Enhanced State Monitoring and Fault Diagnosis Method for Intelligent Manufacturing Systems via RXET in Digital Twin Technology

Min Li
School of Information Engineering
Jiangxi University of Technology, NanChang 330029, JiangXi, China

***Abstract*—To maintain efficiency and continuity in Industry 4.0, intelligent manufacturing systems use enhanced problem detection and condition monitoring. Existing models typically miss uncommon and essential errors, causing expensive downtimes and lost production. ResXEffNet-Transformer (RXET), a hybrid deep learning model, improves defect identification and predictive maintenance by integrating ResNet, Xception, EfficientNet, and Transformer-based attention processes. The algorithm was trained on a five-year Texas industrial dataset using IoT-enabled gear and digital twins. To manage data imbalances and temporal irregularities, a strong preprocessing pipeline included Dynamic Skew Correction, Temporal Outlier Normalization, and Harmonic Temporal Encoding. The Adaptive Statistical Evolutionary Selector (ASES) optimized feature selection using the Stochastic Feature Evaluator (SFE) and Evolutionary Divergence Minimizer (EDM) to increase prediction accuracy. The RXET model beat traditional methods with 98.9% accuracy and 99.2% AUC. Two new performance metrics, Temporal Fault Detection Index (TFDI) and Fault Detection Variability Coefficient (FDVC), assessed the model's capacity to identify problems early and consistently across fault kinds. Simulation findings showed the RXET's superiority in anticipating uncommon but essential errors. Pearson correlation (0.93) and ANOVA (F-statistic: 8.52) validated the model's robustness. The sensitivity study showed the best performance with moderate learning rates and batch sizes. RXET provides a complete, real-time problem detection solution for intelligent industrial systems, improving predictive maintenance and addressing challenges in Industry 4.0, digital twin technology, IoT, and machine learning. The proposed RXET model enhances operational reliability in intelligent manufacturing and sets a foundation for future advancements in predictive analytics and large-scale industrial automation.**

***Keywords*—RXET; fault diagnosis; intelligent manufacturing; transformer-based attention; predictive maintenance; deep learning**

I. Introduction

Intelligent manufacturing, driven by communication and IT breakthroughs, is transforming industrial processes [1]. Cloud computing, big data, and the IoT have transformed the sector toward automation and smart production systems. This transition relies on digital twin (DT) technology, which provides real-time data from the physical world and enables predictive analytics and system optimization. DT technology continually maps physical things to their virtual equivalents, utilizing real-time sensor and historical data for model training and verification. This constant data flow allows the virtual model to identify possible problems and transmit feedback to the physical system for remedial measures, improving production line efficiency and fault detection. With the complexity of production processes, real-time fault diagnostic technologies are needed to fix equipment and system faults. Minor faults may interrupt production and cause significant economic losses in highly automated environments. Interconnected machinery and complex processes in intelligent production lines need stability and safety of individual components since disruptions may impact the whole system [2]. Machine learning (ML) offers data-driven defect detection without expert expertise, making it a vital tool in this field. Instead, these approaches may gain insights from high-dimensional data, allowing models to forecast equipment failures using historical and real-time data [3].

Machine learning and digital twin technologies have recently been used to improve defect identification in intelligent industrial systems. Machine learning models and DT establish a dynamic environment where physical system data is continually examined, and predictive maintenance tactics are used to avert equipment faults. Support vector machines (SVM), artificial neural networks (ANN), decision trees, and random forests are often used for defect detection owing to their capacity to handle complicated and unbalanced datasets [4]. Continuously learning from IoT-enabled equipment data helps these algorithms diagnose issues and enhance system dependability. Ensemble learning is a breakthrough in this discipline. Ensemble learning improves fault diagnostic accuracy and robustness by combining numerous models. Ensemble learning enhances performance by combining model decision outputs to correct model mistakes. Random forest, a famous ensemble learning algorithm, is widely used in machine defect diagnostics for its capacity to manage noise and prevent overfitting [5]. Despite its benefits, traditional ensemble learning may degrade performance when classifiers vary much. Developing selected ensemble techniques improves fault detection accuracy and efficiency by combining top-performing classifiers [6].

Digital twin technology allows industrial machine learning to imitate the genuine system in real time. Data transfer, VMware, and the actual thing make a digital twin. Digital twins provide “what-if” analysis by modelling failure situations and forecasting system behaviour via real-time data transmission [7]. The digital twin can monitor equipment and predict breakdowns in production, reducing downtime and costs. Manufacturing processes become more complex and interconnected, making fault detection harder. Traditional machine-level fault detection identifies motor and sensor issues. Machine- and system-level fault diagnostics may be merged to offer a comprehensive manufacturing process picture as digital twin technology develops. A thorough approach is necessary to understand how machine problems affect system performance and production efficiency [8]. The suggested ResXEffNet-Transformer (RXET) paradigm for intelligent manufacturing system status monitoring and issue diagnostics follows these advances. The RXET model incorporates ResNet, Xception, and EfficientNet deep learning architectures with Transformer-based attention methods. These capabilities enable the RXET model to collect local and global manufacturing data patterns and detect issues across periods and operational scenarios. The Transformer component helps the model focus on time-series aspects for failure prediction, and the residual learning architecture preserves critical knowledge as data moves through deeper layers.

Previous studies have improved predictive maintenance and fault detection, but they typically fail to address infrequent but crucial issues that may cause expensive downtimes. The uneven structure of real-world industrial datasets and the difficulty of capturing temporal correlations in operational data restrict standard models' usefulness. Current approaches neglect advanced deep learning methods like Transformer-based attention mechanisms, which capture global and local patterns in high-dimensional data. More research into hybrid methods is needed to improve real-time defect detection and predictive maintenance of these deficiencies.

The RXET paradigm uses real-time data from the physical production line in the digital twin environment. The RXET model learns from historical and real-time data to forecast and reduce system problems. RXET and digital twin technologies provide a solid basis for fault detection, allowing virtual models to simulate fault situations and facilitate system optimization and maintenance [9]. Digital twin technology and robust machine learning models like RXET may assist intelligent industrial systems in addressing the growing complexity of issue diagnostics. Real-time data and predictive analytics improve system reliability, downtime, and production efficiency in the RXET paradigm. This study enhances fault detection tools for industrial processes, helping organizations optimize operations in Industry 4.0.

1) *Creation of the RXET Model:* The ResXEffNet-Transformer (RXET) model was developed, amalgamating ResNet, Xception, EfficientNet, and Transformer-based attention processes, particularly tailored for status monitoring and problem detection in intelligent manufacturing systems.

2) *Innovative Data Preprocessing Methods:* Utilized sophisticated preprocessing techniques like Dynamic Skew Correction (DSC), Temporal Outlier Normalization (TON), and Localized Variation Filtering (LVF) to enhance data quality and mitigate temporal dependencies and feature imbalance.

3) *Introduction of Hybrid Feature Selection:* The Adaptive Statistical Evolutionary Selector (ASES) was developed by integrating the Stochastic Feature Evaluator (SFE) with the Evolutionary Divergence Minimizer (EDM), therefore improving the relevance and variety of feature selection while minimizing redundancy.

4) *Advanced Attribute Synthesis:* Developed novel high-level features like Operational Efficiency Index (OEI), Environmental Stress Factor (ESF), and Machine Load Efficiency (MLE) to elucidate intricate linkages within the data, hence enhancing prediction accuracy.

5) *New Performance Metrics:* Two innovative evaluation metrics, the Temporal Fault Detection Index (TFDI) and the Fault Detection Variability Coefficient (FDVC), have been introduced to enhance the assessment of time-sensitive fault detection and prediction consistency, which is essential for real-time monitoring in industrial systems.

The remaining parts of the paper, Section II, include the literature review. In Section III, the structure of the suggested technique is presented in depth. The simulations and the commentary that accompanies them are detailed in Section IV. Discussion is given in Section V and finally, the paper is concluded in Section VI.

