Application of Artificial Intelligence techniques to the diagnosis and maintenance of power plants

Fernando del Pino Osborne, ICAI

Abstract—**The maintenance of assets is of critical importance for the economic viability of industrial enterprises, particularly in the context of combined-cycle power plants. This Master's Thesis, entitled "Application of AI Techniques to the Diagnosis and Maintenance of Power Plants," investigates the application of machine learning and artificial intelligence for diagnosing and maintaining a feedwater pump in a combined cycle steam turbine. The study encompasses the processes of characterizing and modeling the pump's typical operational behavior, the early detection of irregularities, and the identification of potential failure modes. A data set spanning three years was employed in the research, which utilized Python and PyTorch Lightning for model development. The study highlights the shift from a reactive to a proactive maintenance strategy through the use of AI, with the objective of reducing downtime and enhancing operational efficiency. The research addresses several key challenges, including data noise, lack of generalization, and effective data collection. The findings illustrate the effectiveness of AI in predictive maintenance, offering a framework that can be applied to a range of industrial assets.**

Index Terms—**Feedwater pump, Predictive maintenance, AI, ML, Anomaly detection, Failure modes, Neural networks**

I. Introduction

In industrial enterprises, the maintenance of assets is a fundamental and indispensable undertaking. Without it, the economic viability of a production facility would be severely compromised. A well-designed maintenance plan serves to extend the lifespan of the machinery in use, while simultaneously preventing unplanned production disruptions. This ultimately results in a reduction of overall operational costs. In this context, maintenance in combined cycle power plants is of paramount importance, as they represent critical infrastructures whose proper functioning is vital for the continuous large-scale production of electrical energy.

This Master's Thesis, entitled "Application of AI Techniques for Diagnosis and Maintenance of Power Plants," is a study of the application of machine learning and artificial intelligence techniques to the diagnosis and maintenance of valuable assets in the energy sector, specifically a feedwater pump of a combined cycle steam turbine. The project's scope encompasses the characterization and modeling of the pump's normal operational behavior, as well as the preventive detection of anomalies and potential failure modes. The ultimate objective is to optimize the maintenance strategy. This feedwater pump is a real asset in an existing combined cycle plant. It is worth noting that this thesis builds on the work previously conducted by doctoral candidate Francisco Javier Bellido López.

Inadequate maintenance can result in unanticipated failures, costly disruptions, and potential safety hazards. It is therefore imperative that maintenance strategies are adopted and enhanced using the most advanced techniques available. An unanticipated failure of a plant can result in a prolonged interruption of supply, which may entail considerable monetary and reputational losses. Although this study focuses on a single component of the plant, the methodology employed is applicable to any other asset for which normal operating data is available.

The application of artificial intelligence, machine learning, and predictive analysis techniques has the potential to transform the traditional maintenance paradigm from a reactive to a proactive and prescriptive approach. The incorporation of these technologies facilitates the achievement of more rapid and precise diagnostics, the detection of anomalies prior to their transformation into failures, and the optimization of maintenance cycles. This results in a notable reduction in downtime and an enhanced operational efficiency of the machines.

A. Main objectives of the project

First, an in-depth examination of potential failure modes of the feedwater pump in a combined cycle power plant will be conducted through a comprehensive analysis of the pump to identify and categorize distinct failure modes.

Secondly, the normal operating condition (NOC) models of the feedwater pumps are to be developed using real data. This encompasses the analysis, processing, and modeling of operational variables, including pressures, temperatures, vibrations, etc. The models will be instrumental in facilitating comparisons and the identification of any deviations that may indicate the presence of anomalies or failures.

Thirdly, anomalies in the historical data series shall be identified through the utilization of the aforementioned normal operation models. The final objective is to employ these models for the continuous monitoring of the status of the feedwater pumps. Anomaly detection techniques will be employed to identify deviations in operational data that may indicate the imminent onset of a failure. This analysis will facilitate the early diagnosis of any underlying issues.

B. Resources Employed

The dataset was constructed through the examination of operational data from 23 variables that are associated with a feedwater pump in a combined cycle steam turbine. The data set encompasses a three-year period, from 2020 to 2022, and provides a comprehensive array of information for in-depth analysis and modeling.

In regard to software development tools, Python was selected as the programming language due to its versatility and comprehensive support for data analysis and machine learning/AI model development. PyTorch and PyTorch Lightning were selected as the machine learning and AI libraries, due to their capacity to facilitate the implementation of intricate models necessitating advanced learning techniques, such as neural networks. PyTorch Lightning is designed to facilitate the coding process, thereby enhancing clarity and efficiency in the management of complex models.

The hardware specifications of the computer used during all training and data analysis are:

- CPU: AMD Ryzen 7 5800H with Radeon Graphics, 3201 MHz, 8 cores, 16 logical processors.
- RAM: 16 GB at 3200MHz, sufficient for handling large datasets and intensive computation processes.
- GPU: NVIDIA GeForce RTX 3060 Laptop, with 6 GB dedicated memory plus 7.9 GB shared memory, essential for efficient deep learning model training.

II. State of the Art

A. Maintenance Strategies

The evolution of industrial asset maintenance strategies has been a significant and ongoing process throughout history, driven by technological advances. In recent years, advances in statistics and machine learning techniques have precipitated profound changes in the field. Historically, the approach to maintenance has evolved from a focus on reactive strategies, such as corrective maintenance, to a greater emphasis on proactive and data-driven techniques, including predictive and prescriptive maintenance. This shift is a response to the necessity of minimizing unplanned downtime and maximizing operational efficiency.

In the field of predictive maintenance, there is a growing tendency towards the integration of machine learning and artificial intelligence with the objective of enhancing the predictive capabilities of these models [1]. Furthermore, there is a growing inclination to integrate these data-driven models with expert knowledge, with the aim of enhancing the predictive capabilities of such models. The following section outlines the key strategies and their respective advantages and disadvantages, as outlined in [1].

1) Corrective maintenance: it is a strategy that involves repairing or replacing parts only when they fail and the equipment in question is unable to operate without intervention. The primary advantage of this strategy is its simplicity. However, this approach carries the disadvantage of potentially high costs due to the necessity of unplanned repairs or replacements.

