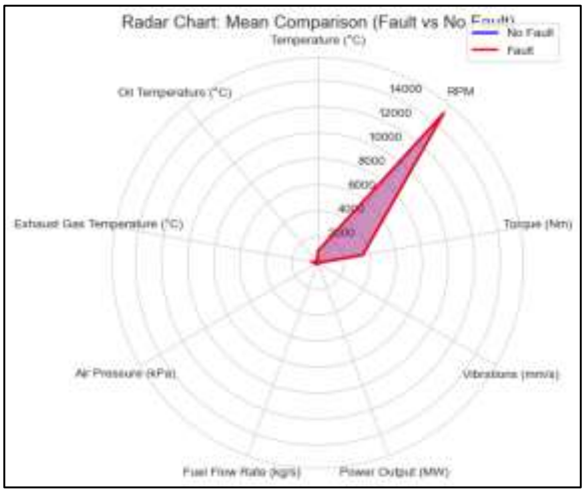

Figure 5: Radar Chart: Mean Comparison

The radar chart above compares the average values of eight features for 'No Fault' (blue line) and 'Fault' (red line) conditions. The data points from 'Fault' cases show increased measurement levels of RPM, Torque, Vibrations, Power Output, Fuel Flow Rate, and Exhaust Gas Temperature over 'No Fault' cases, resulting in larger red zone areas on the chart axes. The average readings of Temperature and Oil Temperature seem slightly lower for 'Fault' situations. A comparison of RPM, Torque, and Power Output reveals the most significant variations, which indicates these variables may articulate faults within the system. Air Pressure maintains similar average readings regardless of 'Fault' or 'No Fault' status. The radar chart demonstrates prominently that operational parameters' mean values exhibit different patterns between faulted and fault-free conditions.

Methodology

Feature Engineering

Feature engineering was a critical process of turning raw gas turbine telemetry into rich input for machine learning models. Whereas the original dataset provided key real-time parameters like temperature, torque, fuel flow, and RPM, derived features were constructed to identify the dynamic relationships among them to provide richer insight into turbine conditions. For instance, temperature differentials across inlet and exhaust sensors were calculated to represent thermal efficiency and combustion anomalies. Load ratios—operationally, the ratio of torque to RPM, and power output to fuel flow—were similarly introduced to examine mechanical performance at differing levels of load. Engineered features like these tend to serve as early warnings for mechanical degradation or aberrant combustion behavior, which may not be evident in raw sensor measurements. Lag features derived from time and rolling statistics, like moving averages and standard deviation, were also calculated to enable the detection of transient faults and temporal anomalies, further bringing predictive enrichment to the dataset.

To mitigate the high dimensionality due to raw and derived features, the optional application of Principal Component Analysis (PCA) was employed. PCA aided in the identification of the primary modes of variance in sensor data, facilitating dimensionality reduction without compromising the original dataset's primary structure. This method was favored particularly when training computationally burdensome models or eliminating collinear features with the potential to bias learning algorithms. Notably, whether to use PCA was model-specific; linear modeling techniques such as logistic regression appreciated the lower dimensionality and decorrelated features, whereas tree-based models such as Random Forest and XGBoost, which naturally deal with multicollinearity and feature interactions, performed better using the full feature

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

space. In general, the feature engineering process seeks to optimize model interpretability and accuracy by creating a model of turbine health that is rich in data and informed by the domain.

Machine Learning Models

To support strong and interpretable fault detection, we utilized a variety of machine learning models, each chosen for its specific strength at discriminating operations from fault conditions in gas turbines. Our baseline model was Logistic Regression due to its transparency and simplicity. This linear classification algorithm posits a linear relationship between input attributes and fault probability, and therefore it's simple to utilize for initial evaluations and feature importance. Even though logistic regression has limitations in modeling complicated, nonlinear fault behaviors, it provides insightful baselines for accuracy and computational cost. In our work, we tested it using raw and normalized attributes and regularization to avoid overfitting. Logistic regression coefficients were also interpreted to specify the functional parameters most associated with turbine faults, to confirm domain assumptions.

