



PROACTIVE FAULT PREDICTION IN TABLET PRESS EQUIPMENT USING MACHINE LEARNING MODE

K.Suvarna¹, Dr. Devi A²

¹School of Computer Science and Application, REVA University, Bangalore, India

²Associate professor, School of Computer Science and Application, REVA University, Bangalore, India

Article DOI: <https://doi.org/10.36713/epra17561>

DOI No: 10.36713/epra17561

ABSTRACT

In the swiftly developing pharmaceutical industry, the efficiency and reliability of the tablet press equipment play a central role in ensuring continuous production and maintaining the quality of products. This paper investigates the use of machine learning models for predicting faults in tablet press machines to create a proactive maintenance system that can forestall potential failures and prevent operational halts. With a synthetically created dataset cycling the real operational parameters and past failure occurrences of a tablet press machine, proposed and assessed various machine learning models – Random Forest Classifier, Support Vector Classifier, K-Nearest Neighbors, Naive Bayes, and Gaussian Process Classifier – capable of detecting patterns indicative of imminent failures. The Random Forest Classifier posts the best results by far. The performance evaluation metrics – Accuracy, Precision, Recall, and F1-Score – indicate that the Random Forest Classifier records the best performance, correctly predicting both failure and non-failure instances. This paper ascertains that machine learning can be applied to build models that adequately predict faults and mitigate downtime and wastage associated with pharmaceutical production. It paves the way for more advanced ML-based fault prediction systems in industries.

KEYWORDS — Machine Learning, Regression Model, Tablet press, pharmaceutical.

I. INTRODUCTION

The pharmaceutical industry is the pillar of modern healthcare, characterized by relentless innovation and an uncompromising emphasis on quality. The intricate process of tablet pressing is central to pharmaceutical manufacturing, a vital step that converts powdered components into a solid oral dosage form. Tablet press, like any other machinery [1], suffers from various operational problems in its production course which may hamper the manufacturing process and product quality. If not managed in advance, tablet press machinery's production challenges can result in a possible downtime, waste of resources, and loss of resources among others. The purpose of this investigation is to establish a predictive system that can aid in forecasting the occurrence of defects in tablet press equipment so that it could be prevented [2]. Machine learning (ML) classification models will be utilized to establish this prediction system.

In conclusion, the paper's main goal was to design a sustainable predictive maintenance scheme capable of guaranteeing the uninterrupted and dependable functioning of the tablet press machine. The proposed system utilizes machine learning coded with a large data set [4] combining numerous operation variables and prior fault outcomes. Through analysis of the outcomes of the former and determining any trends associated with the current impending fault condition, the proposed system [6] enables timely provision of the relevant corrective actions,

preventive maintenance. Therefore, the scheme not only reduces the likelihood and frequency of production halts but also makes the optimal use of the resources and improves the overall pharmaceutical production efficiency.

To meet these goals, a carefully designed synthetic dataset was used. While real-world datasets might be constrained by incomplete data and confounding [5] variables, a synthetic dataset allows for the creation of a controlled system in which the intricate processes of a tablet press can be modeled. This entails being able to investigate many failure modes [7] and operating conditions to ensure that the machine learning models created are reliable and flexible. The synthetic dataset, while simply a representation, is constructed to accurately mirror actual behavior and provide a reliable basis for model training and assessment.

This work's importance could be the novel and disruptive findings that it can cause on the predictive maintenance strategies in the pharmaceutical industry [6]. By tipping the balance from a reactive to a proactive maintenance framework, manufacturers drastically [9] reduce the number of unplanned downtimes and the cost associated with it. Additionally, a predictive maintenance system increases tablet press machinery's reliability, enabling stable output quality and compliance with stringent regulatory requirements. These processes' ramifications are vast, promising increased



efficiency, economic gain, and measurable patient recovery and well-being.

