



Integrating Data Simulation, Machine Learning, and Mathematical Modeling for Predicting Critical Operational Parameters and Evaluating Overall Equipment Effectiveness in Food Production

C. I. Nwoke¹, D. Ikhuele² and O. Diemuodeke³

¹Department of Mechanical Engineering, Imo State University Owerri, Nigeria.

^{2,3}Department of Mechanical Engineering, University of Port Harcourt, Choba, Nigeria.

*Corresponding author email: nwokecollinzo1987@gmail.com

Abstract	Article History
<p>In response to various challenges faced by the food Industry, such as equipment underutilization, high-energy consumption, maintenance inefficiencies, product quality variation, waste production, and inefficient data analytics, which in aggregate represent crucial impact on productivity and operational cost. This study formulated an AI-driven optimization solution for the assessment and optimization of overall equipment effectiveness (OEE) values of machines in the food Industry for enhanced operational efficiency, productivity and profitability. The study investigated real-time sensor and machine setting data, collected from the production environment to simulate the operational conditions of the production line. A Random Forest Regressors were employed to predict main operational parameters of the production system such as downtime, defect rate, and cycle time. After being trained on a portion of the simulated data, these predictions were then utilized to compute individual components of OEE: Availability, Performance, and Quality rates and ultimately determining the overall effectiveness of the production line. The models were tested using unseen samples. A mean square error method was applied to evaluate the accuracy of the model's predictions. The results showed that the true and expected overall equipment effectiveness (OEE) values were similar, ranging from 0.59% to 0.84%, which is relatively low compared to world-class target 85%. This result demonstrates that with the application of data-driven method to enhance the three metrics of OEE can lead to improved system operational efficiency. Additionally, the strong correlation between the actual and predicted scores signifies that the model is capable of accurately evaluating equipment effectiveness by taking multiple factors into account. The results of this article revealed that the downtime, defect rate and cycle time model performed perfectly well, while the validation results showed that the implemented AI model can effectively be explored in Industrial applications. Furthermore, the results of this study will empower maintenance engineers with the tools and insights necessary to enhance operational excellence. By enabling early identification and elimination of potential events, fostering predictive maintenance, supporting continuous improvement and enhancing collaboration within the food and other manufacturing domain.</p> <p>Keywords: OEE, prediction, mathematical modeling, machine learning, machines, productivity.</p>	<p>Received: 05 Jun 2025 Accepted: 20 Jun 2025 Published: 28 Jun 2025</p>  <p>Scan the QR code to view*</p> <p>License: CC BY 4.0*</p>  <p>Open Access article</p>
<p>How to cite this paper: Nwoke, C. I., Ikhuele, D., & Diemuodeke, O. (2025). Integrating Data Simulation, Machine Learning, and Mathematical Modeling for Predicting Critical Operational Parameters and Evaluating Overall Equipment Effectiveness in Food Production. <i>IPS Journal of Engineering and Technology</i>, 1(2), 111–118. https://doi.org/10.54117/ijet.v1i2.19</p>	

Introduction

In order to meet evolving consumers' expectations, maintain a distinctive competitive edge and to address multiple production issues, in the era of Industrial digital globalization, the food Industry must develop and implement advanced data-driven strategies for performance evaluation and optimization to enhance their industrial processes, while maintaining global

manufacturing standards. Zineb et al. [1] emphasized that Industries need to focus on optimizing their operations and improving efficiency by developing various evaluation and improvement models to analyze efficiency across different operational aspects, aligning with technological advancement and Industrial practices to stay competitive and address changing customer needs. Bangla et al. [2] revealed that