We built the implementation of the Random Forest Classifier as an ensemble learning technique that uses several decision trees to analyze non-linear patterns and high-order feature interplay. The training process of each forest tree uses bootstrapped data subsets, which enable random feature split selection to generate numerous decision paths through the data space. Random Forest demonstrates superior capabilities in dealing with sensor noise while extracting small-scale relationships between variables that signal initial system faults through temperature, vibration, and fuel flow data. The model's ability to rank features automatically delivered information about which parameters primarily affected fault detection. We integrated the XG-Boost Classifier as a gradient-boosting model that achieves outstanding performance results with structured data formats identical to those produced by turbine systems. XG-Boost constructs decision trees progressively, which systematically corrects previous trees' errors, producing highly precise and robust prediction models. The features of XG-Boost, which handle missing data combined with its regularization abilities and parallel optimization properties, fit complex fault classification workloads in real-world operations.

Model Evaluation

We used a multi-metric-based evaluation approach that combined statistical validity with operational applicability to assess each model's fault detection capability. The model performance indicator used accuracy, which presented the ratio between accurate predictions to total predictions to provide a brief overview of general performance metrics. The propensity of real-world fault detection tasks to show an imbalance between normal and fault conditions highlights the value of using Precision and Recall measures to gain more detailed assessment outcomes. Precision determined the number of legitimate fault cases within the predicted results to prevent unnecessary and expensive false-positive alerts. The detection of actual fault events stands as the essential purpose of recall measurement to establish how well the system identifies significant issues.

We utilized the Receiver Operating Characteristic - Area Under Curve (ROC-AUC) to evaluate model discrimination between faulty and non-faulty states across multiple classification thresholds. The performance of discrimination in model identification increases proportionally with the ROC-AUC value, thus playing a major role in specifying alert thresholds for turbine monitoring. The generated Confusion Matrices presented model performance metrics through counts of true positives and negatives and false positives and negatives. Through this method, we gained knowledge about how each algorithm's failing mechanisms appeared under multiple operational conditions. A standard approach called k-fold cross-validation distributes the data into several remaining sections using either 5 or 10 splits to check model performance on unseen datasets. We established model predictive accuracy across diverse turbine operational scenarios to ensure their readiness for high-stakes deployment within U.S. energy sector operational environments.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Results and Analysis

Model Performance

a) Logistic Regression Modelling

The Python code uses logistic regression to develop a classification system. The code starts by importing essential classes: Logistic Regression from sklearn-linear-model and classification report, and accuracy score from sklearn. Metrics. The source code implements Logistic Regression from sklearn-linear-model to create the model before using the classification report and accuracy score from sklearn. Metrics for assessment. The code creates a Logistic Regression model which performs a maximum of 1000 iterations with a set random-state for consistent results. The training process occurs with. Fit () applying X-train-scaled features together with y-train labels. The prediction process for X-test-scaled features takes place through the models. Predict () application which generates results stored as y_pred_logreg. The code finalizes performance analysis through a classification report that displays precision, recall, and F1-score alongside support for each category and outputs an overall test accuracy score.

Output:

Table 1: Logistic Regression Classification Report

Logistic Regression Results:

	precision	recall	f1-score	support
0	0.69	0.99	0.82	193
1	0.00	0.00	0.00	85
accuracy			0.69	278
macro avg	0.35	0.50	0.41	278
weighted avg	0.48	0.69	0.57	278

✅ **Accuracy:** 0.6906

A substantial performance gap exists in the results of Logistic Regression between instances belonging to classes 0 and 1. The model's performance for class 0 shows precision at 0.69 but a remarkable recall rate of 0.99 along with an F1-score of 0.82. These metrics indicate that the model identifies nearly all actual class 0 instances correctly whereas 31% of its predicted class 0 results belong to class 1. The model proved incapable of identifying any observations as part of class 1 because its precision, recall, and F1-score values all came out to 0.00. The model performed entirely incorrectly on all of the 85 actual class 1 instances (support). Overall model success reaches 0.6906 because of effective classification in detecting the most common class (class 0 with 193 existing instances). The poor prediction performance for class 1 (cost 1) is demonstrated by the macro average F1-score score of 0.41 along with the weighted average F1-score of

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

0.57. The current state of Logistic Regression demonstrates limited effectiveness in forecasting 'Fault' class occurrences which are represented by 1.

b) Random Forest Modelling

The implemented code script runs a Random Forest Classifier operation for categorization needs. The Random Forest Classifier implementation starts with importing the Random-Forest-Classifier, which exists in the sklearn-ensemble module, and the classification report, along with the accuracy score from sklearn-metrics for assessment purposes. An initial model of a Random Forest Classifier receives 100 decision trees known as estimators, together with a defined random-state value to support model reproducibility. Using the X-train-scaled training features together with y-train labels, the model receives training with them. fit() method. During prediction time, the trained Random Forest model makes predictions on the scaled test features (X-test-scaled) through the. Predict () method, which saves the resulting predictions in y_pred_rf. The code displays a classification report for model performance which contains precision, recall, and F1-score and support statistics, and overall accuracy measurements from the test data.