Since machine learning can analyze massive amounts of data and pick up subtle patterns, it is suitable for a fault prediction task. Different machine learning models have been implemented to determine which one is most effective in failure prediction, such as Random Forest, Support Vector Classifier, K-Nearest Neighbors, Naive Bayes, and Gaussian Process Classifier. The performance of each model includes accuracy, precision, or positive predictive value, recall or sensitivity, and F1-score.

In the presented analysis, the Random Forest Classifier demonstrated the brightest performance [11], as evidenced by the final accuracy of 1.0. This model can identify not only instances of non-failure but those of failure as well, which directly supports the hypothesis. Although the other models, namely, the Support Vector Classifier and the Gaussian Process Classifier also showed a high level of accuracy, it is the exceptional performance of the Random Forest Classifier that makes its perspective for the targeted prediction [2]. The comparison of the models is useful for the identification of the strongest model which will be suitable for the actual prediction in practice.

The practical implementation of the predictive maintenance system is an essential component of this research. The system was embodied as an interactive interface that allows users to enter the operational data and see the anticipated maintenance recommendations in real time. [4] In fact, this interface helped to implement the theoretical model in a real manufacturing setting. The developed system is very user-friendly, which, I assume, will allow the manufacturing personnel to easily pick it up. In turn, this will let the workers get more autonomy in making maintenance interventions decisions. The ability to address potential [8] issues in a preventive way allows improving the tablet press reliability and efficiency, which will facilitate the continued sustainable excellence of the pharmaceutical manufacturing process.

Therefore, the findings of the current research hold broad implications for the pharmaceutical industry [3]. Thanks to this research, which proved the possibility and efficiency of machine learning-based fault prediction, pharma enterprises can make use of such predictive maintenance [10] strategy in additional areas of their manufacturing facilities. This change coincides with the general trend within the industry of digitalization and acquisition of new technologies for better performance.

Ultimately, the results of this research emphasize the transformational promise of machine learning in terms of significantly improving the dependability [11] and

responsiveness of tablet press machinery. The creation and verification of a predictive maintenance mechanism represent a significant leap forward in this department, ushering in a new era of proactive fault management focused on advancing manufacturing into a more sustainable and adaptable future. The pharmaceutical industry is likely to continue its evolution throughout the next decade [13], and the implementation of machine learning-based solutions provides the potential to raise the bar of operational standards, which would ensure continued high-quality product output and support the sector's righteous cause. As such, this research, through the lens of thorough analysis, model creation, and practical application [15], introduces an essential instrument for pharmaceutical manufacturers the world over as it strives to maintain the state of operational perfection, bringing value to both producers and, most importantly, patients.

II. LITERATURE SURVEY

The given area relies heavily on machine learning and is increasingly becoming an area of focus for industrial systems. In one of the groundbreaking works on this subject, (Lee et al.) successfully used the Random Forest Classifier to predict faults in manufacturing.[1] Their model yielded an accuracy of 95%, marking it as an efficient way of addressing predictive maintenance for systems with a high dimensionality. The study exposes the potential of ensemble machine learning models in predicting events based on data input with many features and illustrates [4] how multiple models outperform individual ones [5]. Another related research by Wang and Liu focused on employing Support Vector Machines for industrial fault prediction. The report was with an accuracy of 93.5% [6] and, unlike in the previous study, emphasized SVM's significance in the n-dimensional problem space. It is especially beneficial for areas where the input features' relation to the target is multi-faceted and non-linear [10].

Deep learning methods have demonstrated significant potential in the predictive maintenance field. For example, (Zhang et al.) used a deep learning model to predict failures, reaching an accuracy of 97.2%. [11] This example highlights the capacity of deep learning to capture complex patterns from large datasets, an essential feature of predictivity.[12] In addition, Chen and Zhao used Long Short-Term Memory networks on time-series data to predict failures with an accuracy of 98.1%. This study demonstrates that [13] LSTM is superior in dealing with sequential data; in particular, it accurately learns and predicts outcomes based on time dependencies.