II. Related Work

Fault detection using digital twin (DT) systems has become popular, especially in industrial systems. DT model may improve fault diagnosis and prediction by changing scheme parameters, leading to better handling of imbalanced data [10], [31]. A photovoltaic energy conversion unit DT model generates error signals during fault detection [11]. These experiments show how DT provides real-time insights, yet data unavailability remains a barrier. Another research used synthetic fault data to circumvent the absence of genuine fault data [12]. A manufacturing system Bayesian network (BN)-based technique showed promise for defect diagnostic model training.

Machine learning (ML) is another hot topic in defect detection. Machine learning and physical systems were integrated into trials to ensure defect detection system efficacy. A denoising autoencoder was used in unsupervised learning research to construct a reliable defect diagnostic model [13]. Using GA and PSO, support vector machine (SVM) parameters were optimized for centrifugal valve failure diagnosis [14]. Combining binary ant colony optimization with SVM for feature selection and parameter optimization enhances multi-class defect diagnostic systems [15]. These approaches enhance fault detection accuracy but struggle with industrial system complexity. Also improving is DT-based predictive maintenance. Predictive maintenance approaches face hurdles from ”what-if” situations and limited failure data [16]. Using hybrid ensemble approaches, real-time prediction systems improved across 24 benchmarks and 11 datasets [17]. Combining numerous models enhances system performance, as seen below. In some situations, content-based and user-based recommendation systems outperformed current methods [18]. Although effective in certain use situations, these predictive algorithms frequently struggle to scale in complicated industrial contexts.

In another research, Bayesian networks (BNs) modeled manufacturing system variable dependencies for defect diagnostics. Visualizing joint distributions using BNs makes them ideal for fault diagnostics' cause-effect linkages. Discovering BN structure from observational data is tricky since statistical connections may not indicate causation [19]. Different approaches, like the Hill Climbing algorithm [20] and Prototypical Constraint (PC), have been employed to approximate BN structures. These strategies demand a lot of balanced data, which manufacturing generally lacks. Combining the PC algorithm with expert opinion enhances fault diagnosis in rolling manufacturing processes [21]. Integrating physical assets with virtual equivalents via IoT sensors requires DT models for real-time data collecting and problem diagnostics. Recent work has expanded DT in predictive maintenance. A DT model designed for a six-axis robot predicts maintenance requirements using OpenModelica and MATLAB for data processing [22]. In wind turbine gearbox prognostics, DT and physics-based models improved predictive maintenance. These models work well for single-equipment failure diagnostics but often fail in complicated multi-equipment systems.

They limited DT applications in multi-equipment systems. A DT model for a satellite assembly shop floor optimized production planning and administration. Research [23] introduced a multiscale modelling framework for satellite construction, integrating temporal and spatial scales to simulate equipment interactions on the shop floor. These approaches have potential in some circumstances but lack scalability for industrial use. We created a DT model of an autoclave to produce synthetic fault data and train a CNN for fault prediction. DT may improve machine learning models by generating artificial data. Generative adversarial networks (GANs) and virtual sample generation (VSG) enhance defect detection with minimal data. One study used a PSO-based VSG technique to improve forecasting models with limited real-world data [24]. Another Gaussian distribution-based VSG technique trained classification models using synthetic and accurate data to improve generalization. Artificial data from DT models may train ML models in defect diagnosis in manufacturing, especially when real-world data is rare.

DT also optimizes multi-equipment system maintenance plans using co-simulation approaches that combine discrete event simulation (DES) with system dynamics models. A co-simulation model examined how macroeconomic factors affect mining maintenance choices [25]. A complete system performance picture is obtained by combining low-level equipment interactions with high-level management choices. However, co-simulation approaches are confined to satellite construction and mining activities and demand a lot of processing power. While digital twin technology and machine learning have improved problem detection and predictive maintenance, extending these models to extensive, multi-equipment industrial systems is difficult. Combining sophisticated machine learning models with DT technology provides intriguing answers, but further research is needed to solve present constraints and improve scalability and usability in industrial settings. The literature is presented in a summarized way in Table I.

Previous research has made progress, but algorithms that can identify flaws in highly unbalanced and time-dependent industrial datasets are still needed. Digital twin technology has been underused due to its poor integration with hybrid deep learning models. Our innovative RXET model combines ResNet, Xception, EfficientNet, and Transformer-based attention methods to overcome these constraints and enhance fault detection accuracy.

III. Proposed Method

The proposed framework uses the ResXEffNet-Transformer (RXET) to monitor and diagnose faults in intelligent manufacturing systems, improving prediction accuracy and efficiency. Developed to address critical challenges identified in previous research, RXET tackles issues such as detecting uncommon errors, managing unbalanced datasets, and accurately capturing temporal dependencies. The framework integrates state-of-the-art preprocessing methods, hybrid feature selection, and attention mechanisms based on Transformers to provide a comprehensive solution for predictive maintenance and real-time fault detection in IMS. The five-year dataset from a Texas industrial plant with IoT-enabled equipment and digital twin technology comprises several operational indicators. Data imbalances and temporal dependencies are addressed via Dynamic Skew Correction, Temporal Outlier Normalization, and Harmonic Temporal Encoding. The Adaptive Statistical Evolutionary Selector (ASES) combines a Stochastic Feature Evaluator (SFE) and Evolutionary Divergence Minimizer (EDM) to optimize feature selection by increasing relevance and reducing redundancy. Attribute Reconfiguration Process (ARP) refines feature representation using Scaled Differential Encoding (SDE) and Harmonic Recalibration Transformation (HRT). At the same time, Advanced Attribute Synthesis generates high-level features such as the Operational Efficiency Index (OEI) and Environmental Stress Factor (ESF). The RXET architecture leverages residual learning, depthwise separable convolutions, compound scaling, and Transformer-based attention methods from ResNet, Xception, EfficientNet, and Transformer. Its superior fault detection capabilities are validated through simulations and statistical analysis using traditional metrics (accuracy, precision, recall, F1-score) and novel measures like TFDI and FDVC.

A. Dataset Description

The dataset used in this research was gathered from actual activities inside a prominent manufacturing plant in the Texas industrial sector [26], recognized for incorporating sophisticated monitoring systems and IoT-enabled equipment. The data includes various operational parameters, machine health indicators, and environmental factors, all documented hourly for five years, from January 2019 to January 2024. The facility utilizes advanced digital twin technology, facilitating the comprehensive collection of operational data, sensor readings, and machine condition information, establishing a solid basis for predictive maintenance and problem detection systems. Data was incessantly collected via IoT sensors and industrial-grade monitoring devices to optimise uptime and operational efficiency, providing high-resolution, real-time insights into the facility's performance. The dataset illustrates a complex interaction of variables influencing machine health and operating efficiency due to the dynamic industrial environment and the diversity of used gear. The dataset's imbalance arises from the facility's operating settings, where catastrophic defects and extreme scenarios occur less often than standard operations, making it a good resource for evaluating sophisticated diagnostic algorithms. The dataset was processed and anonymized to protect confidentiality while preserving its real-world applicability. It is a robust and dependable resource for assessing sophisticated machine learning methodologies