2) Preventive maintenance: it involves the periodic inspection of equipment in accordance with a pre-established schedule or work cycles. This strategy is designed to reduce the likelihood of unexpected failures, which represents its primary advantage. One disadvantage of this approach is that it may result in the unnecessary replacement of parts that remain functional, which could lead to inefficiencies.

3) Condition-based maintenance: it employs the use of sensors to facilitate the continuous monitoring of equipment status, thereby enabling the performance of maintenance operations only when the indicators indicate the onset of failure. This approach permits more efficacious interventions based on empirical data, thereby reducing superfluous maintenance. Nevertheless, it requires an initial investment in sensors and is contingent upon the reliability and accuracy of the data obtained.

4) Predictive maintenance: it is a methodology that employs sensors and data analysis to anticipate failures and the remaining useful life of equipment. This enables the dynamic scheduling of maintenance tasks. The principal advantages of this approach are the prevention of failures and the planning of maintenance in an efficient manner. However, this strategy is more complex than previous ones and is contingent upon the quality of the data obtained.

5) Prescriptive maintenance: in addition to forecasting failures, it also proffers corrective measures to enhance the asset's lifespan, aligning with the predicted outcomes. It offers the greatest possible operational efficiency and the most extensive extension of the asset's useful life. A prerequisite for this approach is the existence of a well-developed implementation of condition-based and predictive maintenance strategies.

B. Common algorithms used in fault diagnosis analysis

Having demonstrated the value of artificial intelligence models in predictive maintenance of machines, it is now necessary to identify the algorithms that are most commonly employed. This paper presents a comparative analysis of various AI and ML techniques applied to fault diagnosis in rotating machines, with a particular focus on a combined cycle feedwater pump [2].

1) k-Nearest Neighbour (k-NN) [3]: it is a relatively simple yet effective method for both regression and classification tasks. The physical meaning of the algorithm is straightforward, facilitating intuitive comprehension of its operational principles. However, k-NN requires a substantial amount of memory and computational capacity, and its performance is significantly influenced by the selection of the number of neighbors, k. In the context of rotating machines, k-NN is a prevalent methodology, particularly when discrete variables are involved. It is frequently employed in conjunction with dimensionality reduction techniques, such as principal component analysis (PCA), to enhance efficiency and precision.

2) Naive Bayes classifiers [4]: they are robust to missing values and efficient in terms of storage, which renders them attractive for certain applications. However, these methods rely on the assumption of strong independence among features and require prior knowledge of probability distributions, which can be a limitation. In the context of rotating machines, Naive Bayes is less frequently employed due to the interdependence of variables that is characteristic of industrial machine data.

3) Support Vector Machines (SVM) [5]: renowned for their exceptional accuracy in classification tasks and their capacity to approximate intricate functions with remarkable precision. Notwithstanding these advantages, SVMs are less efficient with large volumes of data and do not provide a direct physical interpretation of the model. In the context of rotating machines, Support Vector Machines (SVMs) are typically employed in conjunction with signal processing techniques for effective feature extraction, thereby enhancing their performance in the diagnosis of machine conditions.

4) Artificial neural networks (ANNs) [6]: they are capable of highly accurate classification and can effectively approximate nonlinear, complex functions. However, they are characterized by a multitude of parameters and are susceptible to overfitting. Furthermore, the opacity of the training process and the absence of a direct physical interpretation can present significant challenges. In the field of rotating machine diagnostics, artificial neural networks (ANNs) are the most commonly utilized technique, particularly in the diagnosis of rotor-related failures. In order for these applications to be effective, it is generally necessary to perform a process of feature extraction prior to their use.

5) Deep learning techniques [7]: they offer the advantage of automatically learning features and handling complex data structures. Furthermore, they necessitate the availability of extensive datasets and prolonged training periods, and they lack a direct physical interpretation, which can be perceived as a limitation. In the field of rotating machines, deep learning is demonstrating growing potential, particularly in the context of large datasets comprising direct mechanical signal data, where its efficacy is becoming increasingly apparent.

In [8] the most commonly utilized machine learning (ML) techniques for the detection of failures in nuclear power plants are delineated. In addition to the aforementioned techniques, the text makes reference to regression algorithms (linear and logistic) as well as random trees (RT). Conversely, Qian (2022) outlines the utilization of Deep Learning Reinforcement (DRL) as a predictive maintenance technique for fault diagnosis, specifically in nuclear power plants and with a particular focus on rotating machines. This technique combines the advantages of automatic feature extraction from deep learning and interactive learning from reinforcement learning, thereby achieving an optimal fit even with a limited number of fault data points. Two distinct DRL models are presented, one based on convolutional neural networks (CNN) and the other on gated recurrent units (GRU). Their performances are then compared with those of three base models: The models under consideration are SVM, CNN, and GRU without DRL. The proposed models are demonstrated to outperform the base models, achieving a diagnostic accuracy exceeding 99%.

The article [9] presents a model based on artificial neural networks (ANNs) for the detection of faults in centrifugal pumping systems. The model is designed to detect a total of 20 types of faults. To develop the model, training and test data were generated under different operating conditions by running the pumping system and creating several real-time faults in an experimental laboratory model. These included problems such as bearing wear and discharge valve leakage. A feature extraction method based on principal component analysis (PCA) was employed to reduce the dimensionality of the input features. The neural network utilized is a feedforward network trained using the backpropagation algorithm, which adjusts the connection weights to minimize the error function, thereby achieving a generalization capability that allows the network to handle previously unseen data. The results demonstrated that the model with extracted features achieved a 100% detection rate in only 170 epochs, while the model with all features achieved a 99.3% detection rate, indicating the efficacy of feature extraction in enhancing model performance and efficiency.

Conversely, [10] proposes an innovative technique for fault diagnosis in photovoltaic (PV) systems using ANNs, developing two different algorithms: one is based on signal thresholds to isolate faults with different combinations of attributes, and another is based on ANNs to identify faults with similar combinations of attributes. The simulation models were experimentally validated with data from a PV array at Jijel University, demonstrating that the algorithms can correctly locate and identify different types of faults. Additionally, the implementation of the technique on an FPGA was presented, showcasing its effectiveness in real and large-scale applications. This approach offers an economical and efficient solution to enhance the reliability of PV systems.