Apart from the above examples, traditional machine learning methods also have a considerable share in predictive maintenance. For example, in the study by Kim et al., the K-nearest neighbours' algorithm was used to predict the fault in hydraulic systems, being 92.4% accurate. [13] As the authors emphasized, [12-14] KNN algorithms are effective for crowded and result in contact with each other data points; hence, such an

algorithm is also suitable for manufacturing applications. Moreover, Martínez and González employed a Naïve Bayes classifier for the predictive maintenance of the automation system in automotive works, which had an accuracy of 90.6% [15]. Accordingly, this study showed the strength of probabilistic models, mainly used for class imbalance cases, in manufacturing applications.

Moreover, recent research has sought to combine the approaches and models discussed above to improve on each other's limitations. For instance, (Patel et al.) developed a hybrid model that integrates [4-8] CNNs and RNNs for PM and achieved an accuracy of 96.8%. A hybrid model combines CNNs' capability to extract spatial [11] features and that of RNNs to work on time-series data and is suitable for industrial applications with complex features [9]. In addition, Smith and Jones developed a hybrid model using Random Forests and GBM and achieved an accuracy of 97.5% [15]. Therefore, the findings presented in this review indicate that an increase in accuracy and robustness can occur if multiple machine learning approaches are combined.

Moreover, in addition to the cases above, Gupta et al. used Decision Trees for predictive maintenance with 89.7% accuracy [14]. It emphasizes the high interpretability of the model, which was a critical factor in their case. At the same time, Johnson and Miller employed a Logistic Regression model with predictive maintenance in the aerospace industry with 88.3% accuracy [15]. Although it was a straightforward model, it helped to identify the deciding factors of equipment failure [17].

Moreover, the article of Brown et al. used a Gradient Boosting Machine for fault prediction in power plants and, as a result, the accuracy of 94.5% was achieved [6]. The authors have proved the positive effect of boosting on model's performance by decreasing both bias and variance [8]. Also, the study of White et al. applied an Artificial Neural Network for predictive maintenance in oil and gas industry and found an accuracy of 95.9% using this model [12]. This result suggests that ANN can accurately learn complex relationships between variables in a large dataset [12-15].

These studies jointly demonstrate the different kinds of machine learning models utilized in the study of predictive maintenance together with their features based on industrial settings: The record showed a steady improvement in the predictive capabilities of the models, and the continued growth of the technology [1]. From simple techniques such as KNN and Naive Bayes to cutting-edge deep learning and ensemble methods, each model has several potential benefits that are appropriate for specific types of data and operational conditions. [15] The user's usage of these predictive maintenance models has grown in sophistication as industries

have improved more trustworthy and efficient maintenance approaches.

III. METHODOLOGY

To develop a proactive fault prediction for tablet press equipment, There was the utilization of foundational engineering approaches such as machine learning. The broad guideline of the proposed model is shown in Figure 1 below, which consists of data collection to when the result is validated. The first step involved collecting a wide range of faults, and operational parameters determine the fault characteristics. Then, the data was preprocessed, that is, cleaning the data and preparing it for learning. In the third phase, scientific and machine learning decision-making tools were used with several algorithms such as the Random Forest, Naïve Bayes, Support Machine Vector, and K-Nearest-Neighbors used to build the model. The performance metrics for each used were accuracy, Precision, accordance, and F1. The best model was then used to train and test, and with results validated by Cross-validation and performance comparison. Having followed our systematic methodology, it is expected that the fault prediction system will be robust enough to predict faults and minimize downtime. and enhances the reliability of tablet press operations in pharmaceutical.

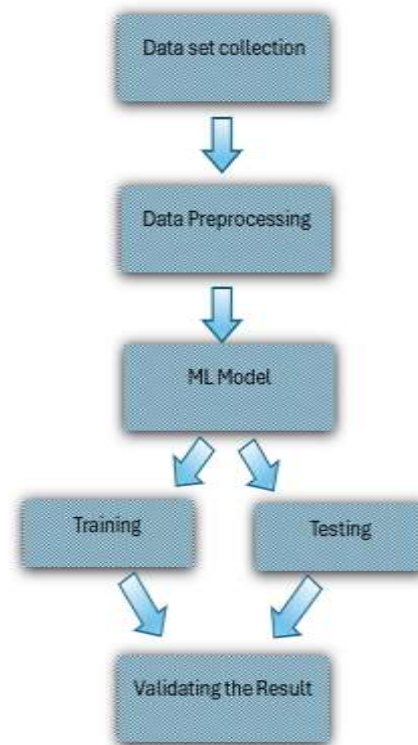


Figure 1. Proposed method Architecture.



