e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 13, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

# Fault Diagnosis in Power Plant Condensers Using Optimized Machine Learning Algorithm

**Ms. C. Kotteeswari[1], R. Tamil Aruvi[2], M. Santhosh[3], M. Muhammed Mukrim[4]**

Assistant Professor, Dept. of C.S.E, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[1]

Student, Dept. of C.S.E, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[2, 3, 4]

**ABSTRACT:** Condensers are vital components in power plants, playing a crucial role in overall efficiency and performance. Effective fault diagnosis in power plant condensers is essential to maintain operational efficiency and prevent costly downtime. Traditional methods often struggle to provide accurate diagnoses under complex conditions. Our novel methodology integrates feature engineering, correlation analysis, and XGBoost classification to enhance diagnostic precision. By extracting and transforming relevant features from fault data, we capture intricate patterns indicative of condenser faults. Subsequent correlation analysis prioritizes key features with strong relationships to fault occurrences, optimizing diagnostic performance. Leveraging the capabilities of XGBoost, we achieve precise fault classification, enabling proactive maintenance and timely interventions. Experimental results demonstrate an outstanding accuracy rate of approximately 99.69%. Comparative evaluations against SVM, KNN, and Naive Bayes highlight XGBoost classification enables accurate and efficient prediction of condenser faults, ultimately improving the reliability, safety, and performance of power generation systems.

**KEYWORDS:** Condenser fault diagnosis, Feature Engineering, Correlation Analysis, XGBoost Classification.

## I. INTRODUCTION

Condenser fault prediction poses a significant challenge in ensuring the seamless operation of power plants. Accurate prediction of faults enables efficient resource allocation and helps prevent costly downtimes. Traditional fault prediction methods often rely on manual analysis and expert knowledge, leading to subjective and time-consuming processes. Leveraging machine learning algorithms has emerged as a promising approach for fault prediction due to their ability to learn patterns and make accurate predictions. However, accurately predicting condenser faults is crucial for preventing operational disruptions. To achieve this, it's imperative to effectively leverage the dataset encompassing various fault types, including power failure, heat dissipation issues, tool wear failure, and overstrain failure. While machine learning-based fault prediction is gaining traction, there is still a need for systematic comparisons of different methods to enhance prediction accuracy. Machine learning algorithms have demonstrated their capability to process complex data and handle high-dimensional datasets efficiently, making them suitable for condenser fault prediction. This paper proposes a novel approach integrates feature engineering, correlation analysis, and XGBoost classification to improve diagnostic accuracy. Through feature extraction and transformation from fault data, we capture nuanced patterns signaling condenser faults. Correlation analysis then prioritizes crucial features strongly linked to fault occurrences, refining diagnostic effectiveness. Harnessing the power of XGBoost, we achieve precise fault classification. The integration of these methods aims to enhance prediction accuracy and facilitate preemptive maintenance interventions, ultimately optimizing the reliability and performance of condenser.