manufacturers are increasingly exploring performance improvement methods particularly in complex manufacturing systems as a result of fiercer competition and the need for optimal productivity. Dobra and Jovani [3] asserted that production systems generate a significant volume of data, necessitating the use of advanced techniques for recording, storage, and processing. Over the years, the Food processing Industries, has encountered various confronts stemming from significant equipment downtime, high power consumption, increased waste production, and maintenance costs. Furthermore, they have had difficulty in effectively leveraging the exponential amount of data generated by their production processes. These has led to excessive waste generation, significant investments and ultimately, low productivity and competitiveness. In this context overall equipment effectiveness (OEE) has been identified as one of the most crucial industrial advancement tools explored to assess and improve the operational efficiency of manufacturing processes. According to Corrales [4], this approach provides maintenance managers with the opportunity to identify the root cause of equipment breakdowns and their contributing factors, enabling targeted improvements in specific areas. OEE takes into account three key factors: availability, performance, and quality rates. The product of these factors yields a single value that represents the overall OEE of any production process. This study aims to enhance operational efficiency in food processing units by accurately evaluating and predicting overall equipment effectiveness (OEE) values. To achieve this, the study investigates the application of data simulation, machine learning and mathematical modeling. Data simulation serves as a foundational tool in this integration, allowing organizations to create virtual models of their production processes. These simulations can replicate real-world scenarios, enabling engineers and managers to test various operational strategies without the risk and cost of physical trials. Machine learning on the other hand, brings a powerful analytical dimension to the study. By utilizing algorithms that can learn from historical data, ML uncover patterns and correlations that may not be immediately apparent through traditional analysis. Mathematical modeling compliments these technologies by providing a structured framework for understanding complex systems. As it represents the relationships between various operational parameters in the study. This integrated approach will serve as a decision support framework for industries seeking to enhance operational performance of their production systems and maintenance practices.

Literature Review

It is important to recognize that failing to reach the ideal overall equipment effectiveness (OEE) level can drain corporate profits and competitiveness. This situation often results in increased maintenance and energy costs, a reduced mean time between repair, production fluctuations and an extended mean repair time, all of which contribute to impaired productivity. Based on this, different performance improvement and optimization techniques have been employed to ascertain highest levels of overall equipment effectiveness (OEE) for different manufacturing processes. Omeji et al. [5] improved the monthly estimated overall equipment effectiveness (OEE) value for a refinery centrifugal

gas compressor system using a Lingo optimization algorithm. The algorithm results revealed losses related to availability and performance as the major contributors to the pump's decreased OEE value. However, a critical attention in terms of effective maintenance was suggested from the generated result for the two metrics. Soul et al. [6] modeled and optimized the OEE of a deep-water disposal pump system using artificial neural network and genetic algorithm (ANN-GA) analysis approach. Authors disclosed that the developed models and analysis can serve as a decision tool for managers to minimize system breakdown, reduce maintenance expenses and increase productivity. Hassani et al. [7] utilized various machine learning algorithms such as support vector machine (SVM), Random Forest (R/F), XGBoost and deep learning to predict the overall equipment effectiveness (OEE) of a production line. Their results indicated that deep learning and Random Forest demonstrated the highest accuracy and overall performance in predicting OEE values. A hybrid analysis approach was adopted by Peter and Janos [8] for optimizing the production process of an automotive Industry, resulting to a significant increase in OEE. Nguyen et al. [9] created machine learning models to forecast OEE in an assembly manufacturing plant, attaining 99% accuracy during training and 90% during testing with the use of Extreme Gradient Boosting, thereby improving productivity and decision-making processes. Juliane et al. [10] introduced a digital data modeling approach for calculating the overall equipment effectiveness (OEE) of machines in the manufacturing sector. Their research aimed to minimize the uncertainties and time-consuming associated with manual data gathering and analysis, which have a direct impact on decision-making processes. In order to provide valuable insights into equipment performance and patterns for process improvement, Souza et al. [11] investigated both supervised and unsupervised machine learning models to analyze OEE on a production line, utilizing historical data. In an extensive study carries out by Rahmat et al. [12] presented how Industries can explore the capabilities of OEE to pinpoint areas for improvement and implement focused strategies to boost process efficacy. They also highlighted how OEE can unveil concealed losses, minimize downtime, and increase product quality, potentially resulting in greater productivity and profitability. To ensure a comprehensive evaluation and optimization of overall manufacturing process, it is important to conduct the assessment proficiently in line with the current Industrial standards and utilizing specific framework. Based on this, Ruly and Amelia [13] implemented an automatic OEE data collection and notification system in a food Industry. The reason for designing the system was to determine in real-time the operational conditions of the production settings and gather large amount of data to accurately calculate OEE, while reducing reliance on human input. Alberuni et al. [14] improved overall equipment effectiveness (OEE) in a manufacturing system through visual stream mapping (VSM) by identifying the root causes of equipment inefficiencies and addressing them during the design stage.