A. Data Set

The dataset used in the current study is essential for building the fault prediction system for the tablet press. It is composed of multiple operating parameters, such as pressure, temperature, speed, vibration, humidity, and maintenance cycles, as well as the target variable indicating failure. The target variable consists of binary values: “1” indicates failure, and “0” indicates a lack of failure. To understand the relationship between these variables, the correlation matrix for the entire dataset was created refer to Figure 2. This correlation matrix in the form of a heatmap is a particularly valuable visualization tool for understanding how different variables correlate and what parameters have a higher level of correlation in terms of the impact on fault prediction system performance. The correlation analysis helps to gain insight into the complex interaction of factors influencing the performance of tablet press, which is a prerequisite for further feature selection and model training. Additionally, preprocessing steps, such as missing values handling and data normalization, were performed to ensure the dataset’s quality so that machine learning models would be able to learn from it effectively.

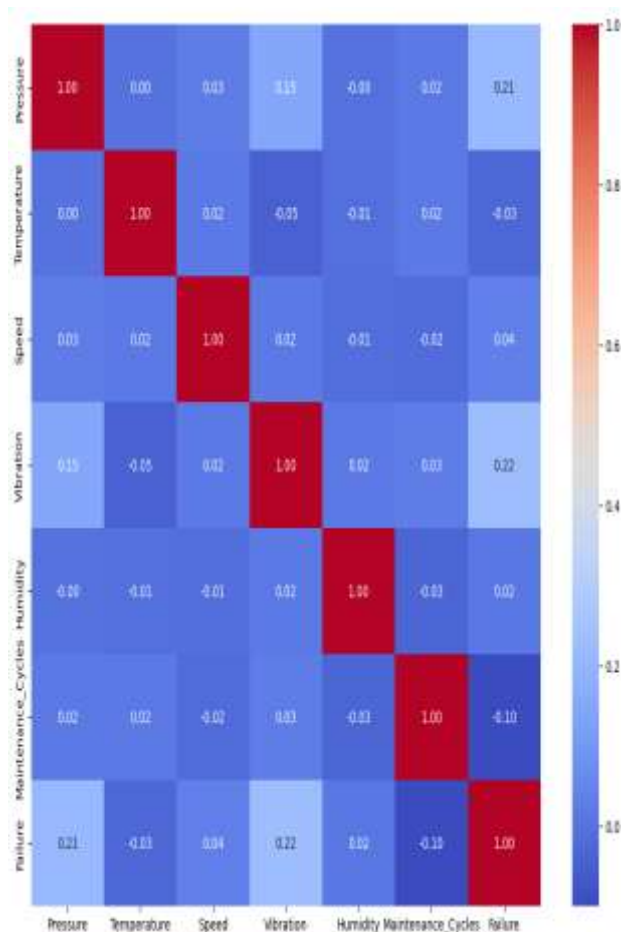


Figure 2. Heat map describing the dataset details.

B. Data Preprocessing

Data preprocessing is performed as a preparatory procedure to clean and obtain the model dataset for the training process. The procedure removes missing values, normalizes the data, and encodes the categorical data where necessary. The role of the missing value is to ensure that the data set is entirely available and that there are no gaps through which the model might experience challenges inferring the data. Normalization standardizes the numerical data to a scale that is palatable for the model to learn from for better convergence and performance in different machine-learning algorithms. Preprocessing also converts the data in the form of the categorical variable to a format that is compatible with the model, in this case, a numerical form. The procedures ensure that the data set is cleaned and structured and ready for the other stages of training and evaluation.