## II. RELATED WORKS

In the domain of fault diagnosis within thermal power plants, researchers have dedicated considerable efforts to developing robust methodologies capable of identifying and addressing potential issues promptly. Among these endeavors, Ali and Mahdi [1] proposed a novel approach centered around independent component analysis, specifically tailored for diagnosing faults within condenser systems—a critical component within thermal power plants. This methodology aimed to enhance the efficiency and accuracy of fault detection, thereby minimizing the risk of operational disruptions and associated losses. Building upon this foundation, Ma et al. [2] introduced a comprehensive methodology that ingeniously combined artificial neural networks with optimal zoom search techniques. This innovative fusion aimed to equip the diagnostic system with the capability to discern and classify various degrees of faults present in power plant thermal systems, even across fluctuating load conditions. By integrating these advanced computational techniques, the diagnostic process could adapt dynamically to changing operational parameters, ensuring robust performance in real-world scenarios. Further expanding the repertoire of fault diagnosis methodologies, Wang et al. [3] proposed a pioneering technique leveraging nearest prototype classifiers for fault detection within power plants. This approach represented a departure from conventional methods, offering a fresh perspective on fault diagnosis by harnessing the inherent strengths of prototype-based classification. By identifying and categorizing faults based on their proximity to prototype representations, this methodology provided a more nuanced and precise diagnostic framework. In a similar vein, Ge et al. [4] delved into the application of a kernel-driven semisupervised Fisher discriminant analysis model—a sophisticated analytical tool designed to classify nonlinear faults encountered in industrial processes. By harnessing the power of kernel methods and semisupervised learning, this approach offered a robust solution capable of effectively addressing the inherent complexities associated with fault diagnosis in dynamic and nonlinear systems. Complementing these efforts, Kang et al. [5] adopted a fuzzy inference system to investigate performance degradation within feedwater heaters—a critical component of power generation equipment. This utilization of fuzzy logic allowed for a more flexible and adaptive approach to fault diagnosis, enabling the system to accommodate uncertainties and variations inherent in real- world operating conditions. Expanding the horizon of fault diagnosis methodologies, Wang et al. [6] introduced a hierarchical depth domain adaptive method designed to facilitate the seamless transfer of classifiers between labeled and unlabeled data under varying loading conditions within power plant thermal systems. This hierarchical approach aimed to enhance the adaptability and robustness of the diagnostic system, ensuring reliable performance across diverse operational scenarios. Machine learning techniques, ranging from backpropagation neural networks to support vector machines and deep learning, have emerged as powerful tools for fault diagnosis. However, the effectiveness of these approaches hinges on the availability of high-quality training data, which can often be challenging to obtain in industrial settings characterized by complex and dynamic operational environments. In response to the challenges posed by incomplete sensor data, Hu et al. proposed a ground breaking solution—a generative adversarial network architecture with trinetworks. This innovative approach aimed to address the complexities of leak detection by leveraging the power of adversarial training and multi-network architectures, thereby enhancing the accuracy and reliability of fault diagnosis processes. Furthermore, Ma et al. [13] introduced a novel graph-theory-based network partitioning algorithm tailored for decentralized fault detection, offering faster response speeds and improved scalability—a crucial advancement in fault diagnosis methodologies, particularly for large-scale industrial systems. The advent of the deep forest (DF) model, pioneered by Zhou and Feng [14], has sparked a surge of interest in fault diagnosis research, leading to innovative approaches by Hua et al. [18], Liu et al. [19], and Ding et al. [20]. Despite these advancements, there remains a notable gap in the literature concerning the application of DF models for condenser fault diagnosis—a critical area warranting further investigation and exploration. Principal component analysis (PCA), a widely used technique for dimensionality reduction, has found extensive application in fault diagnosis. Various modifications and enhancements, such as kernel PCA, nonnegative variable sparse PCA, and sparse PCA, have been proposed to augment its effectiveness in extracting essential features and patterns from complex datasets. Prior to deploying DF for fault diagnosis, leveraging PCA for feature selection can help mitigate the impact of irrelevant features and correlations among data points. By selecting and prioritizing significant features, PCA can enhance the diagnostic process's accuracy and efficiency, laying a solid foundation for effective fault detection and mitigation strategies within thermal power plants.

## III. PROPOSED SYSTEM

In our proposed system, the XGBoost classification algorithm is a pivotal component for predicting fault types in condensers with remarkable accuracy. Renowned for its efficiency and effectiveness in handling intricate datasets, XGBoost stands out as the optimal choice for fault classification tasks due to its ability to discern subtle patterns and relationships within the data. Additionally, our system utilizes feature engineering and correlation analysis techniques to perform feature selection. By extracting relevant data features from historical fault records, we can identify key indicators of condenser faults and prioritize features strongly correlated with fault occurrences. This approach ensures that our classification algorithm is equipped with the most pertinent information, enhancing its predictive accuracy. Ultimately, selecting the appropriate classification algorithm is crucial for achieving accurate fault prediction, and

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

XGBoost emerges as the preferred choice given its proven track record in handling complex datasets and delivering precise classifications.