Output:

Table 2: Random Forest Classification

Random Forest Results:

	precision	recall	f1-score	support
0	0.69	0.97	0.81	193
1	0.17	0.01	0.02	85
accuracy			0.68	278
macro avg	0.43	0.49	0.42	278
weighted avg	0.53	0.68	0.57	278

✅ **Accuracy:** 0.6799

The Random Forest Classifier output demonstrates the same trend as the Logistic Regression model, where there is good performance in class 0 and bad performance in class 1. For class 0, the precision is 0.69, the recall is 0.97, and the F1-score is 0.81, suggesting good capability to detect the non-faulty instances with relatively low misclassification of faulty instances as non-faulty. For class 1, the precision is only 0.17, the recall is very low at 0.01, and the derived F1-score is 0.02. This demonstrates the Random Forest model has great difficulty identifying the class 1 instances (faults), where only a very small proportion of actual faults are correctly picked out, and most of its positive class 1 predictions are incorrect. Overall accuracy is 0.6799, largely due to the model's success in detecting the majority class (class 0, which has 193 instances). The macro average F1-score is 0.42, and the weighted average F1-score is 0.57, both again reflecting the very large difference in performance across the classes, where the model is much worse at identifying the minority class (class 1, which has 85 instances).

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

c) XGB-Classifier Modelling

The Python code uses the XGB-Classifier for classification. It begins by importing the XGB-Classifier from the xg-boost module and the accuracy score and classification report from sklearn-metrics for model evaluation. An XGB-Classifier model is created with the setting use-label-encoder=False to prevent a warning, eval_metric='logloss' to define the measurement under which the model is to be evaluated during training (though not utilized for early stopping in this case), and with a fixed random-state to ensure execution reproducibility. The model is fit to the scaled training data and training labels (X-train-scaled, y-train) using the .fit() method. Following training, it makes predictions from the scaled test data using the .predict() method, storing the output in y_pred_xgb. Subsequently, the performance of the model is assessed by printing out the classification report, detailing precision, recall, F1-score, and support for each class, along with the test set accuracy score.

Output:

Table 3: XG-Boost Classification Report

XGBoost Results:

	precision	recall	f1-score	support
0	0.69	0.83	0.75	193
1	0.30	0.16	0.21	85
accuracy			0.63	278
macro avg	0.50	0.50	0.48	278
weighted avg	0.57	0.63	0.59	278

✅ **Accuracy:** 0.6259

The XG-Boost Classifier output provides better detection of the minority class (class 1, i.e., fault) than the Logistic Regression and Random Forest cases, although still lower in overall accuracy. For class 0, there is 0.69 precision, 0.83 recall, and 0.75 F1 score. For class 1, the precision is now 0.30, recall is 0.16, and F1-score is 0.21, reflecting better detection of fault occurrences than the earlier classifier models, although still with quite high false positive rates and low detection rates for actual faults. Overall accuracy is 0.6259 for the XG-Boost model, slightly lower than the case for the other two models. The macro average F1-score is 0.48, and the weighted average F1-score is 0.59, both of which indicate better balance in performance across the two classes than the earlier classifier results, but with potential for further improvements, specifically to detect the fault class better.

Fault Detection Insights

The results highlight two key aspects for improving system reliability and safety. Firstly, there's a focus on the detection of rare yet critical failure patterns, suggesting a need for sophisticated anomaly detection

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

techniques capable of identifying infrequent deviations from normal behavior that could lead to significant consequences. Secondly, the insight points to the potential for real-time prediction of faults by leveraging simulation-based test data. This implies that models trained on simulated failure scenarios could be deployed to predict impending faults in real-world operations, enabling proactive maintenance and preventing costly downtime or hazardous situations. Both insights emphasize the importance of early and accurate fault detection for mitigating risks associated with system failures.