In [11], three machine learning methods for predicting and diagnosing failures in nuclear power plants are compared: Adaptive Neuro-Fuzzy Inference System (ANFIS), Long Short-Term Memory (LSTM), and Radial Basis Function Network (RBFN). Different Loss of Feed Water (LOFW) events in the reactor are modeled. The results show that ANFIS is superior in predicting the steam generator tube temperature, RBFN excels in predicting the mass flow rate at the reactor core inlet, and LSTM is best in estimating the fault severity, suggesting that a combination of these models could provide a more robust diagnostic system to improve safety and operational efficiency in the nuclear industry.

It is also noteworthy to mention the innovative approach proposed in [12]: the MACOL (Maintenance model Adapted to the Continuous Observed Life of industrial components). This approach is centered on the utilization of discernible risk indicators derived from the actual behavior of wind turbine components under standard operational conditions. The model permits the real-time adjustment of maintenance plans based on risk indicators, thereby reflecting deviations between actual and expected behavior. The optimal time for maintenance actions is determined by the application of fixed and variable health thresholds, thereby providing a flexible mechanism sensitive to operating conditions and increasing the efficiency of the maintenance strategy.

C. Main challenges [1]

1) Errors and noise in collected data: in industrial environments, sensors are exposed to extreme conditions that can generate erroneous or noisy data. Such inaccuracies may result in erroneous diagnoses or predictions. To address this issue, advanced anomaly detection techniques are being investigated that use machine learning algorithms to differentiate between normal data and anomalies resulting from sensor errors.

2) Lack of generalization: prediction models are often designed for specific conditions, particular types of machines or particular parts of machines, which limits their ability to be generalized to different contexts or equipment. Furthermore, the interdependence between different parts is frequently overlooked.

3) The necessity to collect and process data in a measured and effective manner across a vast array of industrial scenarios: this is a significant challenge, particularly in settings where production assets are diverse and geographically dispersed. The integration of technologies that facilitate local data processing, such as the Internet of Things (IoT) and Edge Computing, is regarded as a pivotal solution.

III. Project scope

A. Feedwater pumps in combined cycles

As previously stated, the project entails an analysis of the normal operating conditions of a feedwater pump in a combined cycle. The primary objective of a combined cycle is to leverage the synergies between a steam cycle (Rankine) and a gas cycle (Brayton) by utilizing the high temperature of the gas turbine exhaust gases to heat the inlet water of the steam turbine. The aforementioned heat exchange occurs within the recovery boiler. In this manner, the gas cycle exploits the high-temperature hot spot, while the steam cycle makes highly efficient use of the low-temperature cold spot.

The transfer of heat occurs via phase change at a constant temperature. In contrast to other cycles, the steam cycle does not involve combustion, thus eliminating the necessity for replenishing the working fluid. Furthermore, expansion can be conducted up to vacuum pressures, constrained only by the temperature of the cold focus. The boiler may be constructed with a single pressure level or with multiple levels, and may be configured in a series or parallel arrangement. The combined cycle of this project comprises two pressure levels arranged in a series. The feed water is preheated in the low-pressure boiler and subsequently conveyed to the turbine inlet pressure, where it is transformed into steam at that pressure in the high-pressure boiler, utilizing the higher-temperature gases [13].

The configuration of combined cycle power plants may vary in order to optimize performance. The most common configurations include single-shaft and multi-shaft setups, with variations such as the 1x1, 2x1, and 3x1 arrangements, in which multiple gas turbines feed steam into a single steam turbine. These configurations facilitate enhanced operational flexibility, efficiency, and maintenance accessibility. In a 2x1 configuration of a combined cycle plant, as is the case with the one under study, two gas turbines provide exhaust gases to their respective heat recovery steam generators (HRSGs). Subsequently, steam is obtained by heating the water in the corresponding boiler, which is then used to power a single steam turbine. This configuration provides enhanced operational flexibility and efficiency, particularly at partial loads, as each gas turbine can be operated independently [14].

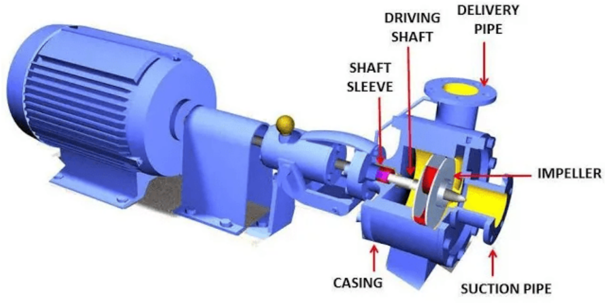

Fig. 1. Horizontal centrifugal pump [15]

A centrifugal pump works by converting mechanical energy from a motor into hydraulic energy to move liquids. As seen in the diagrams (Figures 1 and 2), the electric motor drives the drive shaft, which is protected by a sleeve. This shaft is connected by the bearings of the motor-pump coupling, which ensure smooth and aligned operation. The shaft drives the impeller, which is located inside the casing. Fluid enters through the impeller eye where it is caught by the impeller vanes. As the impeller rotates, it creates a centrifugal force that draws fluid through the suction tube and expels it at high pressure through the discharge nozzle. The stuffing box and packing prevent leakage around the shaft, while the wear ring on the casing minimizes frictional wear of the fluid. In addition, the thrust bearing helps support the axial loads generated by the operation of the impeller. This process allows efficient movement of fluid from a low pressure zone to a high pressure zone through the piping system.

In the context of the steam cycle, the primary function of the feed pump is to facilitate the supply of water to the recovery boiler. In the case study, the feed pump is situated between the low and high-pressure boilers. The feed pump draws water from the LP boiler and feeds it into the HP boiler. Additionally, the feed pump maintains the level of the HP boiler within specified limits and injects high-pressure feedwater between the HP superheater sections to temper the flow. Each boiler has two pumps that alternate, with one pump operating and the other always in standby. A diagram of a combined cycle similar to this can be seen in Figure 3.