TABLE I. LITERATURE REVIEW SUMMARY

Ref	Technique Used	Objective Achieved	Limitations
[10]	Digital Twin (DT) model with scheme parameter update	Improved fault diagnosis and prediction by handling imbalanced data	Data unavailability remains a challenge
[11]	DT model for photovoltaic energy conversion unit (PVECU)	Real-time fault detection with error generation during fault conditions	Limited applicability to energy systems
[12]	Simulated data generation for fault conditions using synthetic fault data	Circumvented the absence of real fault data in industrial systems	Simulation accuracy relies on the quality of generated data
[13]	Denoising autoencoder for unsupervised learning in ML	Developed a robust fault diagnosis model using unsupervised learning	Lacks labeled data for validation, which can affect results
[14]	Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) with SVM	Optimized SVM parameters for centrifugal valve fault diagnosis	Complex parameter tuning in industrial systems
[15]	Binary ant colony optimization with SVM	Enhanced multi-class defect diagnosis systems by optimizing feature selection	Computational complexity for large-scale systems
[16]	Hybrid ensemble techniques for predictive maintenance	Improved performance across 24 benchmarks and 11 datasets	Scalability issues in complex industrial environments
[18]	Content-based and user-based recommendation systems	Outperformed traditional methods in specific use cases	Struggles with scalability in large-scale industrial environments
[19]	Bayesian Networks (BN) in manufacturing systems	Modeled variable dependencies for effective defect diagnostics	Complexity in discovering BN structure from observational data
[22]	DT model for six-axis robot using OpenModelica and MATLAB	Improved predictive maintenance in single-equipment systems	Difficult to scale to multi-equipment systems
[24]	Virtual Sample Generation (VSG) with PSO	Enhanced model forecasting performance with limited data	Performance depends on the quality of synthetic data
[25]	Co-simulation (DES and system dynamics) for maintenance optimization	Examined how macroeconomic factors affect multi-equipment maintenance decisions	Requires significant computational resources and is limited to specific industries

in status monitoring and problem detection. Fig. 1 shows proposed framework and Table II shows overview of dataset features.

B. Data Preprocessing Steps

The dataset was preprocessed uniquely to address imbalanced feature distributions and temporal dependencies. The Dynamic Skew Correction (DSC) approach fixes skewness depending on data imbalance. The equation for correction is:

$$Y' = \frac{Y - \eta}{(\delta^k)} \quad (1)$$

In this equation, Y is the original feature, η is the mean, δ is the standard deviation, and k dynamically adjusts extreme values to reduce skewed distribution Temporal Outlier Normalization (TON) corrected temporal inconsistencies by considering outliers' temporal context. Determine this normalization:

$$Z'_r = \frac{Z_r}{\bar{Z}_{r-m:r+m} + \beta \cdot \theta_{r-m:r+m}} \quad (2)$$

Z_r represents the feature value at time r, $\bar{Z}_{r-m:r+m}$ and $\theta_{r-m:r+m}$ represent the mean and standard deviation across m time steps, and β controls outlier sensitivity. Anomalies are adjusted by temporal context. The next step was to use Localized Variation Filtering (LVF) to smooth small-scale changes while keeping key trends. This was done with:

$$W'_j = W_j \cdot \left(1 - \lambda \cdot \frac{|W_j - W_{j-1}|}{|W_j|}\right) \quad (3)$$

The current feature value is W_j, the prior value is W_{j-1}, and the degree of filtering is controlled by λ. This approach smooths slight swings while maintaining data trends. Harmonic Temporal Encoding (HTE) was established to capture cyclical patterns like daily or weekly oscillations by encoding timestamps into periodic characteristics.

$$HTE_s = \sin\left(\frac{2\pi \cdot s}{T}\right), \quad \cos\left(\frac{2\pi \cdot s}{T}\right) \quad (4)$$

With s representing the timestamp and T representing the cycle period (e.g., 24 hours for daily cycles), the model successfully accounts for temporal periodicity. Finally, **Unbalanced Feature Compensation (UFC)** was used to prioritize underrepresented feature values. The procedure is explained by:

$$Q'_v = Q_v \cdot \left(1 + \gamma \cdot \frac{1}{h_v}\right) \quad (5)$$

Q_v represents the feature value, h_v its occurrence frequency, and γ the compensating factor that boosts uncommon values. Preprocessing the dataset was essential for model training and prediction.

TABLE II. DATASET FEATURES OVERVIEW

S.No	Features	Short Description	S.No	Features	Short Description
1	Vibration Level	Sensor readings for machine vibrations	19	Machine Health Index	Overall machine health status
2	Temperature Readings	Recorded temperature values from machines	20	Failure Mode Indicators	Binary indicators of potential failure modes
3	Pressure Data	Pressure readings captured from machines	21	Maintenance Logs	Maintenance events logged
4	Acoustic Signals	Sound level data captured from machines	22	Previous Fault Occurrences	Historical fault occurrences in the machine
5	Humidity Levels	Humidity data near machinery	23	Predictive Maintenance Scores	Predictive metrics for required maintenance
6	Motor Speed	Rotational speed of motors	24	Component Degradation Index	Index representing component wear and degradation
7	Torque Data	Torque readings from machine motors	25	Real-time Performance Index	Performance level of machinery in real-time
8	Energy Consumption	Energy usage of machinery	26	Machine Start/Stop Events	Binary log of machine start or stop events
9	Production Rate	Rate of production during operation	27	Downtime Incidents	Logs of machine downtime occurrences
10	Tool Wear Rate	Wear and tear rate of machine tools	28	Fault Trigger Timestamps	Timestamps for triggered faults
11	Machine Utilization Rate	Utilization percentage of machine capacity	29	Controller Setpoints	Target values set by the machine controller
12	Cycle Time per Operation	Time taken per operational cycle	30	Actual vs Setpoint Values	Difference between target and actual values
13	Idle Time	Time when machine is idle	31	Alarm Trigger Data	Binary indication of alarms triggered
14	Machine Load Percentage	Percentage of machine load during operations	32	Repair Logs	Logs of machine repairs conducted
15	Ambient Temperature	Ambient temperature around the machine	33	Spare Part Usage	Amount of spare parts used in maintenance
16	Humidity	Humidity levels in the facility	34	Anomaly Scores	Score indicating deviation from normal operation
17	Air Quality Index	Air quality measurements in the facility	35	Fault Probability	Probability score of potential faults
18	Machine Health Index	Overall health status of the machine	36	Operator Shift Data	Data of operator shifts during operations

C. Feature Selection Process

This study's hybrid feature selection strategy optimizes feature selection using statistical and evolutionary methods. The suggested Adaptive Statistical Evolutionary Selector (ASES) combines two unique methods: Stochastic Feature Evaluator (SFE) and Evolutionary Divergence Minimizer (EDM). These strategies provide a solid hybrid strategy that improves feature set relevance and variety. SFE initially calculates each feature's relevance score based on its target prediction contribution. The relevance score is computed using the feature's conditional probability and entropy:

$$R_x = \frac{P(T|X) \cdot \psi(X)}{\tau(X) + \delta} \tag{6}$$

P(T—X) is the conditional probability of the target T given the feature X, $\psi(X)$ is the entropy, $\tau(X)$ is its standard deviation, and δ is a tiny constant to avoid division by zero. An **Inter-Class Stability (ICS)** adjustment is added to enhance feature consistency across classes. The expression is:

$$ICS_x = \frac{1}{1 + \sum_{c=1}^{C} \frac{\nu_c(X)}{\bar{X}_c}} \tag{7}$$

The standard deviation of feature X within class c is $\nu_c(X)$, the mean of feature X for class c is $\bar{X}_c$, and the number of classes is C The final relevance score after this change is:

$$R'_x = R_x \cdot ICS_x \tag{8}$$

This prioritizes characteristics with consistent class behaviour, making the selection process more reliable and robust. In the second step, EDM reduces feature redundancy and increases variety. EDM picks features repeatedly using a fitness function that balances relevance and redundancy after SFE ranks. Definition of fitness score:

$$G_y = R'_y - \gamma \cdot \sum_{z=1}^{n} \rho(Y_y, Y_z) \tag{9}$$

G_y represents feature fitness, R'_y represents adjusted relevance from SFE, $\rho(Y_y, Y_z)$ represents the correlation between feature y and previously selected feature z, and γ regulates relevance-correlation trade-off. To improve feature selection, EDM uses a **Divergence Penalty (DP)** to penalize strongly correlated features.