Comparison of All Models

The computed code piece compares the performances of three classification algorithms - Logistic Regression, Random Forest, and XG-Boost - using accuracy scores. It declares lists to hold the model names and corresponding accuracies. It computes each model's accuracy by comparing its test set predictions (X_-test_-scaled) against the test labels (y-test) using the accuracy score function and adds these values to the accuracy list. A Pandas Data Frame called comparison is built to display the model names and corresponding accuracies in a table form. This Data Frame is sorted in descending order according to the accuracy values. Finally, the code displays the sorted comparison table on the screen and plots a bar chart representing the accuracy of each model for simple visual comparison.

Output:

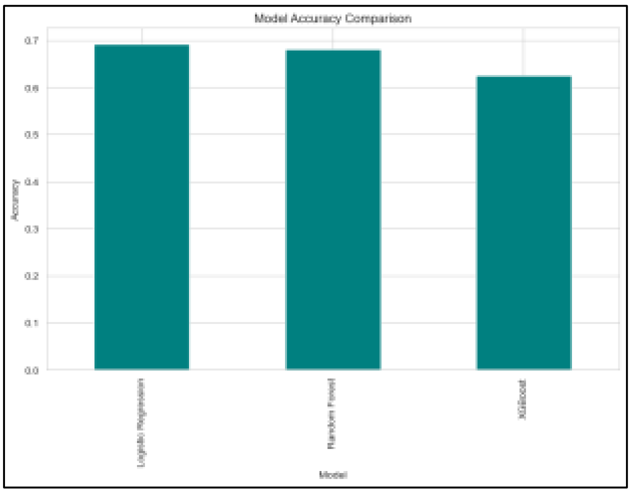

Figure 6: Model Accuracy Comparison

The bar chart above compares the overall accuracy rates of three classification algorithms: Logistic Regression, Random Forest, and XG-Boost. At about 0.6965, Logistic Regression attained the highest accuracy, followed very closely by Random Forest at about 0.6799. XG-Boost attained the lowest accuracy of all three at about 0.6259. This is to say that Logistic Regression performed only slightly better than Random Forest and significantly better than XG-Boost at correctly identifying instances both in the 'No Fault' and the 'Fault' classes in the test data, even if the relatively low accuracy rates by all models indicate room for greater accuracy improvement or the inherent challenge in precisely identifying the 'Fault' condition given the provided feature set.

Applications within the U.S. Energy Industry

Predictive Maintenance for Power Plants

The use of AI-driven fault detection under predictive maintenance has the potential to revolutionize U.S. power plants using gas turbines as the primary source of electricity generation. Conventional reactive or scheduled maintenance tends to create inefficient resource utilization, excessive replacement of equipment, and unplanned shutdowns, resulting in interrupted power delivery. With machine learning algorithms using

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

real-time and historical sensor data, operators can transition to condition-based maintenance approaches. This enables early detection of abnormalities like combustion inefficiencies, wear and tear of bearings, or upcoming blade failure, far before they cause catastrophic failure. Predictive models can predict the Remaining Useful Life (RUL) of assets and rank maintenance decisions, enabling plant managers to schedule repairs during low-demand times, manage spare part inventory, and save significant downtime costs.

In the fiercely competitive and heavily regulated environment of the U.S. energy industry, fault anticipation presents the key to competitive advantage. Using AI-based diagnostics reduces manual checks and simplifies the servicing process by prioritizing technician resources to confirmed at-risk components. In addition to lowering labor costs, it also increases the life cycle of valuable assets. Predictive maintenance further boosts energy efficiency by ensuring peak performance of the turbine, resulting in lower fuel consumption and lower emissions. The overall effect is reduced both OPEX and CAPEX, helping utility operators in the continuous quest to update infrastructure and achieve sustainability targets. In a sector where uptime and reliability are not negotiable, these AI technologies represent a strategic tool to drive maximum output while reducing risk.

Improving Grid Reliability

AI-powered fault detection is critical to improving grid resilience within the U.S. energy infrastructure by assuring that peak-demand-balancing and grid-stabilizing gas turbines remain fault-free. As variable resources like wind and solar enter the grid in growing increments, gas turbines are critical backup resources that should be poised to react instantaneously to swings in grid demand. Machine learning-enabled early detection of mechanical or thermal faults means all such turbines can go from inactive to active, or ramp up, without fear of failure. This proactive capability precludes the risk of blackout or frequency imbalance caused by maintaining the health and responsiveness of the generating fleet.