Methodology

The methodology applied in the study involves collecting data to simulate the operational conditions of a production process, training machine learning models to predict critical

parameters, and calculating the Overall Equipment Effectiveness (OEE). This approach integrates data simulation, machine learning, and mathematical modeling to evaluate equipment performance. Python (2023) programming software was utilized for data analysis and visualization.

Data Simulation

The first step is simulating realistic data for machine and process parameters. This data includes downtime, defect rate, raw material quality, machine settings, sensor readings, and cycle time. These variables serve as input and target values for predictive models.

Collected Data and Equations used in the study

1. Downtime Prediction Inputs: Sensor data and machine settings are used as predictors for downtime.

$$X_{downtime} = \begin{bmatrix} \text{Sensor}_{data} \\ \text{Machine}_{settings} \end{bmatrix} \tag{3.1}$$

2. Defect Rate Prediction Inputs: Raw material quality and machine settings are combined to simulate defect rates.

$$\text{Defect Rate} = 1 - \left(0.9 * \text{raw}_{material_{quality}} + 0.1 * \text{machine}_{settings} \right) + \epsilon \tag{3.2}$$

Where ϵ is random noise added for variability?

3. Cycle Time Prediction Inputs: Sensor data and machine settings are predictors for cycle time.

$$X_{cycle_time} = \begin{bmatrix} \text{Sensor}_{data} \\ \text{Machine}_{settings} \end{bmatrix} \tag{3.3}$$

4. Planned Production Time:

Planned Production Time = Random uniform value between 400 and 500 minutes.

Machine Learning Models

Random Forest Regressor's are employed to forecast downtime, defect rate, and cycle time. After being trained on a portion of the simulated data, these models are tested using

unseen samples. To train models, one must divide data into training and testing sets, match the training data with Random Forests, and make predictions on the test data.

Key Metrics for Validation

The MSE metric is used to determine the precision of predictions.

$$MSE = \frac{1}{n} \times \sum (y_{true} - y_{pred})^2 \tag{3.4}$$

where

n = number of fitted points

y_{true} = actual value

y_{pred} = predicted value [17]

Calculation of OEE Metrics

After obtaining receiving forecasts regarding downtime, defect rate, and cycle time the various components of OEE were computed:

- (i) Availability: The availability of equipment is determined by the proportion of planned production time. It's calculated as:

$$\text{Availability} = \frac{\text{Planned Production Time} - \text{Predicted Downtime}}{\text{Planned Production Time}} \tag{3.5}$$

- (ii) Performance: Performance evaluates the efficiency of the production process relative to the ideal cycle time. It is calculated using:

$$\text{Performance} = \frac{1}{\text{Predicted Cycle time} * \frac{\text{Predicted Total Units}}{\text{Planned Production Time}}} \tag{3.6}$$

- (iii) Quality: Quality measures the proportion of defect-free units produced. It is computed as:

$$\text{Quality} = 1 - \text{Predicted Defect Rate} \tag{3.7}$$

Overall Equipment Effectiveness (OEE): OEE combines availability, performance, and quality into a single metric. The input data used in this model can be seen in Table 3.1.

Table 1. Collected and input data for the study

Parameter	Description	Range/Values	Units/Scale
Planned Production Time	Total planned time for production	400–500	Minutes
Downtime Data	Predicted unproductive time due to delays	10–50	Minutes
Raw Material Quality	Measure of the quality of raw materials	0.8–1.0	Scale (0–1)
Machine Settings	Settings or configurations of machines	0.5–1.0	Scale (0–1)
Sensor Data	Sensor readings indicating operational conditions	0.3–0.9	Scale (0–1)
Cycle Time Data	Time taken to produce one unit	0.45–0.6	Minutes per unit
Defect Rate	Proportion of defective products	Simulated (based on input)	Scale (0–1)
Predicted Total Units	Number of units expected to be produced	800–1000	Units

Notes:

- (i) Defect rate calculation was derived from the combination of raw material quality and machine settings, with added noise for realism.
- (ii) Planned production time is a uniformly distributed random values representing the scheduled time for production operations.