IV. Failure modes

THE foundation for the standard operational conditions models is an exhaustive examination of the potential

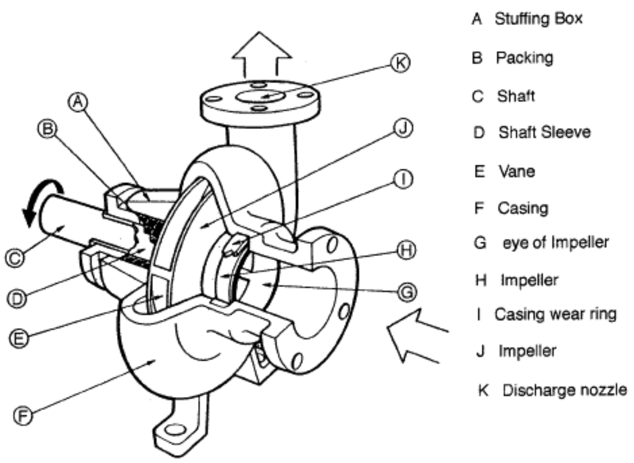

Fig. 2. Main components of a horizontal centrifugal pump [15]

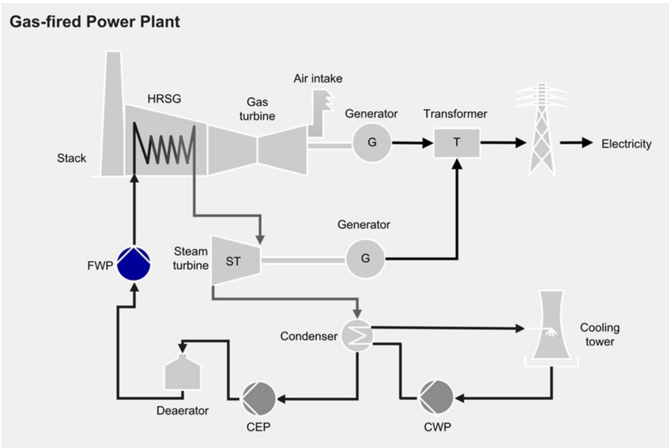

Fig. 3. Combined cycle diagram with a feed water pump [16]

failure modes associated with the pump. This is essential for the accurate selection of variables to be modelled, with a view to assessing the pump's correct behavior. A comprehensive examination of potential failure modes facilitates the identification of critical variables, including pressure, temperature, vibration, and flow rate, which are indispensable for evaluating the pump's performance. By grasping these variables, we can construct precise models that forecast the pump's conduct in diverse operational scenarios, thus guaranteeing dependability and efficacy in its operation. As outlined in [17], three distinct categories of failure can be identified.

A. *Hydraulic failures*

Hydraulic failures in pumps are caused by an inability to efficiently transfer fluid from the suction side to the discharge side. Such occurrences can be attributed to a number of factors, including but not limited to issues such as cavitation, blockages, hydraulic instability, and recirculation. Cavitation occurs when the local pressure within the pump drops below the vapor pressure of the liquid, resulting in the formation of vapor bubbles that collapse abruptly, causing damage to the impeller and other internal components. Blockages impede the flow of fluid, thereby reducing efficiency and potentially causing damage. Hydraulic instability is defined as fluctuations in the pump's performance due to irregularities in fluid dynamics, which often result in the generation of vibrations and noise. The phenomenon of recirculation, occurring in both the suction and discharge sides of the pump, is defined as the backflow of fluid within the pump, which can result from inadequate sizing or design flaws. In order to effectively monitor the occurrence of hydraulic failures, it is essential to closely observe several key variables, including the discharge flow, the pressure at both the discharge and suction points, and the temperature of the discharged liquid.

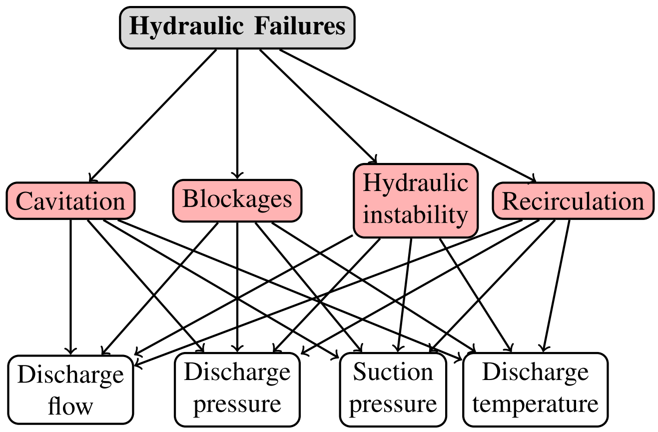

Fig. 4. Common hydraulic failure modes and related features

B. *Mechanical failures*

Mechanical failures encompass a broad range of issues primarily related to the deterioration of pump components due to the effects of wear and tear. Such issues include seal malfunctions and leakage, inadequate lubrication, worn bearings, and misalignment of the shaft. Bearing failures are frequently attributable to inadequate lubrication, misalignment, or contamination, which precipitate augmented friction and heat generation. This can culminate in pump seizure. Seal leaks may result from the degradation of seal material or improper installation, which can lead to fluid leakage and a reduction in pump efficiency. The deterioration of bearings can result in the generation of vibrations and an increase in noise levels, which may be indicative of the commencement of wear or misalignment in the components. Shaft misalignment can result in abnormal bearing vibrations, which in turn can have a detrimental impact on the overall performance and longevity of the pump system. Vibration and noise levels are crucial indicators of potential mechanical failures, as they frequently signal the early stages of component wear or misalignment.