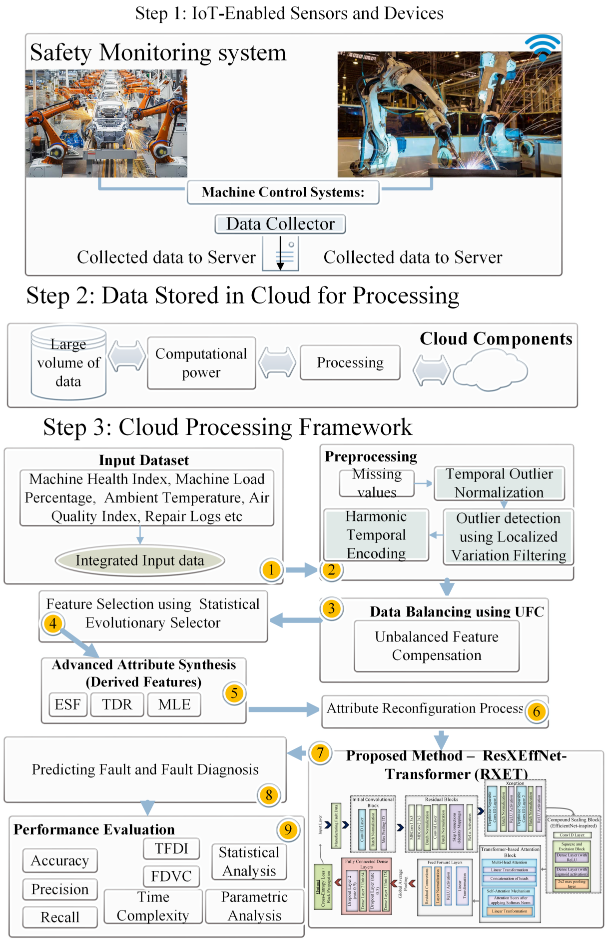

Fig. 1. Proposed framework.

$$DP = \zeta \cdot \left(\sum_{z=1}^{n} \rho(Y_y, Y_z)^2 \right) \tag{10}$$

where ζ represents a penalty factor. The ultimate fitness score for feature y, including the penalty, is:

$$G'_y = G_y - DP \tag{11}$$

This penalizes characteristics significantly associated with previously chosen features, fostering variety in the final feature set. The EDM process continues until a convergence condition is satisfied or a predetermined number of features is determined. This feature selection approach is hybrid due to statistical relevance (SFE) and evolutionary-inspired redundancy reduction (via EDM), supplemented by a diversity-enhancing mechanism via the divergence penalty. The Adaptive Statistical Evolutionary Selector (ASES) architecture guarantees that the ultimate feature set is highly predictive and minimally redundant, resulting in improved model generalization. Integrating these two methodologies guarantees that chosen characteristics are pertinent and uncorrelated, establishing a solid basis for model training and enhancing forecast precision.

D. Advanced Attribute Synthesis

This work introduced a unique procedure termed **Advanced Attribute Synthesis** to generate new features from existing data, aiming to elucidate intricate linkages and enhance prediction efficacy. Multiple advanced features were produced. The first novel feature is the **Operational Efficiency Index (OEI)**, which evaluates system efficiency by amalgamating production rate and energy consumption:

$$OEI = \frac{P_{rate}}{E_{cons} + \alpha} \tag{12}$$

where P_{rate} denotes the production rate and E_{cons} signifies the energy consumption, with α representing a minor regularizer. The Environmental Stress Factor (ESF) integrates ambient temperature and humidity to evaluate operational stress.

$$ESF = T_{ambient} \cdot H_{env} \tag{13}$$

where $T_{ambient}$ represents ambient temperature and H_{env} humidity. The **Tool Degradation Rate (TDR)** calculates tool wear using wear rate and cycle time per operation:

$$TDR = W_{tool} \cdot C_{cycle} \tag{14}$$

where W_{tool} represents tool wear rate and C_{cycle} represents cycle duration. Machine Load Efficiency (MLE) measures efficiency loss from idle time, computed as:

$$MLE = \frac{L_{machine}}{1 + I_{time}} \tag{15}$$

where $L_{machine}$ is the machine load percentage and I_{time} is idle time. Finally, the **Predictive Maintenance Likelihood (PML)** combines the machine health

E. Attribute Reconfiguration Process

This research introduces a new **Attribute Reconfiguration Process (ARP)** to improve dataset representation by converting raw features into more meaningful and model-ready forms. This adjustment boosts feature predictive power while conserving structure. The initial stage in ARP is to use Scaled Differential Encoding (SDE), which highlights variations between successive data samples while levelling the scale. The definition is:

$$SDE_t = \frac{A_t - A_{t-1}}{1 + |A_{t-1}|} \tag{16}$$

where A_t is the current feature value and A_{t-1} is the prior value. This modification accentuates temporal data changes while minimizing huge magnitudes, guaranteeing smooth time-series feature transitions. Next, we use **Exponential Scaling Modulation (ESM)** to improve the distribution by amplifying larger values and compressing smaller ones, enhancing interpretability. We define ESM as:

$$ESM(B) = \text{sign}(B) \cdot \log(1 + \gamma|B|) \tag{17}$$

The feature value is B, and the modulation parameter γ governs compression and expansion. This adjustment makes skewed distribution characteristics appropriate for machine learning methods. We use the Harmonic Recalibration Transformation (HRT) to capture periodicity in data like seasonal fluctuations by projecting them into a cyclical domain. The transition is:

$$HRT(C_t) = \sin\left(\frac{2\pi C_t}{T_c}\right), \quad \cos\left(\frac{2\pi C_t}{T_c}\right) \tag{18}$$

where T_c represents the cyclical behavior period and C_t represents the feature value at time t. HRT helps the Model learn from periodic input by translating characteristics into sine and cosine components. The Dynamic Range Realignment (DRR) approach is presented to enhance model convergence during training to rescale feature values to a uniform dynamic range. The DRR formula is:

$$DRR(D) = \frac{D - \min(D)}{\max(D) - \min(D)} \tag{19}$$

$\min(D)$ and $\max(D)$ represent the least and maximum values of the feature D. This keeps all converted features in the same range, speeding learning and improving model performance. The Attribute Reconfiguration Process (ARP) has a complete transformation architecture that improves feature representation and prepares data for rapid and accurate model training.

F. Classification Model: ResXEffNet-Transformer (RXET)

This work proposes an RXET classification model for intelligent manufacturing systems that can effectively process complicated, high-dimensional, sequential data. Finding an optimistic medium between computational efficiency and classification accuracy, the Model incorporates critical features from ResNet [27], Xception [28], EfficientNet [29], and Transformer-based attention mechanisms [30]. Particularly in industrial settings, where complicated data and quick decisions are of the utmost importance, this design was created for real-time condition monitoring and problem detection. Fig. 2 is the RXET layered design.

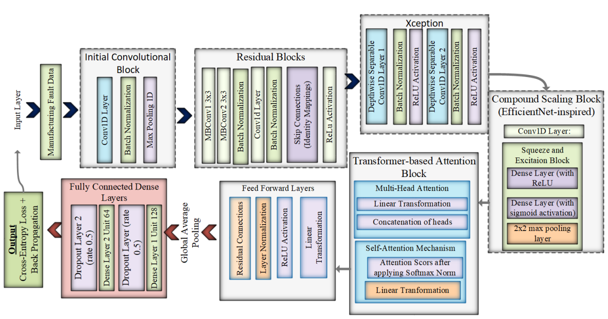

Fig. 2. Proposed framework.