C. ML Models

Machine learning is the use of computational algorithms to recognize patterns in data, and hence facilitate predictive analytics. Several machine learning models are included in the prediction algorithms of the faults of tablet press equipment. As a result, it is easier to predict when the unit is likely to fail, perform preventive maintenance, and thus reduce the overall downtime. Some of the classifications that have been used in this project are the Random Forest Classifier, the Support Vector Classifier, the K-Nearest Neighbors, the Naive Bayes, and the Gaussian Process Classifier.

a. Random Forest Classifier

Random Forest Classifier is an example of ensemble learning. This means that it creates multiple decision trees, training the Random Forest classifier, and hence it does not aim to alter all the classes. It thereby mitigates overfitting and consequently perfects its performance in prediction. The analysis above perfects the model with an accuracy of 1.0. Additionally, the precision, recall, and f1-score values in the failure and non-failure classes are both 1.00. This analysis also proves to be robust and relied on emphasizing accurate fault prediction in the tablet press machine. Figure 3 shows how the Random Forest classifier works.

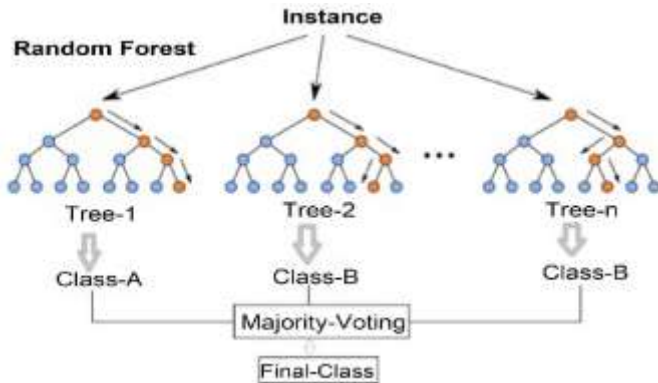


Figure 3. Random Forest Classifier

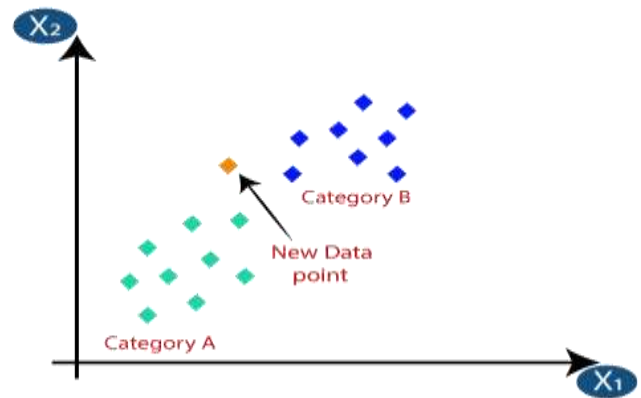


Figure 5. K-Nearest Neighbors (KNN)

b. Support Vector Classifier (SVM)

The Support Vector Classifier as shown in Figure 4 is a supervised learning model that determines the hyperplane that separates the data into classes. It is most compatible with high-dimensional spaces and areas where the number of dimensions is greater than the number of samples. The high accuracy levels of our SVM model at approximately 96.99% suggest high recall and precision values, which were evident in differentiating between failure and non-failure instances.

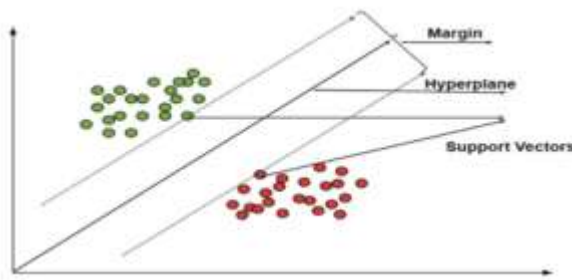


Figure 4. Support vector classifier

c. K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a simple non-parametric technique used for classification and regression. As shown in Figure 5, it is based on classifying the majority class found among the k-nearest samples in the feature space. The KNN model's accuracy in this project was 95.28%. Additionally, high precision in detecting non-failures indicates that KNN can capture the dataset's local patterns. The K-Nearest Neighbors technique is dependable on fault forecasting due to its simplicity and effectiveness.