Results and Discussion

Overall Equipment Effectiveness assessment of machines from the food processing industry was done with an AI-driven optimization methodology, an attempt is to be made for predicting major parameters like downtime, defect rates, and cycle time, which can then be used as forecasts for the individual components in evaluating OEE: Availability, Performance, and Quality. These, in turn, are used to determine the overall effectiveness of the production line. Results in this work were derived through machine learning models, which have been trained on realistic data reflecting real scenarios of operation at a food processing plant.

Downtime, defect rate, and cycle time predictive models yielded a robust performance, and their output was therefore used to compute availability, performance, and quality rates. This discussion highlights the trends, insights, and implications that can be derived from the results. This gives an overview of how efficient the production line is and the opportunities for improvement. Results are summarized in the Table 2, below for the first five samples of the test data showing availability, performance, and quality and OEE values. The corresponding graph,

Figure 1, represents the trend of availability, performance, and quality across the test samples. In Figure 1, each component

shows a different kind of trend in variation, reflecting their sensitivities due to operational and predictive parameters.

Validation of the Results from the metrics evaluation in food processing Industry

On applying the performance of the AI-driven optimization methodology to assess OEE measures, in the effort to validate the obtained results, the values are compared with those of ground truth values. The values which are used as validation measures are Available (Avail), performance (Perf), quality (Qual) and overall OEE. In comparison with the performance of the AI model in these parameters, it is shown that it can effectively be used in industrial applications.

OEE metrics were computed through the use of real-time and historical data training conducted on the AI model to forecast downtime, defect rates, and cycle times. Tabulating the validation results allows for a direct comparison between true and predicted values and to indicate both trends and deviations. The validation process was conducted across five samples as shown in Table 3, below, which illustrates how the true and predicted values for availability, performance, quality, OEE are compared across these five sets. These findings are also illustrated graphically as shown in Figure 2, including true and predicted values for availability, performance, quality, and OEE metrics across the five samples.

Table 2. Samples from Test Data, Showing Availability, Performance, Quality, and OEE Values

Sample	Availability	Performance	Quality	OEE
1	0.95	0.81	0.86	0.66
2	0.92	0.89	0.89	0.73
3	0.94	1.01	0.82	0.78
4	0.92	0.96	0.95	0.84
5	0.91	0.82	0.79	0.59