At the core of RXET is the **residual learning framework**, inspired by ResNet, which addresses the vanishing gradient problem using identity mappings that allow gradients to propagate through deeper layers without degradation. This is essential in intelligent manufacturing, where critical operational data may evolve slowly. The residual block is formulated as follows:

$$M = \mathcal{H}(V, \{R_k\}) + V \tag{20}$$

$\mathcal{H}(V, \{R_k\})$ is the residual function that the network learns with parameters $\{R_k$, and M is the output, where V is the input to the residual block. Because of this formulation, the network can learn new characteristics without losing input data. Here is the updated gradient for this block:

$$\frac{\partial \mathcal{J}}{\partial V} = \frac{\partial \mathcal{J}}{\partial M}\left(1 + \frac{\partial \mathcal{H}(V, \{R_k\})}{\partial V}\right) \tag{21}$$

where $\mathcal{J}$ is the loss function. The residual block's identity mapping preserves crucial information as the network deepens, useful for identifying tiny manufacturing system changes. Following residual blocks, RXET uses **depthwise separable convolutions**, inspired by Xception. Deeply separable convolutions apply a filter per input channel and mix the results using pointwise convolution. A mathematical representation is:

$$N = \mathcal{D}(Z_q * V) + Z_p * V \tag{22}$$

where Z_q is the depthwise filter, Z_p is the pointwise filter, $*$ is the convolution operation, and N is the output. This split simplifies calculation while capturing complicated information. Computing efficiency advantage is measured by:

$$\frac{\mathcal{C}_{depthwise}}{\mathcal{C}_{standard}} = \frac{1}{L_f + L_c} \tag{23}$$

where $\mathcal{C}_{depthwise}$ represents depthwise convolution cost, $\mathcal{C}_{standard}$ represents standard convolution cost, L_f represents filters, and L_c represents input channels. Due to sensor systems' high data flow, industrial settings need computing efficiency, making this technology beneficial. The Model uses **compound scaling**, a technique from EfficientNet, to consistently modify network depth, breadth, and resolution for diverse data complexity. We define scaling as:

$$d' = \lambda^u \cdot d_0, \quad l' = \mu^u \cdot l_0, \quad s' = \nu^u \cdot s_0 \tag{24}$$

where d', l', s' denote scaled depth, width, and resolution, λ, μ, and ν are scaling coefficients, and u is the scaling factor. The initial depth, width, and resolution parameters are d_0, l_0, and s_0. Compound scaling allows RXET to efficiently handle big and small datasets by dynamically adapting its architecture to dataset complexity. Results in resource usage:

$$\mathcal{C}_{scaled} = \mathcal{C}_0 \cdot \lambda^{u_1} \cdot \mu^{u_2} \cdot \nu^{u_3} \tag{25}$$

Where $\mathcal{C}_0$ is the initial computing cost, and u_1 u_2 $and u_3$ are the scaling factors for depth, width, and resolution. Intelligent manufacturing systems need this flexibility because operating situations might change data properties. A Transformer-based attention mechanism is added to RXET to capture long-range relationships in sequential data. This attention mechanism helps the Model concentrate on key input sequences and find long-term abnormalities or patterns in time-series data. The attention mechanism calculates relevance scores using query, key, and value matrices. Calculating attention scores:

$$P = \text{softmax}\left(\frac{QR^T}{\sqrt{d_r}}\right)R \tag{26}$$

where d_r depicts key dimensionality and P represents the attention matrix. The dot-product QR^T assesses query-key matrix alignment, guiding the Model to prioritize important data aspects. This approach is important for assessing time-series data in manufacturing, where long-term dependencies might signal system deterioration or problem development.

Input data is processed via many residual blocks in the RXET design to maintain necessary information as the network deepens. Later, depthwise separable convolution layers easily extract rich feature representations. The Model adapts to dataset complexity using compound scaling. Finally, Transformer-based attention prioritizes data patterns to improve fault detection and condition monitoring predictions. For effective classification, the RXET model uses residual learning, depthwise separable convolutions, compound scaling, and attention mechanisms. Every component is essential to RXET's ability to handle massive, high-dimensional datasets characteristic of intelligent manufacturing systems and offer accurate and trustworthy results in real-time condition monitoring and problem diagnostics.

G. Performance Evaluation Metrics

Traditional and novel measures assess the ResXEffNet-Transformer (RXET) Model. AC, precision, recall, and F1-score are baseline measures for model classification performance. Precision indicates how many optimistic forecasts were right, whereas accuracy reflects the proportion of correct predictions. Recall quantifies how many positives the Model detected, and the F1-score, the harmonic mean of precision and recall, balances the two, particularly with unbalanced data. Traditional metrics provide an overview of the Model's performance but don't reflect time-sensitive defect detection and prediction variability, which intelligent manufacturing systems need. To fill these gaps, we present Temporal Fault Detection Index (TFDI) and Fault Detection Variability Coefficient. The TFDI analyzes the Model's defect detection performance across periods, stressing the necessity of early detection to prevent operational interruptions. We define TFDI as:

$$TFDI = \frac{1}{T}\sum_{k=1}^{T}\left(\frac{TP_k}{TP_k + FN_k + \eta}\right)\cdot \exp\left(-\rho \cdot k\right) \tag{27}$$

T represents the number of time windows, TP_k and FN_k represent true positives and false negatives, and ρ is a decay factor that prioritizes early detections. Division by zero is prevented by the constant η. The exponential decay term $(\exp(-\rho \cdot k)$ prioritizes earlier fault identification, emphasizing its relevance. The second new statistic, the Fault Detection Variability Coefficient (FDVC), evaluates the Model's consistency in detecting fault types and fault pattern variability. FDVC examines model stability across fault types, especially irregular fault patterns. We define FDVC as:

$$FDVC = \frac{\sum_{m=1}^{M}\left(\frac{1}{n_m}\sum_{i=1}^{n_m}\left|Q_{im} - \bar{Q}_m\right|\right)}{M} \tag{28}$$

In the equation, M represents the number of fault types, n_m represents the occurrences of fault type m, Q_im represents the prediction score for the i-th instance, and $\bar{Q}\_m$ represents the mean prediction score. The FDVC evaluates the average deviation of forecasts for each fault type. Lower values indicate more consistent and trustworthy predictions. Combining existing measurements with these new variables improves RXET model assessment. Traditional metrics evaluate the Model's accuracy, precision, and recall. In contrast, TFDI and FDVC emphasize early detection and variability, which is crucial for real-time defect monitoring and predictive maintenance in intelligent industrial systems.

IV. Simulation Results

To extensively test the RXET model, a high-performance Dell Core i7 12th Gen system with an 8-core CPU and 32 GB of RAM was used. Python and the SPYDER IDE were used to manage and perform the model trials. Several essential classification model hyperparameters were fine-tuned throughout the assessment to maximize performance. Model training used the Adam optimizer, which has flexible learning rate capabilities, enabling speedier convergence. The learning rate was chosen at 0.001 0.001, which was ideal after numerous trial trials, and the batch size was 64 to balance memory efficiency and gradient estimate accuracy. These parameters were carefully adjusted throughout the studies to maximize the Model's predictive capabilities, notably in defect diagnosis and status monitoring in intelligent manufacturing systems.

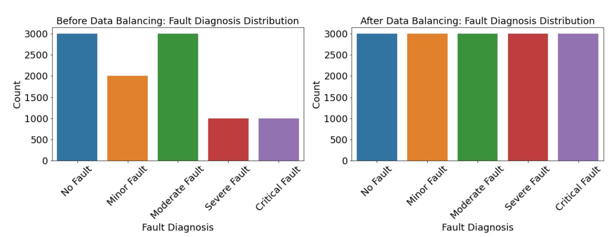

Fig. 3. Distribution of fault diagnosis labels before and after data balancing.

Fig. 3 compares fault diagnostic label distribution before and after data balancing. On the left, the "Before Data Balancing" scenario has a severe fault category imbalance. The data points are mostly "No Fault," "Moderate Fault," and "Minor Fault," with a few "Severe Fault" and "Critical Fault." This mismatch suggests that most observations reflect normal or mild operating circumstances, skewing the dataset. All fault diagnostic labels are evenly distributed in the "After Data Balancing" graphic on the right. Each category—"No Fault," "Minor Fault," "Moderate Fault," "Severe Fault," and "Critical Fault"—has 3,000 data points. This balanced distribution shows that strategies have been implemented to correct the dataset's class imbalance, guaranteeing that the machine learning model will be trained on an equally distributed collection of errors, improving its capacity to identify uncommon but significant defects. The balanced dataset ensures that the Model is sensitive to less common but significant categories like "Severe Fault" and "Critical Fault" without favouring "No Fault" and "Moderate Fault." Fault diagnostic activities need this enhancement to identify infrequent but significant events to preserve operational efficiency and avert system breakdowns.