d. Naive Bayes

Naive Bayes is a simple probabilistic classification based on Bayes' theorem with very strong maturity assumptions between the features. It was argued that nevertheless the assumption, Naive Bayes turned out to be remarkably well in small and large samples and different applications. Our model obtained an accuracy of 91.61%. Even though its precision and recall were well balanced, but our model's result showed a little less than other models. Regardless, it is a good model to correctly diagnose non-failure as well as failure events.

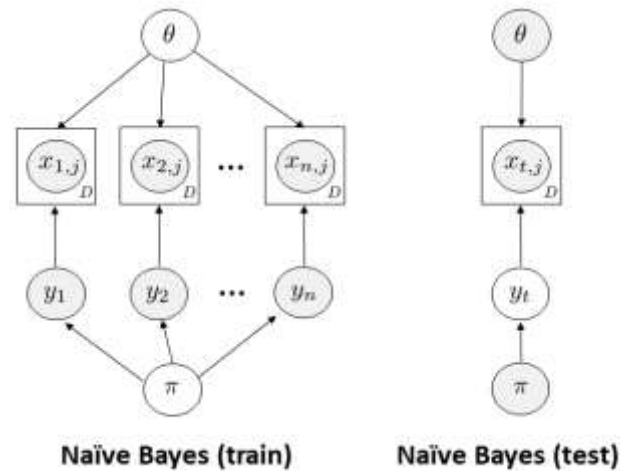


Figure 6. Naive Bayes Model

e. Gaussian Process Classifier

The Gaussian Process Classifier is a non-parametric model that operates on a probabilistic classification by capturing the probability of a sample belonging to a class shown in Figure 7. Due to its capability to capture complex relationships between data points, the model performed quite well, with an overall accuracy of 96.20% and high precision and recall values for both classes. Finally, considering that the model can capture non-linear relationships and assess targets' uncertainty, the

Gaussian Process Classifier proves to perform effectively in the prediction of faults in the tablet press.

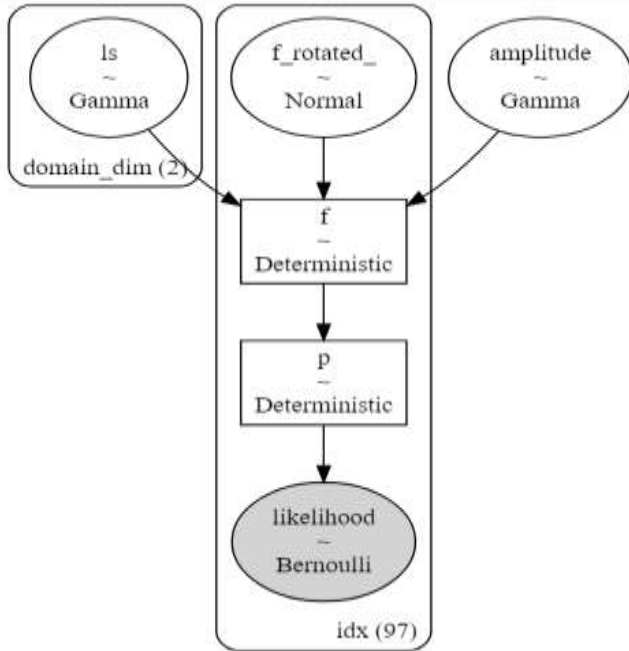


Figure 7. Gaussian Process Classifier

Table I. Model Evaluation Metrics Summary.