### A. *Data Collection*

Data collection entails capturing records of past occurrences, facilitating subsequent data analysis to uncover recurrent patterns. This process enables us to extract valuable insights from historical data, aiding in the identification of recurring trends and patterns essential for predictive analysis and decision-making. Fig 1. shows the overview of the fault dataset.

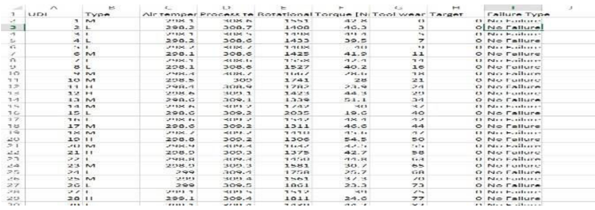


**Figure 1 Overview of the Fault dataset**

From these recurring patterns, predictive models are constructed using machine learning algorithms, which seek ou**t**

### B. Data *Preprocessing*

Data preprocessing is a vital step in data analysis where raw data is cleaned, transformed, and organized to enhance its quality and usability. One crucial aspect is outlier detection and removal, as outliers can skew analytical results and model performance. By efficiently identifying and eliminating outliers, preprocessing ensures that the data used for analysis is more representative and reliable, contributing to improved accuracy and efficiency in subsequent task**s.**

### C. *Feature Engineering*

Feature engineering is pivotal in our project, enabling us to extract valuable insights from the dataset concerning condenser faults. One of the key features engineered is the Temperature Differential Feature, which calculates the variance between 'Process temperature' and 'Air temperature,' resulting in the creation of the ‘Temperature Differential’ feature. This feature provides potential insights into temperature-related dynamics within the system. Additionally, we introduce the Power Feature, which is computed by combining 'Torque' and 'Rotational speed' to capture the system's power output or energy consumption. This feature offers valuable information regarding the operational characteristics of the condenser. Moreover, we derive the Cumulative Effect Feature by amalgamating 'Torque' and 'Tool wear,' representing their cumulative impact over time. These engineered features aim to enhance the predictive capabilities of our model by capturing relevant patterns and relationships within the data, ultimately improving the accuracy of condenser fault prediction.

### D. *Feature Selection*

Feature selection is a critical process in machine learning aimed at enhancing model performance and efficiency by identifying and selecting the most relevant features from a dataset. This process helps reduce data dimensionality by eliminating irrelevant or redundant features, improving the model's predictive power, interpretability, and computational efficiency. By selecting informative features, feature selection mitigates the risk of overfitting, enhances model generalization to unseen data, and accelerates both training and inference processes. In our feature selection module, we employed the Spearman correlation coefficient—a robust measure of monotonic relationships between numerical features and the target variable. This technique enabled us to assess the strength and direction of association between each feature and the target, guiding the selection of relevant predictors for our model. Fig 2. Shows the Correlation Analysis. Fig 3. Show the Selected Features.

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

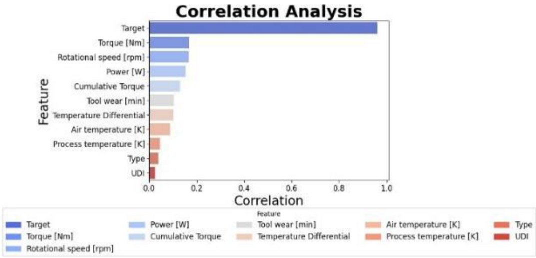



Figure 2: Correlation Analysis

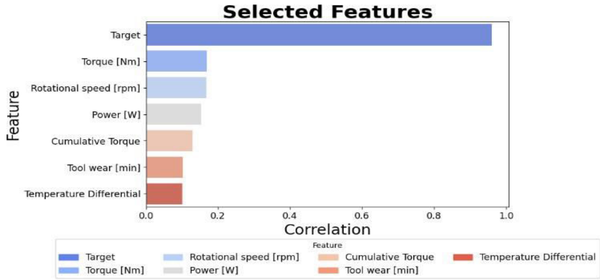



Figure 3: Selected Features

*A. Predictive Modeling for Fault Diagnosis*

Classification is a fundamental task in machine learning, where the goal is to assign predefined categories or labels to input data based on their features. Here, we explore various classification algorithms to predict fault types in condensers. We begin by implementing Support Vector Machine (SVM), a versatile algorithm that separates data points by finding the optimal hyperplane to maximize the margin between different classes. Next, we employ K-Nearest Neighbors (KNN), which classifies data points based on the majority class among their nearest neighbors in feature space. Additionally, we utilize Naive Bayes, a probabilistic classifier that applies Bayes' theorem with strong independence assumptions between features. Finally, we leverage XGBoost (Extreme Gradient Boosting), an ensemble learning technique known for its efficiency and high predictive accuracy.