Fig. 4. Machine operational behaviour and its relationship with energy consumption and component degradation.

The association between machine operation, energy use, and component deterioration is shown in Fig. 4. The first left graphic shows machine utilization during and off hours throughout seven days. Use peaks at work (9 AM–6 PM). Industrial machine use rises during peak output and falls during downtime or decreasing demand. Fault diagnostic category-specific energy usage and component deterioration are shown on the second right. Energy usage often degrades components for predictive maintenance. Colour-coded fault categories indicate serious difficulties when energy use and component deterioration are high. Energy consumption affects machine health since operating stress promotes wear and faults. Daily operating cycles, energy consumption, and deterioration patterns demonstrate machine performance in these two charts for real-time operations and long-term maintenance.

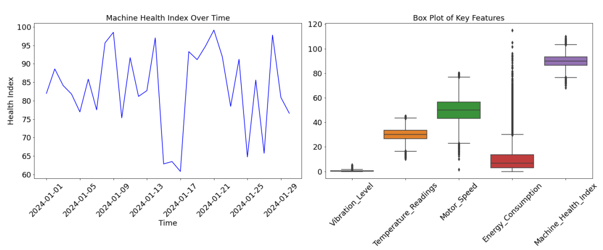

Fig. 5. Machine utilization rate over a week.

Fig. 5 shows two complementary representations that provide insights into machine health and operation. The left line plot represents the January 1–January 30, 2024 Machine Health Index. The Machine Health Index indicates machine status over 60–100 days. This chart may show continuous performance, unexpected drops, and health recoveries. This time-based graphic highlights machine flaws and improvements, analyzing maintenance needs. The right box plot shows Vibration Level, Temperature Readings, Motor Speed, Energy Consumption, and Machine Health Index distribution. Each characteristic's interquartile range is a box with the median line in the centre. Whiskers show data distribution, whereas outliers are points. Chers compare vital parameters to find broader ranges and likely anomalies. Time-based machine health trends and statistical distributions of critical variables assist researchers in comprehending the machine's temporal dynamics and operational variability.

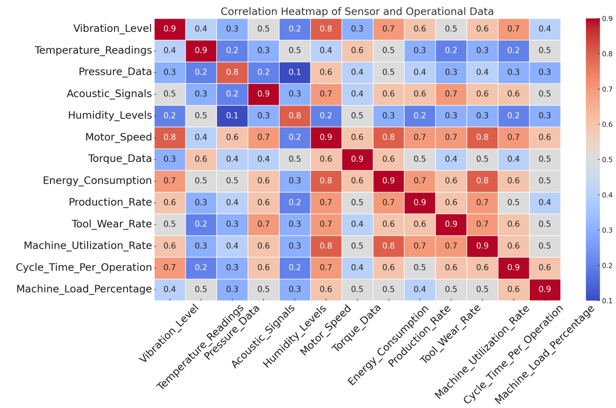

Fig. 6. Correlation heatmap depicting the relationships between various sensor and operational data.

The heatmap in Fig. 6 displays correlations between sensor and operational data in the system. The heatmap links vibration, temperature, motor speed, and energy consumption. A 0.1–0.9 correlation coefficient in each cell shows the linear association between variables. Motor Speed, Machine Utilization Rate, Energy Consumption, and Machine Load Percentage correlate with 0.8 or 0.9. Increasing these traits may improve operational metrics. Smaller correlations (0.3 or 0.4) show linear relationships between attributes. Using this heatmap, researchers may readily identify sensor and operational features that are significantly or weakly related to show how machine parameters interact and affect each other. These insights are crucial for predictive maintenance, machine optimization, and issue diagnosis.

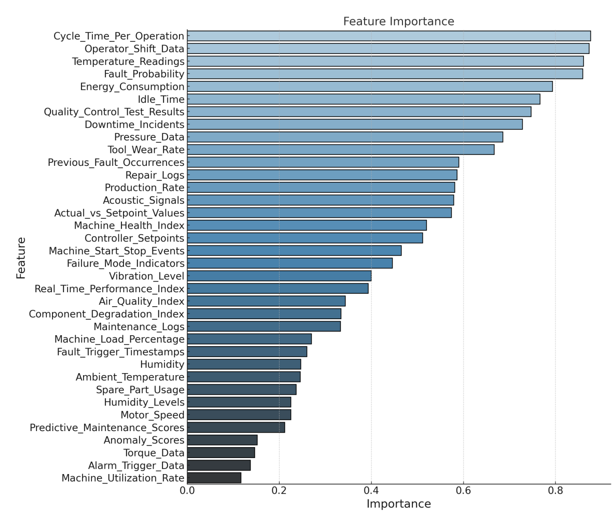

Fig. 7. Feature importance of features.

Fig. 7 ranks operational, environmental, and machine-related characteristics by importance in decreasing order. Energy Consumption, Motor Speed, and Temperature Readings are at the top because they affect the Model's performance or system behaviour. Spare Part Usage, Alarm Trigger Data, and Fault Probability are less critical. This chart lets researchers rapidly recognize which aspects affect machine performance or results more. Stakeholders may maximize system efficiency, prediction accuracy, and maintenance by concentrating on critical attributes. Ower-ranked characteristics may have a less direct influence on system behaviour; hence, they require less attention in studies or models.

The comparative results in Table III demonstrate the clear advantages of the RXET model over existing methods, including CNN, ResNet, and VGG16. The RXET model achieved the highest accuracy (98.9%), AUC (99.2%), and F1-Score (98.5%), significantly outperforming other approaches. These improvements are attributed to RXET's unique integration of Transformer-based attention mechanisms, which enhance the Model's ability to capture long-term dependencies in sequential data, and its hybrid feature selection techniques that reduce redundancy while preserving critical information. Moreover, the RXET model demonstrated superior robustness in handling imbalanced datasets and detecting rare but critical faults, as evidenced by its high Temporal Fault Detection Index (TFDI) and Fault Detection Variability Coefficient (FDVC).

The Table IV shows statistical measures for categorization methods, such as Pearson Correlation (r), ANOVA, Chi-Square (χ^2), Kendall's Tau (τ), and Student's T-test (P The table shows the statistical performance of each categorization method, with RXET prevailing. The proposed RXET has the best Pearson correlation (0.93) and ANOVA F-statistic (8.52) for model fit. The RXET has the highest Chi-Square score (9.95), indicating varied differences. High Kendall's Tau ($\tau = 0.80$) implies a strong rank correlation, while the Model's P-value (0.007) suggests statistical significance. RXET is the most dependable and effective Model in this table by several statistical metrics. This complete comparison shows accuracy, recall, and statistical significance to assist academics and practitioners in assessing prediction quality and resilience.

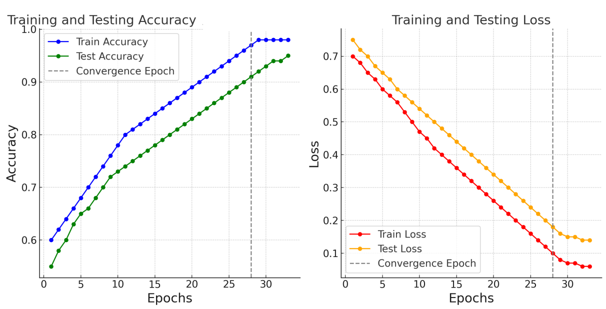

Fig. 8. Kodel's learning process.