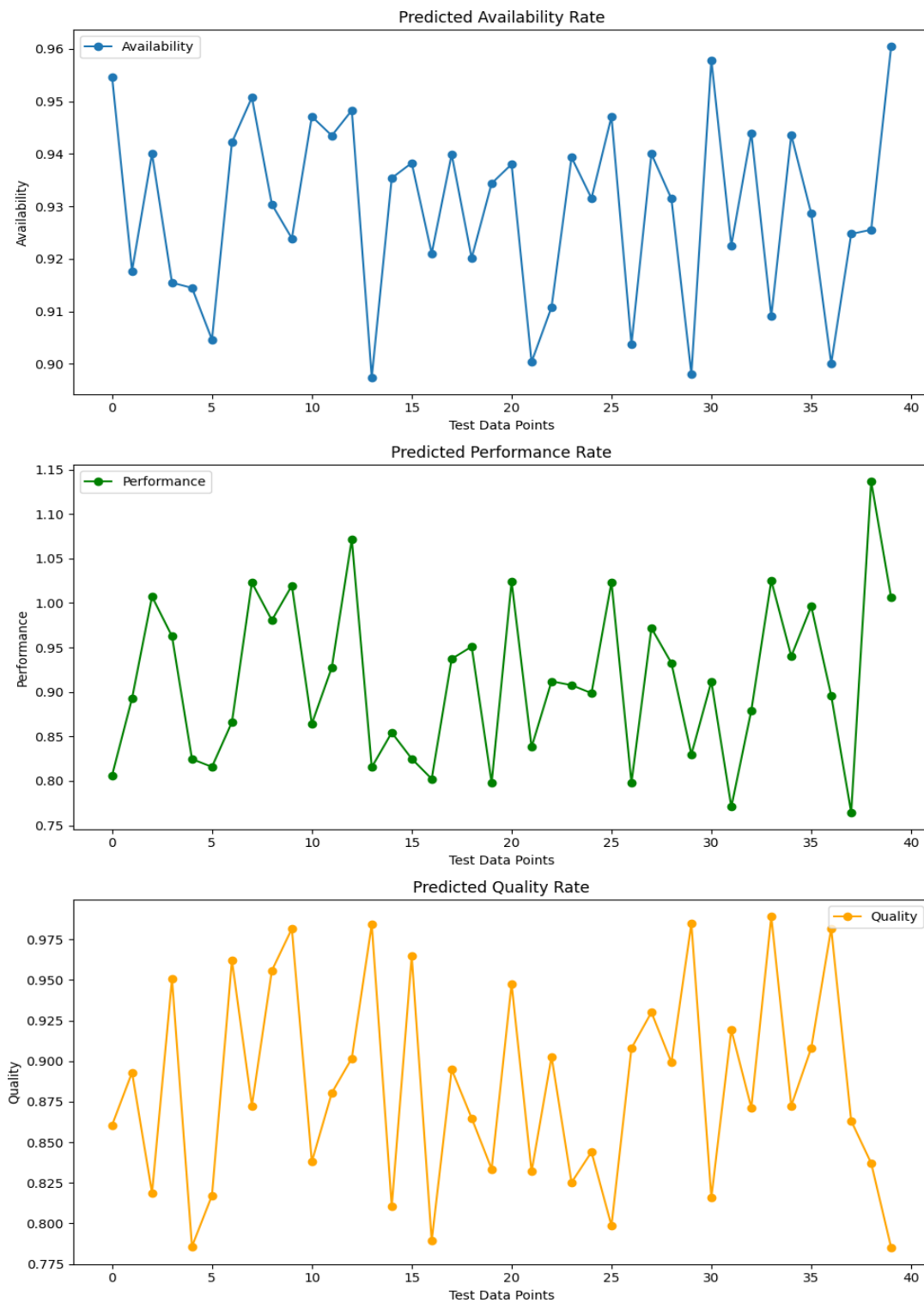


Figure 1: Illustration of the variations in availability, performance, and quality across the test samples

Table 3. Results from the validation Approach

Sample	T_Avail	P_Avail	T_Perf	P_Perf	T_Qual	P_Qual	T_OEE	P_OEE
1	0.95	0.95	0.81	0.81	0.86	0.86	0.66	0.66
2	0.92	0.92	0.89	0.89	0.89	0.89	0.73	0.73
3	0.94	0.94	1.01	1.01	0.82	0.82	0.78	0.78
4	0.92	0.92	0.96	0.96	0.95	0.95	0.84	0.84
5	0.91	0.91	0.82	0.82	0.79	0.79	0.59	0.59