C. *Electrical failures*

The majority of electrical failures in pumps are attributable to issues within the motor or control systems. The most common electrical failures include overloads, imbalances, and short circuits. An overload can result from

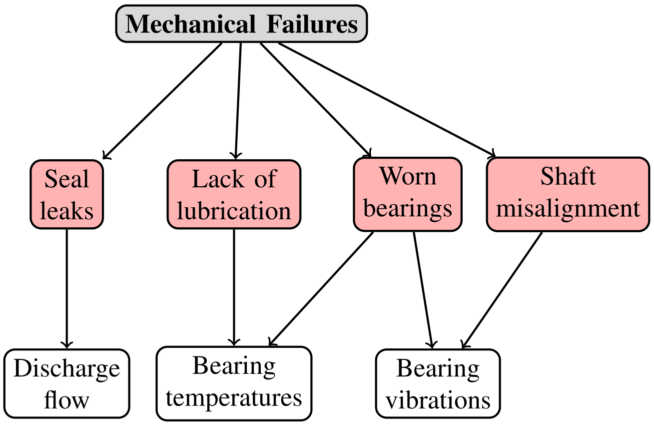

Fig. 5. Common mechanical failure modes and related features

excessive currents in the motor, which may lead to overheating and potential failure of the motor windings. Imbalances in the power supply, such as phase imbalances, have the potential to impair the motor's functionality, leading to inefficiencies or even complete cessation of operation. Short circuits, frequently attributable to insulation failure or electrical surges, can precipitate elevated motor temperatures and subsequent motor winding failures. It is essential to monitor key variables such as motor currents and temperatures in order to identify potential electrical failures.

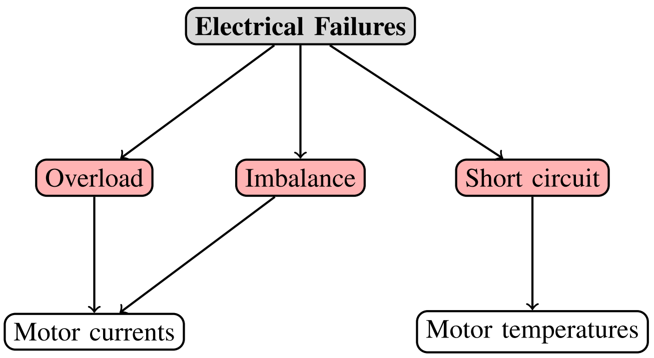

Fig. 6. Common electrical failure modes and related features

V. Models' development

Once the various modes of pump failure have been subjected to analysis, the development of models representing normal behavior may then commence. The following section will delineate the steps necessary to carry out this process, with each step illustrated by a real normal operational model for the pump subject of study. The model under consideration is the discharge flow rate, which may serve as an indicator of pump leakage.

A. Data preprocessing and filtering

The initial phase of this process entails the preprocessing and filtering of the data. The primary objective of this stage is to prepare the raw data for analysis, ensuring its quality and relevance to the model to be developed. This involves several essential steps that facilitate the isolation and structure of the information in an appropriate manner. The specific steps involved in this phase are detailed below.

1) Isolation of the operating period: at this stage, the time interval during which the system was operating at a constant rate is selected. It is of the utmost importance to prevent the introduction of non-representative data, such as periods of start-up or shutdown, into the analysis. Such data would not correspond to the normal mode of operation and would therefore introduce noise into the analysis. Figure 7 depicts the unfiltered data for one of the variables, discharge flow rate. It illustrates the presence of both operating and non-operating periods, as well as transient states. To proceed to the subsequent stage, the data from transient and non-operating states must be removed. The outcome of this filtering process is illustrated in Figure 8.

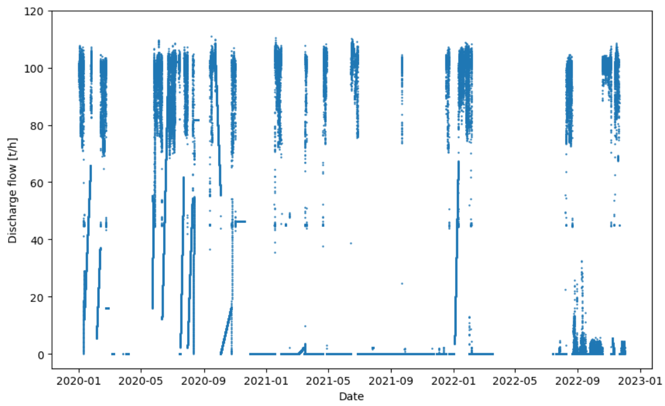

Fig. 7. Raw discharge flow data

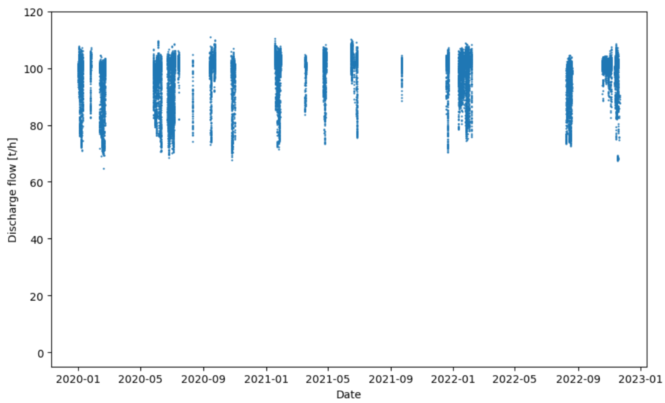

Fig. 8. Filtered discharge flow data

2) Identification of a valid training period: once the operating period has been isolated, it is necessary to identify a sub-period within it that is suitable for model training. This sub-period should be sufficiently lengthy to encompass significant system variations while excluding outlier events that do not align with the norm. To identify the optimal subset of data for training purposes, the pump failure and maintenance history is observed. In the case study, it was observed that no failures or major maintenance events occurred between January 2020 and March 2021. Therefore, this period was selected as the basis for obtaining the model. The data set is divided into two portions: 80% for training and 20% for testing. From the 80% designated for the training set, a 20% subset will be extracted as the validation set to prevent overfitting of the model. This division is illustrated in Figure 9. The discharge flow rate (x-axis) is plotted against steam power (y-axis) for each of the training, test, and all the available data.