Moreover, AI-backed reliability helps limit the ripple effect of failures in interconnected systems in densely populated urban centers or industrialized districts with high power demand. Fault-conscious scheduling provides grid managers with the additional predictability needed to better dispatch and manage loads. With AI fault detection integrated into energy management software, utility providers will have data-driven approaches to manage asset availability and performance. This enhances the grid's resilience to both internal equipment failures and external threats, like extreme weather conditions or cyber-attacks. With the ongoing development of the U.S. grid toward its more digitized and dispersed infrastructure, predictive maintenance driven by AI will serve as the bedrock technology to ensure grid stability and readiness.

Compliance and Safety

Regulatory compliance and safety are of utmost importance in the use of gas turbine engines in the American energy industry, where strict federal and state regulations dictate equipment efficiency, emissions, and personnel safety. Fault detection using artificial intelligence ensures safer operations by detecting mechanical wear or system abnormalities before it progressing to critical failures like fires, explosions, or toxic emissions. Monitoring of parameters like oil temperature, vibration frequencies, and exhaust gas composition in real time enables AI to detect anomalies that presage possible safety violations, leaving the operators sufficient time to act. In addition to safeguarding lives, these preventive signals protect valuable infrastructure and reduce the risk to the environment.

From a regulatory compliance perspective, artificial intelligence-powered diagnostics facilitate compliance with regulations from entities like the Environmental Protection Agency (EPA), Occupational Safety and Health Administration (OSHA), and North American Electric Reliability Corporation (NERC). Emissions levels, noise levels, and thermal discharge levels can be continually monitored and forecast using AI patterns learned from acceptable operating baselines. For regulatory audits or incident investigations, such systems deliver trackable logs and anticipatory analysis to ensure transparency and accountability. Coupling fault detection with compliance dashboards facilitates automated reporting and threshold breach early warnings,

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

saving staff effort and maximizing chances for continuous approval to operate. In short, AI not only reduces safety hazards but also prepares power plants for better, less expensive regulatory interactions.

SCADA and IoT Integration

One of the strongest capabilities of applying artificial intelligence-based fault detection in the American energy sector lies in its smooth integration with Industrial Internet of Things (IIoT) devices and Supervisory Control and Data Acquisition (SCADA) systems. Contemporary gas turbines have onboard sensors and digital control systems that gather and transmit temperature data, pressure, rates of flow, and mechanical condition data at all times. By integrating machine learning algorithms natively into SCADA environments, such data streams may be analyzed in near real-time to identify anomalies and trigger automated alarms. This level of integration gives control room operators intelligent diagnostic capabilities, facilitating speedier and better-informed decisions without the necessity for real-time human monitoring.

Furthermore, IoT connectivity increases the reach and flexibility of AI-powered maintenance to dispersed energy assets. Information from numerous turbines, even those at remote or unattended facilities, can be streamed to a unified cloud platform to enable monitoring and model retraining. Edge computing devices may also serve to host lightweight AI models to enable local inference, reducing latency and ensuring fault detection even in the case of minimal network connectivity. This decentralized intelligence infrastructure enables adaptive maintenance schemes as well as scalable implementation across the disparate turbine fleets utilized across the U.S. power grid. Ultimately, integration with SCADA and IoT not merely increases the level of operational intelligence but provides a platform for autonomous maintenance ecosystems to support the digital transformation initiatives of the American electric power industry.

Strategic Considerations and Future Work

Limitations

Despite promising results exhibited by fault detection models incorporating artificial intelligence, there are limitations to be recognized—namely, limitations in the quality and granularity of available input data. Successful deployment of predictive algorithms is contingent on access to high-frequency, high-resolution sensor data reflecting the rapid changes characteristic of turbine operations. Legacy turbines, particularly in older U.S. infrastructure, are typically outfitted with low-bandwidth data acquisition systems that cannot sense minute, brief faults appearing in millisecond-level vibration spikes or micro-variations in fuel-air ratio. Insufficient data of this type can restrict model sensitivity, lowering early fault detection accuracy and interfering with the detection of pre-failure behaviors. Sensors themselves will further degrade or drift with time, injecting noise or bias into the data that will degrade model performance unless rigorous calibration and verification processes are imposed.