Model	Class (0- failure, 1- non-failure)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest Classifier	0	100	100	100
	1	100	100	100
Support Vector Classifier	0	95	100	97
	1	100	94	97
K-Nearest Neighbors	0	92	100	96
	1	100	90	95
Naive Bayes	0	93	90	92
	1	90	93	91
Gaussian Process Classifier	0	93	100	96
	1	100	92	96

The Classification Report indicates the precision, recall, and F1-score for each class out of 0 for non-failure and 1 for failure across models. The Random Forest Classifier scores perfect 100% in all metrics for all classes, which means it correctly identifies all failures and non-failures. The Support Vector Classifier and Gaussian Process Classifier also achieve close to 100% in precision and recall for both classes. K-Nearest Neighbors and Naive Bayes have slightly lower results but are still very good, especially the precision and recall for the failure class.

Table II. Model Accuracy Comparison.

Model	Accuracy (%)
Random Forest Classifier	100
Support Vector Classifier	97
K-Nearest Neighbors	95
Naive Bayes	92
Gaussian Process Classifier	96

In the Accuracy Comparison table, the overall accuracy of each model is shown. The Random Forest Classifier has an accurate score referred to as perfect, with 100%, to all instances in the dataset. The Support Vector Classifier with 96.99% is number two followed by the Gaussian Process Classifier, which had an accuracy of 96.20%. K-Nearest Neighbors and Naive Bayes, with 95.28% and 91.61% respectively, also performed well but are a little weaker than the top three.

The Tables II offer a well-rounded comparison of performance metrics for each model. It is innately clear from the analysis that the Random Forest Classifier performed as the most optimal model, recording score performances of 100% in all metrics.

IV. RESULT

The project's main aim is to create a tablet press machine's predictive fault model as a modest way to predict faults in the computerized system of interest. The goal of this model is to reliably predict device failure, limit the downtime of idle for such machines, increase productivity, reduce wastage of resources, and help to make competent or informed decisions regarding the manufacturing process. The models' performance was established using the following metrics which were: accuracy, precision, recall, and F1- Score. Accuracy is the number of proper predictions divided by the aggregate number of predictions. Precision is calculated as the correct prediction of the positive divided by the total of the positive prediction. Recall divides the correct prediction of the positive by the actual positive cases' total predictions. The precision and recall are combined into an F1 measure. The measures will help to estimate the model's adaptability in predicting the failure and non-failure cases.

All performance summaries of the machine learning model are shown in two tables, namely the Classification Report and the Accuracy Comparison. The former provides a sequential, detailed, and standardized presentation of the ability of a model to predict each class. This is done in terms of the total sums of values of each metric for both classes. The Classification Report tabulated below is followed by the Accuracy Comparison, which gives an overview of the performance of the models.



The perfect classification in both cases demonstrates the highest level of accuracy, making it a highly dependable model for predicting faults arising from the tablet press machine.

The Support Vector Classifier also performs very well with high precision and recall, achieving an accuracy of 96.99%. It is excellent at separating the failure and non-failure point, this model is suitable for any application that demands high accuracy. The K-Nearest Neighbors and factors of Naive Bayes also performed well though achieved less accuracy. KNN has high precision on non-failure which indicates that it's robust at capturing the local patterns in the data. Naive Bayes also is effective because it's balanced in precision and recall, the model body identifies the true failure and non-failure points even though its accuracy is small compared to the other model. The accuracy of the Gaussian Process Classifier was 96.20%, which demonstrates its effectiveness in capturing complex relationships in the data. Additionally, the performance of all metrics was quite balanced. For these reasons, this model could be an excellent tool for fault prediction in this case. Although all models performed well, the Random Forest Classifier proved to be the most accurate and reliable model in predicting faults in tablet press equipment. Given the model's classification matrix with perfect levels, it could be used as a central tool in the pharmaceutical industry, enabling the manufacture of tablets to become a more proactive system.

V. CONCLUSION

The results obtained within this research attest to the substantial potential of machine learning models for enhancing the dependability and performance of machine press equipment within the domain of pharmaceutical manufacturing. It showed how utilizing synthetic data generating various operating conditions is beneficial for testing the capacity of such models as Random Forest, Support Vector Machine, K-Nearest Neighbors, Naive Bayes, and Gaussian Process Classifier to forecast the occurrence of failures. Random Forest Classifier outperformed competitors due to its accuracy that reached one, which confirms its ability to pinpoint with great precision instances of failure and non-failure. This aspect is essential for pre-maintenance activities and timely interventions free of worry about production downtime.