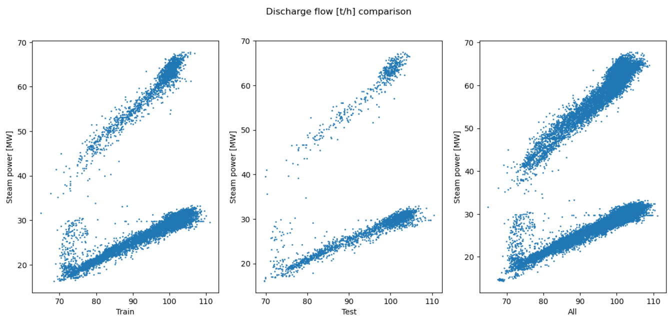

Fig. 9. Train and test sets vs all the data

3) Creation of the required virtual variables and data normalization: the subsequent phase is the generation of virtual or derived variables, which can enhance the predictive capacity of the model. The objective is to identify patterns and relationships that are not immediately apparent in the raw data. In this particular instance, two such variables have been constructed. The first variable is the mean phase intensity of the motor. This procedure yields a variable that is representative of all the currents and less dependent on possible network imbalances that may occur. The second variable is binary in nature and serves to indicate the operational state of the steam power within the pump. As illustrated in Figure 9, the two operational modes are separated by a limiting value of approximately 35 MW. Additionally, the variables have undergone normalization, with the training set exclusively utilized for this purpose. This transformation aims to enhance the model's performance and efficiency by preventing features with larger values from exerting an undue influence on the loss function and model parameters.

B. *Target feature and input feature selection*

Once the data has been subjected to preprocessing and filtering, the subsequent phase in the development of the model is to determine which variables are to be modeled. The selection of the target variable is contingent upon the specific failure mode that is to be identified. The variable to be selected is the one whose potential anomalous behavior serves as an indicator of the occurrence of the failure. In the illustrative example, the objective is to identify a potential leakage issue. Consequently, the discharge flow rate has been selected as the target variable for modeling.

The subsequent phase is the selection of the input variables for the model. To achieve this, it is necessary to identify those pump variables that have a physical relationship with the target variable. This selection process is a dynamic one that will require adjustment according to the results obtained during modeling and validation. It is possible that the initial hypothesis regarding input variables will be disproven by the data. Three input variables have been selected for the model under study: suction pressure, steam power, and mean phase intensity of the motor, due to their direct correlation with the discharge flow rate. Suction pressure serves as a pivotal indicator of pump inlet conditions. Low pressure may indicate supply issues or obstructions within the inlet piping, directly influencing the flow rate. Steam power quantifies the energy supplied to the system, and fluctuations in this power may reflect alterations in pump operational parameters. Finally, motor mean phase current represents the electrical current drawn by the pump motor. An increase in this current may suggest the necessity for additional motor effort to maintain flow.

C. *Selecting and training the model*

Once the target variable has been identified and the input variables that can explain the behavior of this output variable have been selected, the next step is to select and train the model. A multilayer perceptron (MLP) neural network has been chosen as base model for its capacity to capture intricate non-linear relationships between input variables and the target variable. Multilayer perceptrons are particularly well suited for this type of problem because they are capable of approximating highly complex functions and of discovering patterns in the data that other modeling techniques might fail to identify. Moreover, MLPs are robust and flexible, enabling the integration of multiple input variables with disparate characteristics. The hyperparameters for each model were as follows: the number of layers, the number of neurons in each layer, the learning rate, the batch size, and the dropout rate. An optimization of these hyperparameters has been conducted in order to identify the combination that produces the lowest error.

At the time of training, the mean square error (MSE) was employed as a metric. To prevent overfitting, a neuron dropout has been incorporated between each layer, and the MSE of the validation set has been monitored during training to enable early stopping when no improvement in this metric is observed. The following hyperparameters were employed for the model of the discharge flow rate:

- Number of layers: 2
- Number of neurons layer 1: 16
- Number of neurons layer 2: 13
- Learning rate: 0.0082
- Batch size: 95
- Dropout rate: 0.21

Figure 10 illustrates the outcome of the training process. The x-axis represents the actual values, while the y-axis represents the predicted values. The closer the point cloud is to the x = y line, the more accurate the model is. It can be observed that the model demonstrates a high level of fit. The mean square error (MSE) was 0.08 for the training set and 0.02 for the validation set.

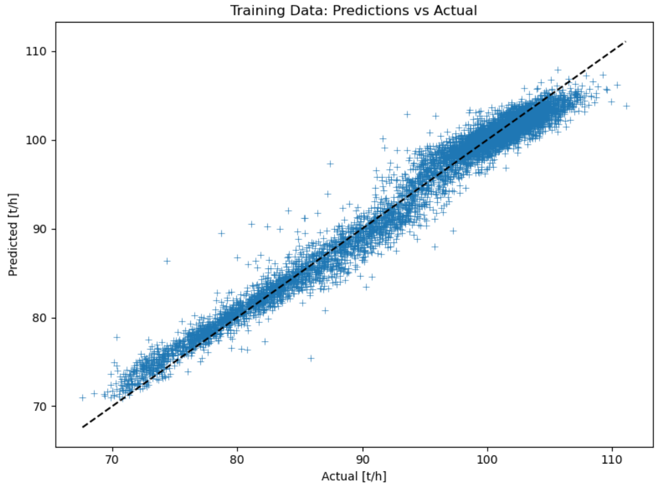

Fig. 10. Discharge flow training: predictions vs actual

D. Validation and residual analysis

Once the model has been trained, it is essential to validate it to ascertain whether it has been able to generalize the typical behavior of the pump. First, it is verified that the distribution of the residuals is approximately normal. Figure 12 illustrates that the residuals obtained from the training of the discharge flow rate model satisfy this condition. Subsequently, the model is employed in the test set to ascertain whether overfitting has occurred, yielding the results depicted in Figure 11. The MSE for the test set is 0.02, similar to the ones obtained in training. It is then evident that the model exhibits a high degree of fit with the test set, thereby providing compelling evidence that it has effectively generalized the flow behavior.