To evaluate the performance of these algorithms, we consider metrics such as accuracy, precision, recall, and F1- score. Accuracy measures the overall correctness of the classification, while precision and recall provide insights into the algorithm's ability to make correct positive predictions and capture all positive instances, respectively. The F1-score provides a unified measure of a model's performance by incorporating both precision and recall, ensuring a well-rounded evaluation of its effectiveness. Fig 4. Shows the Classification Metrics.

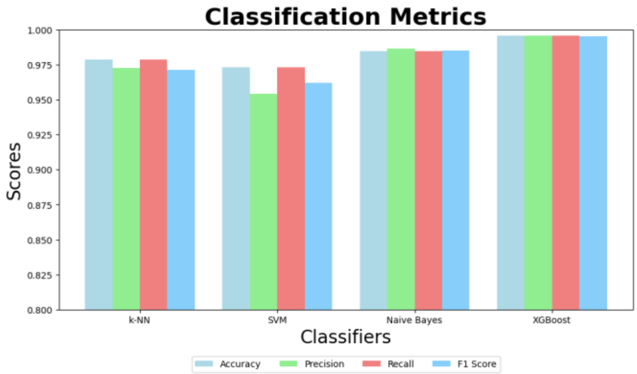



Figure 4: Classification Metrics

After comprehensive evaluations, we conclude that XGBoost outperforms SVM, KNN, and Naive Bayes in terms of accuracy, precision, recall, and F1-score. Its ensemble learning approach, coupled with efficient handling of complex datasets, makes XGBoost the most suitable algorithm for predicting fault types in condensers with high accuracy and reliability.

## IV. CONCLUSION AND FUTURE WORK

The purpose of this research endeavor is to advance the diagnosis of condenser faults through the application of sophisticated methodologies encompassing feature engineering, correlation analysis, and classification algorithms. Our modified approach aims to elevate fault diagnosis accuracy, thereby facilitating the implementation of proactive maintenance strategies and bolstering the overall reliability of power generation systems. Our proposed methodology demonstrates promising outcomes in condenser fault diagnosis. However, there remains ample opportunity for refinement and further exploration within this domain. Future endeavors could concentrate on refining feature engineering techniques to extract more subtle fault indicators, optimizing correlation analysis methodologies to enhance the precision of feature selection, and exploring the integration of additional classification algorithms to augment predictive accuracy.

Moreover, the methodologies devised in this study hold potential for extension beyond condenser fault diagnosis, extending to other industrial domains such as pipeline system fault detection and bearing fault diagnosis. By perpetuating the refinement and expansion of these techniques, we can contribute significantly to advancements in fault diagnosis across a spectrum of industrial applications, ultimately enhancing operational efficiency and bolstering system reliability.

## REFERENCES

[1] The purpose of this research endeavor is to advance the diagnosis "Independent component analysis approach for fault diagnosis of condenser system in thermal power plant," published in the Journal of Central South University in March 2014 by A. Ali and D. Mahdi

[2] An intelligent power plant defect diagnosis for varied degree of severity and loading situations was published in the June 2010 issue of IEEE Transaction on Energy Conversation by L. Ma, Y. Ma, and K. Y. Lee.

[3] "An optimized nearest prototype classifier for power plant fault diagnosis using hybrid particle swarm optimization algorithm," published in the International Journal of Electrical Power & Energy Systems in June 2014 by X. Wang, L. Ma, and T. Wang.

[4] "Diagnosis of feedwater heater performance degradation using fuzzy inference system," Expert Systems with Applications, March 2017, Y. K. Kang,
H. Kim, G. Heo, and S. Y. Song.