An additional constraint is the infrastructure needed to enable real-time analytics. Fault detection using AI requires strong computational power, whether at the edge or in data centers, to process and reason about massive amounts of streaming data. Smaller power providers or legacy power plants might not have the technical and financial means to implement such platforms, constraining scalability. Interoperability is still another problem to be encountered in integrating AI into SCADA architectures of various kinds and proprietary nature. In the absence of standardized data protocols, as well as a common cybersecurity model, integrating AI into mixed-vendor environments is risky and time-consuming. Therefore, while there is indeed a clear road ahead with AI fault detection, its deployment, in reality, is presently hampered by technical roadblocks to be addressed rationally through modernization, investment, and policy alignment.

Research Extensions

Future steps for the development of fault detection in gas turbines using AI include deeper integration of time-series predictive models and deep learning architectures with improved capabilities to identify long-term temporal relationships and sophisticated fault development patterns. RNN, LSTM, and TCN architectures are particularly suitable for modeling temporal turbine sequences and have the potential to

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

predict fault incidence, as well as the time and development of mechanical failures. Their integration will expand fault prognostic capabilities toward real-time dynamic scheduling of maintenance as the condition of the equipment evolves. Another direction is to merge deep learning with conventionally statistical diagnostic approaches to produce explainable and high-performing systems, combining domain-specific rules with data-driven reasoning.

An important research focus in the future is the detection of multi-fault and cascade failure regimes, wherein multiple interdependent components degrade concurrently or initiate one another sequentially. Existing classification frameworks are mainly pre-trained to identify single-point anomalies; yet, real-world turbine failures are typically systemic, encompassing a sequence of events spanning across the combustion, mechanical, and thermal subdomains. Extending model capabilities to identify such compound failure regimes needs to be supported by augmented datasets annotated with rich failure histories, along with new model architectures compellingly modeling hierarchical dependencies and cause-effect couplings. Simulated worlds and digital twins of turbine systems may also be exploited to create synthetic multi-fault data for training. Developing these capabilities will markedly increase fault detection accuracy and response strategies, paving the way for predictive maintenance frameworks truly competitive with human expertise.

Policy and Investment Implications

The deployment of AI-based fault detection in the US energy sector is well-aligned with national initiatives in infrastructure upgrading, energy resilience, and environmentally friendly operations. As the US Department of Energy (DOE) and its allied government entities drive the development of smart grids and industrial digitalization, investments in intelligent maintenance solutions will find strong support through competitive grants and government-private sector collaboration. By proving verifiable improvements in efficiency, safety, and emissions management, fault detection using AI technology will qualify under initiatives like the Infrastructure Investment and Jobs Act (IIJA), which distributes considerable funding to critical grid assets and technology development. Utilities implementing such systems will, in addition, have firmer grounds to request rate recovery and regulatory clearance for expenditures in the form of supported efficiency and cost savings through data-informed strategies. Policy support is also provided in the areas of workforce transformation and cyber security.

As AI modifies the working environment of the gas turbine, there is a corresponding requirement for training courses to support technicians and engineers with capabilities to interpret machine learning output and cope with SCADA systems integrated with AI. Federal support can be provided through funding for education and skills development initiatives to bridge the skills gap. Concomitantly, national policy will have to set criteria and standards for the secure implementation of AI within critical energy infrastructure, preventing data breaches and model tampering risks. Strategic investment in such initiatives guarantees not just deployment success for AI in gas turbine monitoring but, by so doing, the development of a future-proof operative environment supporting long-term energy self-sufficiency, sustainability aspirations, as well as technological supremacy in the industrial deployment of AI.

VIII. Conclusion

The central objective of this research was to build and test an AI-driven fault detection system to identify early signs of failure in gas turbine engines and apply specifically to deployment within the energy sector of the U.S. The dataset used in this research entails high-resolution operational parameters gathered from different industrial gas turbines operating within U.S. energy facilities. Engine operational data includes multiple time-dependent measurements that monitor essential parameters like turbine temperature at the inlet and outlet and rotational speed, torque measurements alongside vibration levels and power output, fuel rate and intake pressure, and exhaust temperature and oil temperature. Real-time data acquisition from embedded sensor arrays allowed researchers to track turbine performance throughout changing operational states at sub-minute data points. To support strong and interpretable fault detection, we utilized a variety of machine learning models, each chosen for its specific strength at discriminating operations from fault conditions in gas turbines. We used a multi-metric-based evaluation approach that combined statistical validity with operational applicability to assess each model's fault detection capability. Logistic Regression