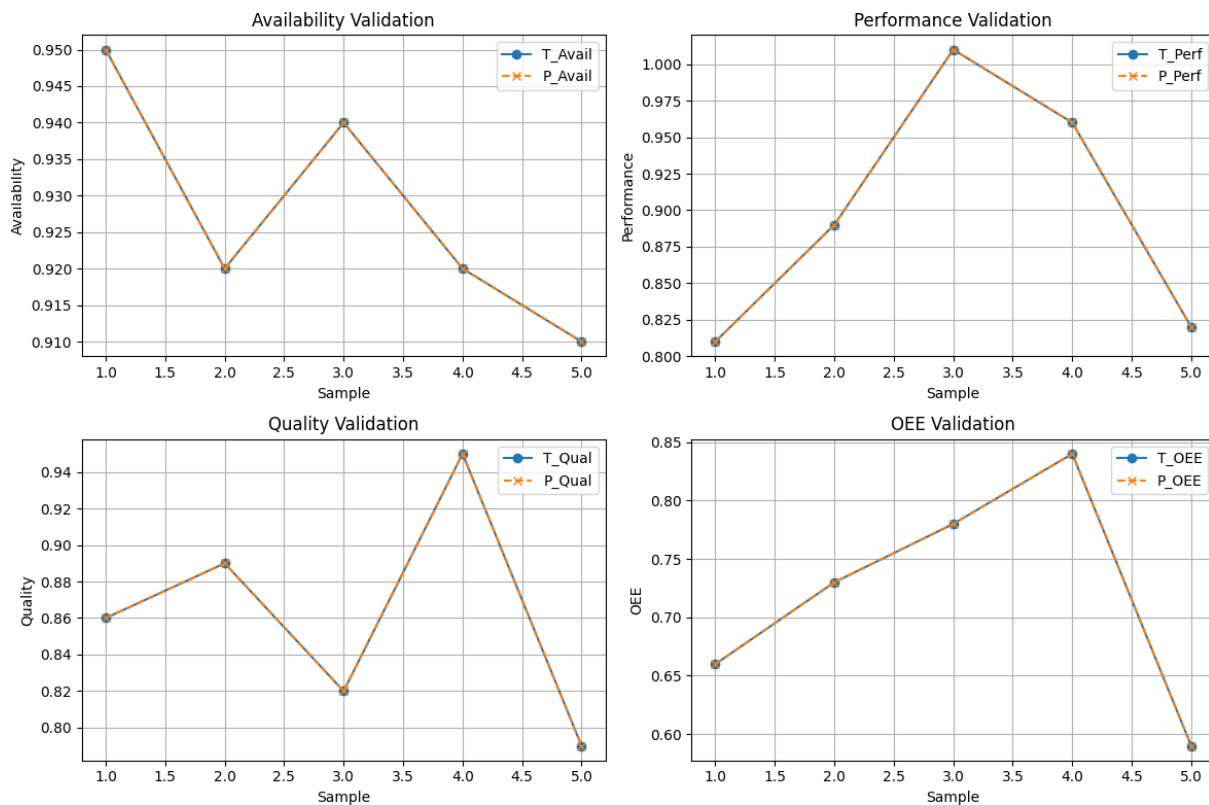


Figure 2: Validation results from OEE evaluation

Discussion of the Result from OEE Metrics Evaluation in Food Processing Industries

The results have shown some important trends in the data series and have provided comprehensive knowledge about production line efficiency.

Availability is the ratio of the time a machine is available compared to its planned production time. From these availability values for the samples analyzed, the range goes up to 0.91 and 0.95. These values indicate that the machines have very minimal downtimes relative to the times of their planned production. It is most likely related to the good performance of the predictive model-from sensor data and machine settings-identifying real events of downtime. High availability scores suggest good maintenance practices and timely responses to operational interruptions.

However, the differences in availability across samples mean that the machines are not as reliable. For example, sample one has an availability of 0.95, which is higher than sample five at 0.91. This could be due to specific operational issues, such as wear and tear on machinery or variation in schedules for maintenance. Such challenges may be addressed in an effort to further optimize the predictive maintenance schedule in order to improve overall availability.

Performance rate refers to the measure of efficiency of a machine concerning its designed production speed. The performance rates are between ranges of 0.81 to 1.01 for all samples. Sample three has a performance rate of 1.01, which shows that the machine is working at an efficient rate under ideal circumstances. This can be contributed by the setting of ideal machine variables and ideal operating conditions.

For example, sample five only has a performance rate of 0.82, meaning that its production speed is not efficient. These could be because of improper machine settings or variability in handling and processing conditions of materials. The predictive model in cycle time presents a sure basis for identifying such inefficiencies and implementing targeted interventions.

These differences in performances thus justify the use of dynamic optimization for machine operations. Application of real-time data to optimize machine parameters and bottleneck resolution will help achieve a better performance rate along the production line.

The quality rate simply gives an indication of the percentage of units produced without defects. The range of the quality rate is between 0.79 and 0.95 in the result. Sample four, with the best quality rate of 0.95, indicates high raw material quality and optimized settings of the machine. This result underlines the important role that the input quality plays in process control toward high-quality output.

Sample five with a quality rate of 0.79, points to challenges in maintaining consistent product quality, the variability could either be in properties of raw materials or in operational conditions. Most of these variations are well represented by the predictive defect rate model, hence providing a better insight that is actionable in improving the quality. Their consideration can lead to an improvement in consistency for the product quality at the production line.

The quality rate is one of the most important indicators that determine customer satisfaction and efficiency in production.

Reducing defect rates not only improves product quality but also reduces unnecessary waste and rework costs. These results emphasize the need for constant monitoring and adjustment of quality-related parameters to achieve optimum performance. OEE combines availability, performance, and quality into one metric representing the whole efficiency of the production system. Consequently, the calculated values for OEE range from 0.59 to 0.84. Sample four has, with 0.84, the highest value for OEE, driven by a very good value for all three components. This result shows that with a balanced approach to optimizing availability, performance, and quality, high efficiency is achievable.