[5] "A hierarchical deep domain adaptation approach for fault diagnosis of power plant thermal system," published in September 2019 in IEEE Transactions on Industrial Informatics, by X. Wang, H. He, and L. Li.

[6] In April 2022 publication in Electrical Engineering,
P. Zhang, Z. Cui, Y. Wang, and S. Ding "Introduce a Fault Diagnosis Approach for Electrical Machine Drive Systems".

[7] The July 2004 paper "Fault diagnosis based on Fisher discriminant analysis and support vector machines" by L. H. Chiang, M. E. Kotanchek, and A. K. Kordon Compute Chemistry and Engineering.

[8] T. G. Dietterich, "An experimental comparison of three decision tree ensemble construction methods: boosting, randomization, and bagging," 139–157 in Mach. Learn., vol. 40, no. 2, 2000.

[9] G. E. Hinton and R. R. Salakhutdinov, "Using neural networks to reduce the dimensionality of data," Science, March 2006, vol. 313, no. 5786, pp. 504-507.

[10] The article "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review" was published in IEEE 2019.

[11] A tnGAN-based leak detection technique for pipeline networks considering partial sensor data, X. Hu, H. Zhang, D. Ma, and R. Wang, IEEE Trans. Instrumentation Measurements, vol. 70, pp. 1–10, 2021.

[12] [Z. Zhou and F. Ji, "Deep forest: Towards an alternative to deep neural networks," Nat. Sci. Rev., vol. 6, no. 1, October 2019, pp. 74-86.

[13] Ma et al. presented a hierarchical event detection method for power systems using spectral theory in IEEE Transactions on Systems, Man, and Cybernology (April 2019).

[14] Pan et al. introduced a novel deep learning network for fault detection in IEEE Transactions on Industrial Informatics (September 2019).

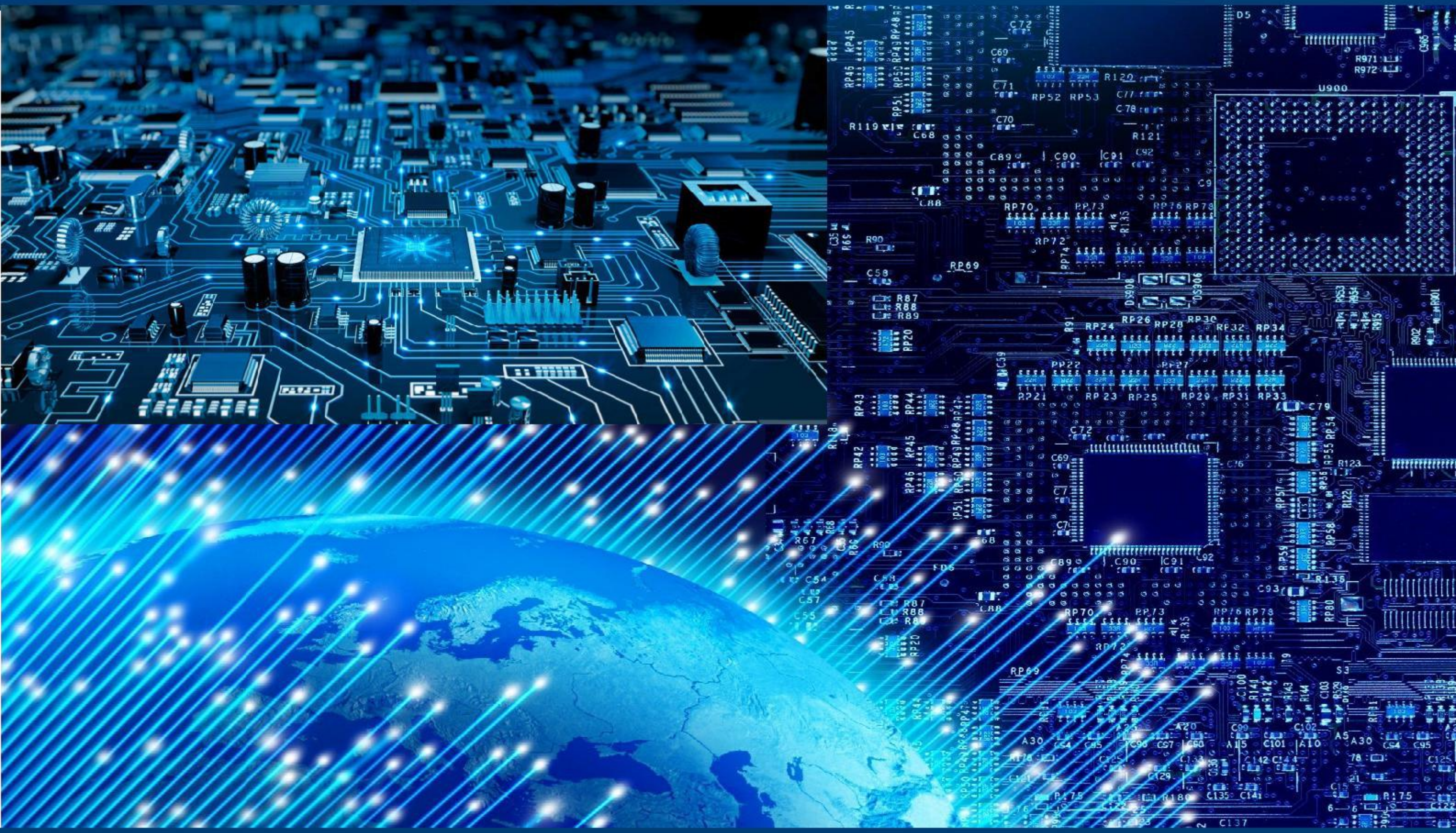
[15] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Deep learning model with limited data sample for gearbox fault diagnosis," December 2020; IEEE Trans. Ind. Informat., vol. 16, no. 10, pp. 6263–6271