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

attained the highest accuracy, followed very closely by Random Forest. XG-Boost attained the lowest accuracy of all three at about. The use of AI-driven fault detection under predictive maintenance has the potential to revolutionize U.S. power plants using gas turbines as the primary source of electricity generation. AI-powered fault detection is critical to improving grid resilience within the U.S. energy infrastructure by assuring that peak-demand-balancing and grid-stabilizing gas turbines remain fault-free. One of the strongest capabilities of applying artificial intelligence-based fault detection in the American energy sector lies in its smooth integration with Industrial Internet of Things (IIoT) devices and Supervisory Control and Data Acquisition (SCADA) systems. Future steps for the development of fault detection in gas turbines using AI include deeper integration of time-series predictive models and deep learning architectures with improved capabilities to identify long-term temporal relationships and sophisticated fault development patterns. RNN, LSTM, and TCN architectures are particularly suitable for modeling temporal turbine sequences and have the potential to predict fault incidence, as well as the time and development of mechanical failures.

References

Abdelhamid, K., Ammar, T. B., & Laid, K. (2021, January). Artificial Intelligent in Upstream Oil and Gas Industry: A Review of Applications, Challenges and Perspectives. In International Conference on Artificial Intelligence and its Applications (pp. 262-271). Cham: Springer International Publishing.

Adebiyi, A. A.& Habyarimana, M., (2025). A Review of Artificial Intelligence Applications in Predicting Faults in Electrical Machines. Energies, 18(7), 1616.

Ahmed, A., Jakir, T., Mir, M. N. H., Zeeshan, M. A. F., Hossain, A., hoque Jui, A., & Hasan, M. S. (2025). Predicting Energy Consumption in Hospitals Using Machine Learning: A Data-Driven Approach to Energy Efficiency in the USA. Journal of Computer Science and Technology Studies, 7(1), 199-219.

Agbaji, A. L. (2021, March). An Empirical Analysis of Artificial Intelligence, Big Data and Analytics Applications in Exploration and Production Operations. In International Petroleum Technology Conference (p. D101S043R001). IPTC.

Barua, A., Karim, F., Islam, M. M., Das, N., Sumon, M. F. I., Rahman, A., ... & Khan, M. A. (2025). Optimizing Energy Consumption Patterns in Southern California: An AI-Driven Approach to Sustainable Resource Management. Journal of Ecohumanism, 4(1), 2920-2935.

Chelliah, P. R., Jayasankar, V., Agerstam, M., Sundaravadivazhagan, B., & Cyriac, R. (2023). The Power of Artificial Intelligence for the Next-Generation Oil and Gas Industry: Envisaging AI-inspired Intelligent Energy Systems and Environments. John Wiley & Sons.

Chowdhury, M. S. R., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Barua, A., Chouksey, A., & Chowdhury, B. R. (2024). PREDICTIVE MODELING OF HOUSEHOLD ENERGY CONSUMPTION IN THE USA: THE ROLE OF MACHINE LEARNING AND SOCIOECONOMIC FACTORS. The American Journal of Engineering and Technology, 6(12), 99-118.

Chouksey, A., Shovon, M. S. S., Islam, M. R., Chowdhury, B. R., Ridoy, M. H., Rahman, M. A., & Amjad, M. H. H. (2025). Harnessing Machine Learning to Analyze Energy Generation and Capacity Trends in the USA: A Comprehensive Study. Journal of Environmental and Agricultural Studies, 6(1), 10-32.

Fazle, A. B., Prodhan, R. K., & Islam, M. M. (2023). AI-Powered Predictive Failure Analysis In Pressure Vessels Using Real-Time Sensor Fusion: Enhancing Industrial Safety And Infrastructure Reliability. American Journal of Scholarly Research and Innovation, 2(02), 102-134.

Gamal, H. (2025). Technological Integration and Innovative Strategies Harnessing Artificial Intelligence for Operational Excellence. In Building Business Knowledge for Complex Modern Business Environments (pp. 237-270). IGI Global.

Gazi, M. S., Barua, A., Karim, F., Siddiqui, M. I. H., Das, N., Islam, M. R., ... & Al Montaser, M. A. (2025). Machine Learning-Driven Analysis of Low-Carbon Technology Trade and Its Economic Impact in the USA. Journal of Ecohumanism, 4(1), 4961-4984.

Hasan, M. R. (2024). Revitalizing the electric grid: A machine learning paradigm for ensuring stability in the USA. Journal of Computer Science and Technology Studies, 6(1), 141-154.