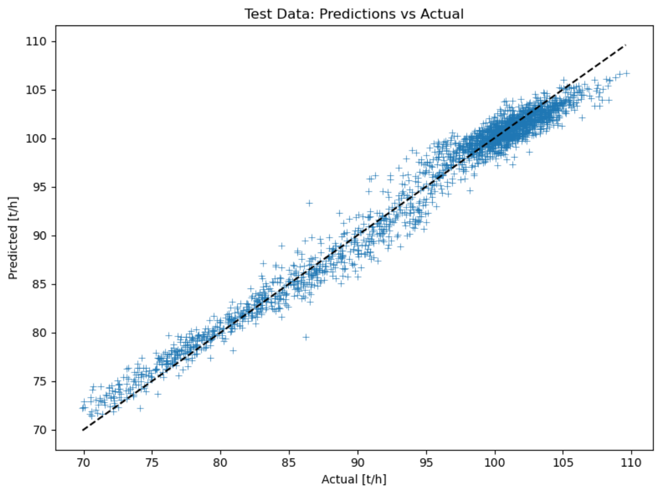

Fig. 11. Discharge flow test: predictions vs actual

A further issue that requires examination is the relative importance of each input variable in the model. To this end, the integrated gradients [18] algorithm has been employed. Figure

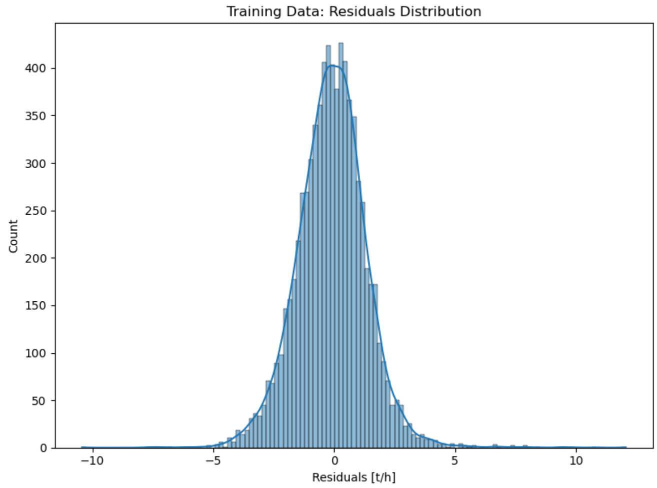

Fig. 12. Discharge flow training residuals distribution

13 illustrates the relative importance of each of the three input variables for the discharge flow model. As can be observed, the average intensity is the most significant variable, aligning with the initial hypothesis that changes on the intensity are strongly reflected on the discharge flow value.

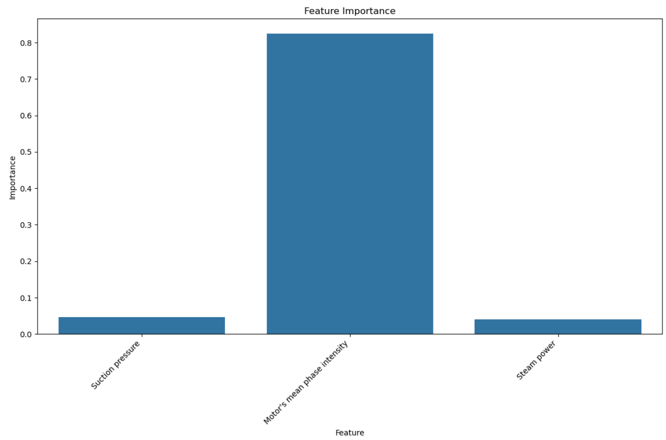

Fig. 13. Discharge flow feature importance

At this point in the analysis, an appropriate model is available for the examination and identification of pump failures, as will be explained in the following section.

VI. Failure risk assessment

In order to identify pump failures using the normal behavior model, it is first necessary to obtain the Gaussian mixture that best fits the model errors (Figure 14). Subsequently, the 99% confidence intervals for normal operation are calculated from the normalized training residuals distribution (Figure 15). When the model is applied to new data, any behavior that falls outside the aforementioned intervals will be considered anomalous. Conversely, if the data falls within the specified range, it will be considered to exhibit normal behavior.

In order to achieve a temporal representation that demonstrates the state of this potential failure mode that has been

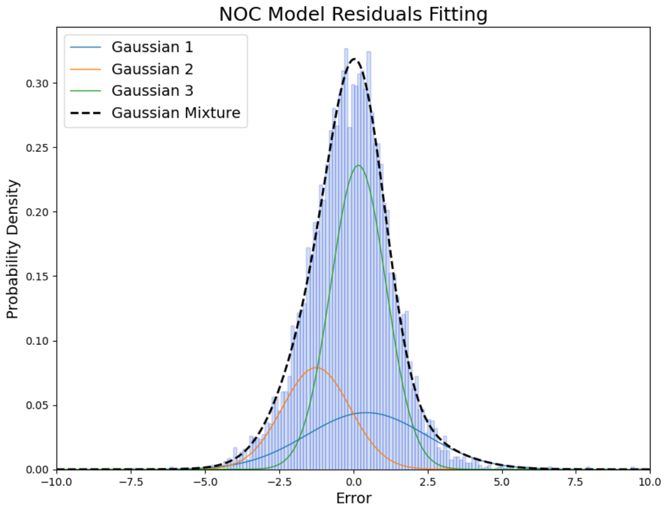

Fig. 14. Gaussian mixture of discharge flow model residuals

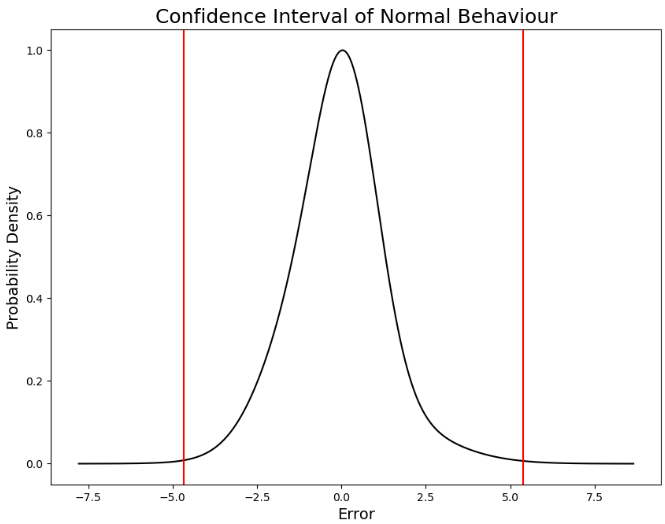

Fig. 15. Confidence interval of normal behavior for discharge flow model

analyzed, the deviations of the residuals with respect to the normal behavior model are represented, as illustrated in Figure 16.