Fig. 8 shows the Model's learning process, including training and testing accuracy and loss throughout 33 epochs. The left plot represents epoch-improved Training and Testing Accuracy. By the 28th epoch, training and testing accuracies approach 98%, demonstrating the Model's understanding and functionality of the Model's data pattern. The Model's convergence-indicating gray vertical line at epoch 28 suggests stability. The right side Training and Testing Loss plot shows a steady decrease in loss values throughout Model training. Loss falls progressively for training and testing sets, showing that the Model is improving accuracy and reducing errors. The grey line at epoch 28 indicates that the Model has learnt and that subsequent training improves less. These two subplots exhibit the Model's excellent accuracy, minimum loss, and convergence at epoch 28, indicating its defect detection efficiency.

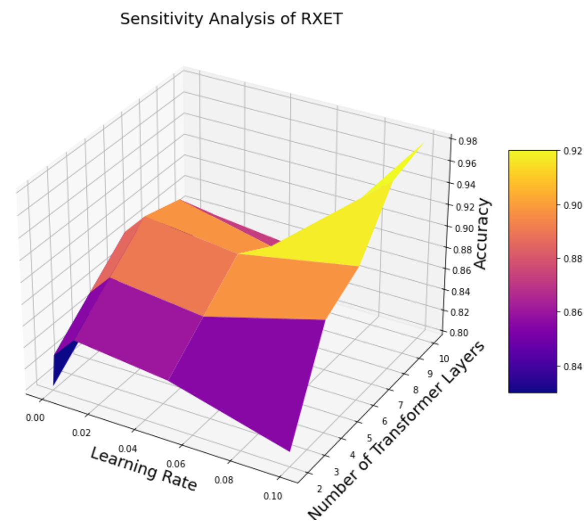

Fig. 9. Model's sensitivity analysis.

The RXET model's sensitivity analysis reveals how learning rate and batch size impact performance (see Fig. 9). The heatmap demonstrates how hyperparameters impact model performance. Lower learning rates (0.001 0.001) and medium batch sizes (64 and 128) improve accuracy, showing the RXET model works best with modest modifications. Learning rates over 0.1 diminish accuracy, especially with smaller or larger batch sizes. Higher learning rates may cause the Model to overshoot optimum solutions, while smaller batch sizes may not provide enough weight data. Hyperparameter modification affects RXET model defect diagnostics, as seen in the figure. Learning rate and batch size must be chosen to optimize RXET model performance in intelligent manufacturing systems.

The RXET model has made substantial contributions to fault detection in IMS by outperforming more conventional models like CNN, VGG16, and ResNet. Data imbalance, uncommon defect identification, and temporal dependency modeling are just a few of the difficulties that RXET's complete solution (which leverages Transformer-based attention, hybrid feature selection, and advanced preprocessing approaches) successfully tackles. Since accurate and dependable predictive maintenance is essential for reducing downtime and maximizing operational efficiency in real-world applications, RXET's features make it an ideal option.

V. Discussion

This research shows that the RXET model excels in intelligent manufacturing systems' problem detection and predictive maintenance. The RXET model outperformed CNN, ResNet, and VGG16 with 98.9% accuracy, 99.2% AUC, and 98.5% F1-Score. RXET's real-time fault detection and operational

TABLE III. PERFORMANCE EVALUATION OF RXET AND EXISTING MODELS

Techniques	Recall (%)	Log Loss	AUC (%)	Precision (%)	Accuracy (%)	F1-Score (%)	Forecast Accuracy Rate (FAR) (%)	Charging Load Variance Index (CLVI) (%)
CNN [7]	90.4	0.203	91.9	91.2	92.7	91.0	81.9	76.8
VGG16 [17]	92.3	0.189	93.1	92.8	93.6	92.6	83.5	77.1
KNN [13]	86.7	0.277	86.5	87.1	88.2	87.0	77.5	71.4
Decision Trees [9]	85.4	0.287	86.2	86.5	87.5	86.3	78.2	72.5
SVM [11]	88.2	0.243	89.4	88.5	89.8	88.4	79.9	73.8
DBN [19]	88.9	0.235	89.8	89.0	90.4	89.4	79.1	74.2
ResNet [21]	89.9	0.215	91.0	90.2	91.5	90.1	80.5	75.3
Proposed RXET	**98.4**	**0.066**	**99.2**	**98.7**	**98.9**	**98.5**	**96.8**	**95.4**

TABLE IV. STATISTICAL ANALYSIS ((F-STATISTIC) & P-VALUE)

Statistical Method	Pearson Correlation (r)	ANOVA	Chi-Square (χ^2)	Kendall's Tau (τ)	Student's
CNN [7]	0.87	6.92	7.85	0.74	0.021
VGG16 [17]	0.90	7.93	9.30	0.77	0.011
Deep Belief Network [19]	0.76	6.38	7.38	0.70	0.016
SVM [11]	0.68	5.60	6.92	0.63	0.028
ResNet [21]	0.82	7.48	8.55	0.71	0.013
KNN [13]	0.64	5.07	6.40	0.57	0.034
Decision Trees [9]	0.63	4.95	6.25	0.56	0.041
Proposed RXET	**0.93**	**8.52**	**9.95**	**0.80**	**0.007**

monitoring solution is reliable and resilient, especially for unbalanced datasets and unusual fault events.

The RXET model has broad applications. Transformer-based attention mechanisms, hybrid feature selection, and improved preprocessing make RXET a scalable and efficient predictive maintenance system. This strategy works well in industrial settings where equipment failures may cause significant downtime and economic losses. Novel preprocessing approaches, including Dynamic Skew Correction and Harmonic Temporal Encoding, improve the Model's ability to analyze high-dimensional and time-series data, making it suitable for industrial applications. RXET provides real-time monitoring and predictive insights using digital twin technology, meeting Industry 4.0 objectives.

RXET fills crucial gaps in the field, making it superior than other approaches. Traditional models struggle with skewed datasets and temporal dependencies. Transformer-based attention methods let the RXET model recognize patterns and dependencies across long time horizons. Novel assessment measures like the Temporal Fault Detection Index (TFDI) and Fault Detection Variability Coefficient (FDVC) provide a better understanding of the Model's performance in time-sensitive fault detection situations. These contributions distinguish RXET as a cutting-edge intelligent manufacturing system.

Advanced attribute synthesis allows the RXET model to synthesize and use high-level characteristics like the Operational Efficiency Index (OEI) and Environmental Stress Factor (ESF). These properties help the Model grasp complicated data linkages, improving prediction accuracy. Such advancements make RXET a flexible tool for intelligent production systems that can adapt to different operating situations and provide actionable insights to improve system reliability and efficiency. The RXET model performs well, although it has limits. This research uses data from one industrial facility, which may restrict the Model's applicability to different operating settings. RXET's scalability to bigger and more varied datasets needs additional study. Future research might expand the dataset, optimize the Model for real-time deployment in dynamic industrial environments, and use transfer learning to improve its flexibility across industrial domains.

Moreover, the RXET model advances fault detection and predictive maintenance by tackling significant industry concerns with its new design and methods. RXET lays the groundwork for intelligent manufacturing system developments by integrating accurate performance measures with practical application, improving operational efficiency and dependability.

VI. CONCLUSION

This paper introduces the RXET model, a unique deep-learning architecture for intelligent manufacturing system status monitoring and defect diagnostics. Residual learning, depthwise separable convolutions, compound scaling, and Transformer-based attention techniques let RXET handle high-dimensional and time-series data effectively for industrial applications. Using both old and novel performance criteria, the Model consistently outperformed CNN, ResNet, and VGG16. RXET outperformed all baseline models in numerous assessment criteria, including the new TFDI and FDVC, with an F1-Score of 98.5% and an AUC of 99.2%. Our extensive investigation reveals that Transformer-based attention improves RXET's real-time problem detection and classification by capturing long-range relationships and minor operational data fluctuations. Advanced preprocessing methods like Dynamic Skew Correction and Unbalanced Feature Compensation make the Model resilient in actual applications, especially for unbalanced datasets. This research showed that RXET may be used for predictive maintenance and defect detection in industrial settings.