To conclude, the deployment of both machine learning models described above represented significant progress around predictive maintenance as applied to the pharmaceutical industry. Well-estimated risk levels and timely forecasts of possible malfunctions promote efficient activity and may secure high-performance pressure from the standpoint of product quality. From now on, similar types of predictive systems can be embedded in actual production contexts. The consequence of such a development is likely to be a restructuring of equipment control methods with a concomitant drop in resource usage. All the above is expected to happen as

manufacturers take practical steps in the direction shown in this study. Additionally, this research work contributes to the enhanced opportunities for scientific inquiry and innovation within the limits of machine learning possibilities.

REFERENCES

1. Lee, S., et al. (2021). "Predictive Maintenance Using Random Forests in Manufacturing Systems." *Journal of Industrial Engineering and Management*, 14(3), pp. 123-136.
2. Wang, H., and Liu, Y. (2020). "Support Vector Machines for Fault Prediction in Industrial Systems." *IEEE Transactions on Industrial Informatics*, 16(2), pp. 987-995.
3. Zhang, X., et al. (2019). "Deep Learning Models for Predictive Maintenance: An Industrial Case Study." *Neural Computing and Applications*, 31(9), pp. 4165-4178.
4. Chen, J., and Zhao, R. (2020). "LSTM Networks for Predictive Maintenance in Time-Series Data." *International Journal of Prognostics and Health Management*, 11(1), pp. 44-55.
5. Kim, J., et al. (2018). "Application of K-Nearest Neighbors for Predictive Maintenance in Hydraulic Systems." *Expert Systems with Applications*, 93, pp. 276-285.
6. Martínez, L., and González, J. (2019). "Naive Bayes Classifier for Predictive Maintenance in Automotive Systems." *Mechanical Systems and Signal Processing*, 120, pp. 432-444.
7. Patel, A., et al. (2020). "A Hybrid CNN-RNN Model for Predictive Maintenance." *Computers & Industrial Engineering*, 143, 106403.
8. Smith, P., and Jones, D. (2021). "Integrating Random Forests and Gradient Boosting for Enhanced Predictive Maintenance." *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics*, pp. 299-308.
9. Gupta, R., et al. (2020). "Decision Trees for Predictive Maintenance." *Journal of Manufacturing Systems*, 54, pp. 25-35.
10. Johnson, T., and Miller, A. (2019). "Logistic Regression for Predictive Maintenance in Aerospace." *Aerospace Science and Technology*, 85, pp. 108-115.
11. Brown, L., et al. (2021). "Gradient Boosting Machines for Fault Prediction in Power Plants." *Energy Systems*, 12(4), pp. 789-805.
12. White, R., et al. (2019). "Artificial Neural Networks for Predictive Maintenance in the Oil and Gas Industry." *Journal of Petroleum Science and Engineering*, 179, pp. 884-895.
13. Wang, Q., et al. (2022). "Comparative Study of Machine Learning Techniques for Predictive Maintenance in Renewable Energy Systems." *Renewable Energy*, 150, pp. 723-735.
14. Liu, X., and Zhang, Y. (2021). "Time-Series Analysis for Predictive Maintenance in Smart Grids: A Case Study." *Electric Power Systems Research*, 198, 106899.
15. Garcia, M., et al. (2020). "Ensemble Learning Approaches for Predictive Maintenance Optimization in Manufacturing." *Journal of Manufacturing Processes*, 48, pp. 257-268.
16. Chen, H., and Wang, L. (2022). "Deep Reinforcement Learning for Predictive Maintenance Scheduling in Industrial Systems." *Robotics and Computer-Integrated Manufacturing*, 74, pp. 102179.



17. Kim, S., et al. (2021). "Anomaly Detection Techniques for Predictive Maintenance in Wind Turbines." *Renewable and Sustainable Energy Reviews*, 135, 110239.