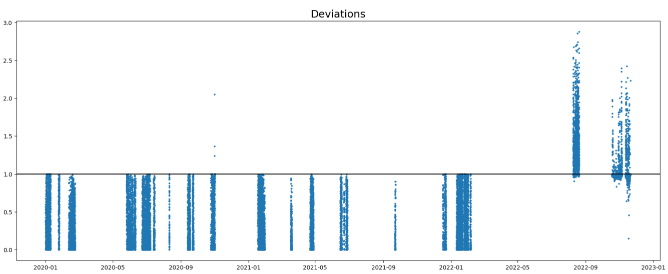

Fig. 16. Deviations using discharge flow model

The accumulation of deviations exceeding the unit yields the risk curve, which is employed for the detection of anomalies. Figure 17 illustrates the risk curve for the discharge flow. It can be observed that the behavior is within the normal range until the end of 2022, at which point an increase in anomalous values is noted. This may be indicative of a potential leakage. The model developed would have enabled the identification of this failure prior to its occurrence, thereby allowing the implementation of measures to mitigate its impact.

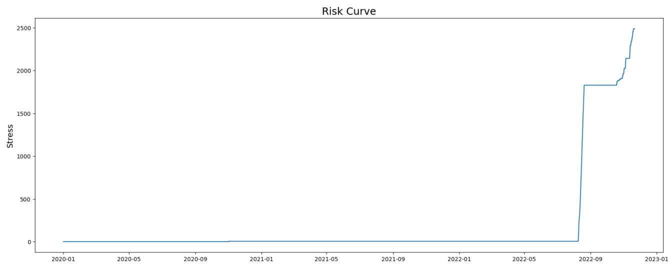

Fig. 17. Risk curve of discharge flow model

VII. Conclusion

THIS Master's Thesis has demonstrated the significant potential of artificial intelligence and machine learning techniques in the diagnosis and maintenance of critical assets within combined cycle power plants. By focusing on a feed-water pump, the study has showcased how normal operational behavior can be modeled, and how anomalies and potential failures can be preemptively detected. The application of AI methodologies enables a transition from a reactive to a proactive maintenance approach, thereby markedly reducing downtime and enhancing operational efficiency.

The research employed a comprehensive three-year data set and advanced AI tools, underscoring the benefits of using Python, PyTorch, and PyTorch Lightning for such applications. The study addressed key challenges, including the issue of data noise and the need for effective data collection across a range of industrial environments. The findings emphasize the importance of integrating machine learning models with expert knowledge to enhance predictive capabilities.

The developed model exhibited high accuracy in predicting the normal behavior of the discharge flow of the feed-water pump and identifying deviations that indicate potential failures. The approach used in this study is adaptable to other industrial assets, provided that normal operating data is available, thus making it a versatile tool for improving maintenance strategies across various sectors.

Future research could explore the integration of more sophisticated AI models and techniques, such as reinforcement learning, to further enhance predictive maintenance capabilities. Additionally, the implementation of real-time monitoring systems and the continuous updating of models with new data will ensure ongoing accuracy and reliability in failure prediction and maintenance optimization.

Acknowledgment

The author would like to express his gratitude to his thesis supervisor, Antonio, for his invaluable guidance, support, and encouragement throughout the course of this research. His expertise and insights have been instrumental in the development of this thesis. Furthermore, the author would like to express his gratitude to Javier, whose previous research provided a solid foundation for this study. His contributions have constituted a significant source of inspiration and direction.

REFERENCES

[1] 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.

[2] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[3] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited," *Mechanical Systems and Signal Processing*, vol. 70, pp. 201–208, 2016.

[4] V. Muralidharan and V. Sugumaran, "A comparative study of naïve bayes classifier and bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis," *Applied Soft Computing*, vol. 12, no. 8, pp. 2023–2029, 2012.

[5] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (svms)," *Applied Soft Computing*, vol. 11, no. 6, pp. 4203–4211, 2011.

[6] J. Rafiee, F. Arvani, A. Harifi, and M. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical systems and signal processing*, vol. 21, no. 4, pp. 1746–1754, 2007.

[7] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[8] G. Hu, T. Zhou, and Q. Liu, "Data-driven machine learning for fault detection and diagnosis in nuclear power plants: A review," *Frontiers in Energy Research*, vol. 9, p. 663296, 2021.

[9] S. Rajakarunakaran, D. Devaraj, and K. Suryaprakasa Rao, "Fault detection in centrifugal pumping systems using neural networks," *International Journal of Modelling, Identification and Control*, vol. 3, no. 2, pp. 131–139, 2008.

[10] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. M. Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renewable Energy*, vol. 90, pp. 501–512, 2016.

[11] A. M. Mwaura and Y.-K. Liu, "Prediction of the loss of feed water fault signatures using machine learning techniques," *Science and Technology of Nuclear Installations*, vol. 2021, no. 1, p. 5511735, 2021.

[12] R. J. de Andrade Vieira and M. A. Sanz-Bobi, "Failure risk indicators for a maintenance model based on observable life of industrial components with an application to wind turbines," *IEEE Transactions on Reliability*, vol. 62, no. 3, pp. 569–582, 2013.

[13] J. I. L. Hurtado, "Eficiencia en ciclos combinados," 2024, ingeniería Energética, Master en Ingeniería Industrial, ICAI. Accessed: 2024-07-12.

[14] S. G. Garrido, "Tipos de centrales de ciclo combinado: configuraciones habituales," 2024, accessed: 2024-07-12. [Online]. Available: https://www.cicloscombinados.com/index.php/articulos/articulos-avanzados/centrales-termicas-de-ciclo-combinado/tipos-de-centrales-de-ciclo-combinado

[15] N. Pump, "Horizontal centrifugal pump manufacturer," 2024, accessed: 2024-07-16. [Online]. Available: https://ntgdpump.com/horizontal-centrifugal-pump-manufacturer/

[16] S. Ltd, "Feed water pump (fwp) for gas-fired plants," 2024, accessed: 2024-07-12. [Online]. Available: https://www.sulzer.com/en/shared/applications/feed-water-pump-for-gas-fired

[17] K. McKee, G. Forbes, M. I. Mazhar, R. Entwistle, and I. Howard, "A review of major centrifugal pump failure modes with application to the water supply and sewerage industries," in *ICOMS Asset Management Conference Proceedings*. Asset Management Council, 2011.

[18] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 3319–3328.