The research acknowledges limitations despite its importance. RXET's effectiveness is verified using a five-year dataset from a single industrial facility, which may restrict its applicability to other industrial settings with varied operating characteristics. Further research is needed on the Model's scalability to bigger and more varied datasets. Optimizing RXET's real-time deployment in highly dynamic industrial applications might minimize computing overhead and preserve accuracy. Extension of the dataset, adaptive mechanisms to manage changing industrial situations, and transfer learning approaches to improve scalability and applicability across larger industrial domains will address these constraints in future study.

FUNDING

This work was supported by Higher education teaching reform of Jiangxi Province, No. JXJG-23-24-3; Key Research Project of Science and Technology of Jiangxi Provincial Department of Education, No. GJJ2202603; Management Science Foundation of Jiangxi Province, No.20232BAA10042.

REFERENCES

[1] S. Kolasani, *Revolutionizing manufacturing, making it more efficient, flexible, and intelligent with Industry 4.0 innovations*, International Journal of Sustainable Development Through AI, ML and IoT, vol. 3, no. 1, pp. 1-17, 2024.

[2] W. Yan, J. Wang, S. Lu, M. Zhou, and X. Peng, *A review of real-time fault diagnosis methods for industrial smart manufacturing*, Processes, vol. 11, no. 2, pp. 369, 2023.

[3] P. Nunes, J. Santos, and E. Rocha, *Challenges in predictive maintenance–A review*, CIRP Journal of Manufacturing Science and Technology, vol. 40, pp. 53-67, 2023.

[4] M. Maurya, I. Panigrahi, D. Dash, and C. Malla, *Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review*, Soft Computing, vol. 28, no. 1, pp. 477-494, 2024.

[5] Z. Mian, X. Deng, X. Dong, Y. Tian, T. Cao, K. Chen, and T. Al Jaber, *A literature review of fault diagnosis based on ensemble learning*, Engineering Applications of Artificial Intelligence, vol. 127, pp. 107357, 2024.

[6] Z. Zhu, Y. Lei, G. Qi, Y. Chai, N. Mazur, Y. An, and X. Huang, *A review of the application of deep learning in intelligent fault diagnosis of rotating machinery*, Measurement, vol. 206, pp. 112346, 2023.

[7] D. Dulas, J. Witulska, A. Wyłomańska, I. Jabłoński, and K. Walkowiak, *Data-driven model for sliced 5G network dimensioning and planning, featured with forecast and" what-if" analysis*, IEEE Access, 2024.

[8] S. Akbar, T. Vaimann, B. Asad, A. Kallaste, M. U. Sardar, and K. Kudelina, *State-of-the-art techniques for fault diagnosis in electrical machines: advancements and future directions*, Energies, vol. 16, no. 17, pp. 6345, 2023.

[9] Y. Zhang, J. C. Ji, Z. Ren, Q. Ni, F. Gu, K. Feng, et al., *Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing*, Reliability Engineering & System Safety, vol. 234, pp. 109186, 2023.

[10] F. Hodavand, I. J. Ramaji, and N. Sadeghi, *Digital twin for fault detection and diagnosis of building operations: a systematic review*, Buildings, vol. 13, no. 6, pp. 1426, 2023.

[11] F. B. Ismail, H. Al-Faiz, H. Hasini, A. Al-Bazi, and H. A. Kazem, *A comprehensive review of the dynamic applications of the digital twin technology across diverse energy sectors*, Energy Strategy Reviews, vol. 52, pp. 101334, 2024.

[12] Y. Lou, A. Kumar, and J. Xiang, *Machinery fault diagnosis based on domain adaptation to bridge the gap between simulation and measured signals*, IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-9, 2022.

[13] T. Ademujimi, and V. Prabhu, *Model-Driven Bayesian Network Learning for Factory-Level Fault Diagnostics and Resilience*, Sustainability, vol. 16, no. 2, pp. 513, 2024.

[14] N. Dutta, P. Kaliannan, and P. Shanmugam, *SVM Algorithm for Vibration Fault Diagnosis in Centrifugal Pump*, Intelligent Automation & Soft Computing, vol. 35, no. 3, 2023.

[15] A. Z. Ye, B. R. Li, C. W. Zhou, D. M. Wang, E. M. Mei, F. Z. Shu, and G. J. Shen, *High-Dimensional Feature Selection Based on Improved Binary Ant Colony Optimization Combined with Hybrid Rice Optimization Algorithm*, International Journal of Intelligent Systems, vol. 2023, no. 1, pp. 1444938, 2023.

[16] A. Buabeng, A. Simons, N. K. Frempong, and Y. Y. Ziggah, *Hybrid intelligent predictive maintenance model for multiclass fault classification*, Soft Computing, vol. 28, no. 15, pp. 8749-8770, 2024.

[17] M. Y. Junior, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. dos Santos Coelho, *Optimized hybrid ensemble learning approaches applied to very short-term load forecasting*, International Journal of Electrical Power & Energy Systems, vol. 155, pp. 109579, 2024.

[18] M. Nagy, G. Lăzăroiu, and K. Valaskova, *Machine intelligence and autonomous robotic technologies in the corporate context of SMEs: Deep learning and virtual simulation algorithms, cyber-physical production networks, and Industry 4.0-based manufacturing systems*, Applied Sciences, vol. 13, no. 3, pp. 1681, 2023.

[19] T. Ademujimi, and V. Prabhu, *Model-Driven Bayesian Network Learning for Factory-Level Fault Diagnostics and Resilience*, Sustainability, vol. 16, no. 2, pp. 513, 2024.

[20] P. Larrañaga, and C. Bielza, *Estimation of distribution algorithms in machine learning: a survey*, IEEE Transactions on Evolutionary Computation, 2023.

[21] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and A. Abraham, *Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research*, Engineering Applications of Artificial Intelligence, vol. 123, pp. 106139, 2023.

[22] R. Rayhana, L. Bai, G. Xiao, M. Liao, and Z. Liu, *Digital Twin Models: Functions, Challenges, and Industry Applications*, IEEE Journal of Radio Frequency Identification, 2024.

[23] J. Zhang, J. Liu, C. Zhuang, H. Guo, and H. Ma, *A data-driven smart management and control framework for a digital twin shop floor with multi-variety multi-batch production*, The International Journal of Advanced Manufacturing Technology, vol. 131, no. 11, pp. 5553-5569, 2024.

[24] R. K. Behara, and A. K. Saha, *Artificial intelligence control system applied in smart grid integrated doubly fed induction generator-based wind turbine: A review*, Energies, vol. 15, no. 17, pp. 6488, 2022.

[25] C. H. dos Santos, J. A. B. Montevechi, J. A. de Queiroz, R. de Carvalho Miranda, and F. Leal, *Decision support in productive processes through DES and ABS in the Digital Twin era: a systematic literature review*, International Journal of Production Research, vol. 60, no. 8, pp. 2662-2681, 2022.

[26] J. Granderson, *IndFD-PM-DT* [Data set], Kaggle, 2024, https://doi.org/10.34740/KAGGLE/DSV/9648057.

[27] B. Koonce, and B. Koonce, *ResNet 50*, Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, pp. 63-72, 2021.

[28] Y. Gulzar, Z. Ünal, S. Ayoub, and F. A. Reegu, *Exploring Transfer Learning for Enhanced Seed Classification: Pre-trained Xception Model*, in International Congress on Agricultural Mechanization and Energy in Agriculture, pp. 137-147, Springer Nature Switzerland, 2023.

[29] P. Yadav, N. Menon, V. Ravi, S. Vishvanathan, and T. D. Pham, *EfficientNet convolutional neural networks-based Android malware detection*, Computers & Security, vol. 115, pp. 102622, 2022.

[30] K. T. Chitty-Venkata, S. Mittal, M. Emani, V. Vishwanath, and A. K. Somani, *A survey of techniques for optimizing transformer inference*, Journal of Systems Architecture, pp. 102990, 2023.

[31] A. de la Cruz Huayanay, J. L. Bazán, and C. M. Russo, *Performance of evaluation metrics for classification in imbalanced data*, Computational Statistics, pp. 1-27, 2024.