Hossain, M., Rabbi, M. M. K., Akter, N., Rimi, N. N., Amjad, M. H. H., Ridoy, M. H., ... & Shovon, M. S. S. (2025). Predicting the Adoption of Clean Energy Vehicles: A Machine Learning-Based Market Analysis. Journal of Ecohumanism, 4(4), 404-426.

Hossain, A., Ridoy, M. H., Chowdhury, B. R., Hossain, M. N., Rabbi, M. N. S., Ahad, M. A., ... & Hasan, M. S. (2024). Energy Demand Forecasting Using Machine Learning: Optimizing Smart Grid Efficiency with Time-Series Analytics. Journal of Environmental and Agricultural Studies, 5(1), 26-42.

Hossain, S., Hasanuzzaman, M., Hossain, M., Amjad, M. H. H., Shovon, M. S. S., Hossain, M. S., & Rahman, M. K. (2025). Forecasting Energy Consumption Trends with Machine Learning Models for Improved Accuracy and Resource Management in the USA. Journal of Business and Management Studies, 7(1), 200-217.

Hossain, M. S., Mohaimin, M. R., Alam, S., Rahman, M. A., Islam, M. R., Anonna, F. R., & Akter, R. (2025). AI-Powered Fault Prediction and Optimization in New Energy Vehicles (NEVs) for the US Market. Journal of Computer Science and Technology Studies, 7(1), 01-16.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 658 – 678
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6773

Kabashkin, I., Perekrestov, V., & Pivovar, M. (2025). AI-Driven Fault Detection and Maintenance Optimization for Aviation Technical Support Systems. Processes, 13(3).

Kamkar, M., Leonard, K. C., Ferrer, I., Loo, S. C. J., Biddinger, E. J., Brady, D., ... & Serrano, J. F. (2024). Artificial Intelligence (AI) for sustainable resource management and chemical processes. ACS Sustainable Resource Management, 1(2), 178-180.

Koroteev, D., & Tekic, Z. (2021). Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. Energy and AI, 3, 100041.

Lee, H. (2024). An AI-Driven Strategy for Non-Invasive Fault Detection: Techniques, Applications, and Innovations (Doctoral dissertation, Oakland University).

Mohamed Almazrouei, S., Dweiri, F., Aydin, R., & Alnaqbi, A. (2023). A review on the advancements and challenges of artificial intelligence-based models for predictive maintenance of water injection pumps in the oil and gas industry. SN Applied Sciences, Nor, A. K. M., Pedapati, S. R., & Muhammad, M. (2021). Application of Explainable AI (Xai) For Anomaly Detection and Prognostic of Gas Turbines with Uncertainty Quantification. Preprints, 2021, 2021090034.5(12), 391.

Rahman, M. K., Dalim, H. M., Reza, S. A., Ahmed, A., Zeeshan, M. A. F., Jui, A. H., & Nayeem, M. B. (2025). Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis: Insights from Shell's Market Dynamics. Journal of Business and Management Studies, 7(1), 44-61.

Reza, S. A., Hasan, M. S., Amjad, M. H. H., Islam, M. S., Rabbi, M. M. K., Hossain, A., ... & Jakir, T. (2025). Predicting Energy Consumption Patterns with Advanced Machine Learning Techniques for Sustainable Urban Development. Journal of Computer Science and Technology Studies, 7(1), 265-282.

Sharma, P., Cirrincione, M., Mohammadi, A., Cirrincione, G., & Kumar, R. R. (2024). An overview of artificial intelligence-based techniques for PEMFC system diagnosis. IEEE Access.

Shil, S. K., Chowdhury, M. S. R., Tannier, N. R., Tarafder, M. T. R., Akter, R., Gurung, N., & Sizan, M. M. H. (2024). Forecasting Electric Vehicle Adoption in the USA Using Machine Learning Models. Journal of Computer Science and Technology Studies, 6(5), 61-74.

Shovon, M. S. S., Gomes, C. A., Reza, S. A., Bhowmik, P. K., Gomes, C. A. H., Jakir, T., ... & Hasan, M. S. (2025). Forecasting Renewable Energy Trends in the USA: An AI-Driven Analysis of Electricity Production by Source. Journal of Ecohumanism, 4(3), 322-345.

Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. K., Pant, L., & Debnath, P. (2024). Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models. Journal of Economics, Finance and Accounting Studies, 6(5), 112-122.