[16] "Deep forest based intelligent fault diagnosis of hydraulic turbine," J. Mech. Sci. Technol., vol. 33, no. 5,
pp. 2049–2058, May 2019, X. Liu, Y. Tian, X. Lei, M. Liu, X. Wen, H. Huang, and H. Wang.

[17] "Simultaneous Fault Detection and Diagnosis Using Adaptive Principal Component Analysis and Multivariate Contribution Analysis," by L. M. Elshenawy, T. A. Mahmoud, and C. Chakour Innov. Eng. Chem. Res., Nov. 2020; vol. 59, no. 47; pp. 20798–20815.

[18] Y. Du and D. Du, "Using empirical mode decomposition based principal component analysis for fault detection and diagnosis," July 2018, Comput. Chem. Eng., vol. 115, pp. 1–21.

[19] "Fault identification for process monitoring utilizing kernel principal component analysis," J. H. Cho, J. M. Lee, S. W. Choi, D. Lee, and I. B. Lee, Chem. Eng. Sci., vol. 60, no. 1, pp. 279–288, 2005.
[20] "Nonnegative sparse PCA," by R. Zass and A. Shashua, in Proc. Adv. Neural Inf. Process. Syst., Jan. 2006, pp. 1561–1568.
[21] "A Novel Basics for Principal Component Analysis," by F. Chen and K. Rohe arXiv:2007.00596 (2020).
[22] S. Zou, "Using principal component analysis and RBF neural network to diagnose condenser faults," M.S. thesis, Huazhong University of Science and Technology, School of Energy Power Engineering, Hubei, China, 2014.
[23] J. E. Jackson, A Principal Components User's Guide. Wiley, New York, NY, USA, 2005.
[24] "Real-time fault detection and diagnosis using sparse principal component analysis," by S. Gajjar, M. Kulahci, and A. Palazoglu J. Process Control, July 2018, vol. 67, pp. 112-128.
[25] "Spectrum of variable-random trees," F. T. Liu, M. T. Kai, Y. Yu, and Z. H. Zhou, J. Artif. Intell. Res., vol. 32, no. 1, pp. 355–384, May 2008.
[26] Random forests, L. Breiman, Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.





INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com