Sample five, with an OEE of 0.59, has considerable scope for improvement. The relatively low value of OEE is due to its suboptimal performance and quality rates, while the availability rate is relatively high. This again points to the fact that for effective production optimization, all three elements must be considered together.

Since the samples are not consistent in OEE, the influence of a complex interaction among operational factors and machine conditions is apparent. Using predictive analytics and machine learning helps the production line locate which area to focus on and intervene, targeting improvements in overall efficiency.

Discussion of the Validation Results from OEE Evaluation

As seen in the Table 3 and Figure 2, there's a distinct separation between the actual and predicted values for availability, performance, quality, and OEE across all five samples. This alignment demonstrates the precision of the AI-driven approach to capture operational dynamics and predict equipment effectiveness.

Availability is the duration of time required for equipment to function. Both the actual and predicted availability values are equivalent, with a range of 0.91 for sample 2 and 0.95 for example, when compared to sample 1. This consistency is based on real-time monitoring and historical data, which accurately predicts outages using the model. The model's measurement of uptime for the equipment is demonstrated in sample 4, where T-Avail and P-Above are both 0.92.

Conclusion

In today's industrial landscape, the application of AI-based optimization methodology proves to be a fundamental framework for the food and other manufacturers in predicting critical operational parameters and evaluating overall equipment effectiveness (OEE) in production processes. This novel approach enables manufacturers to uncover and predict patterns that may not be immediately apparent through traditional analysis, allowing for timely interventions that minimize downtime and maintenance costs. Furthermore, it provides evolutionary step for firms seeking to maximize their operational efficiency with minimal resources and to drive continuous improvement objectives, leading to enhanced productivity and competitiveness in the global market.

Equipment's efficiency during its uptime, including cycle times and throughput rates, is the determining factor of performance. Each sample contains a consistent true and predicted performance value.' Both T-Perf and P-P-F are 0.89 in sample 2 but peak efficiency is 1.01 when tested the same way in sample 3 experiment. The model can accurately predict variations in operational speeds and throughput, which are influenced by factors such as operator efficiency and machine settings, due to this agreement.

The quantity of output on the other hand is determined by assessing whether it meets the required standards.' The accuracy of the model's prediction for quality rates is demonstrated in Sample 4, with T-Qual and P-Qual being 0.95. In sample 5, the values of T-Qual and P-Qual are 0.79, which implies lower output quality. Not only does it measure historical defect rates, but the model also provides real-time quality metrics. The importance of consistent quality and customer satisfaction can be maintained in an industry where product quality is a top priority through precise forecasting.

Equipment availability, performance and quality are measured by a single unit, OEE. True and predicted OEE values are in agreement with each sample. Considering availability, performance, and quality, the OEE in sample 4 is at its highest point, with T and P being 0.84 and 0.85 respectively. Nevertheless, OEE in sample 5 is the lowest at 0.59, leading to lower values across all contributing metrics. However, this is not true for only sample 2. There is a strong correlation between actual and expected OEE values, which means that the model can accurately evaluate the effectiveness of equipment by considering multiple factors. The validation outcomes indicate several crucial points and this makes the model highly reliable as there is no distinction between true and predicted values. Despite variations in operation, the results are consistent across all measures due to their adaptability under different operating conditions, such as changing cycle times and defect rates. The model's ability to predict availability enables proactive maintenance planning, which leads to reduced unplanned downtime and improved productivity. Secondly, the alignment of OEE values supports the model's ability to convey practical recommendations for boosting production efficiency.

Contribution to Knowledge

The uniqueness of this study is the integration of data simulation, machine learning and mathematical modelling in OEE assessment methodologies to obtain a dynamic and predictive assessment model. In the light of conventional methods of OEE calculation, which are normally static and involve past values, this methodology leads to a forward-looking view in real-time by predicting main operational parameters. This allows for inefficiencies to be targeted and identified ahead of time a uniqueness compared to the traditional approach.

This work has contributed toward manufacturing optimization in the following ways: machine learning models predicting downtime, defect rates, and cycle times with high accuracy predictive power that increases the reliability and timeliness of OEE assessments. Second, this methodology provides a

