IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14833

IOT Based Quality Control And Classification Analysis in Manufacturing: An Insights from Production Line Movements

ASWINI C[1], SUGUNA T[2], MALAVIKA R[3], GLADSON OLIVER S[4]

Information Technology, Government College of Technology, Coimbatore, India[1]

Information Technology, Government College of Technology, Coimbatore, India[2]

Information Technology, Government College of Technology, Coimbatore, India[3]

Information Technology, Government College of Technology, Coimbatore, India[4]

Abstract: An Internet of Things (IoT)-driven quality control mechanism within the manufacturing sector commences with the visualization of item movements throughout production lines, aimed at identifying critical stations and pathways that have a substantial impact on product quality. This preliminary phase yields valuable insights into the production flow and highlights potential bottlenecks or opportunities for enhancement. Feature Engineering is employed to derive pertinent information through the selection and transformation of features, thereby augmenting the efficacy of machine learning models. The performance of the model is evaluated in comparison to other classification methodologies, such as Support Vector Machine, Naive Bayes, Random Forest, and Gradient Boosting, predicated on the chosen features. Through the examination of the interrelations among features, stations, lines, and the response variable, a deeper comprehension of the most influential factors affecting product quality and defect occurrence is attained. By leveraging visualizations of production line movements, feature importance rankings, and classifier performance metrics, this IoT-driven framework furnishes actionable insights for manufacturers to enhance product quality and mitigate defects.

Keywords: Internet of Things, Sensors, Machine Learning, Classification, Manufacturing.

I. INTRODUCTION

In the context of IoT-based quality control and classification analysis in manufacturing, several key performance indicators (KPIs) are essential for evaluating operational effectiveness. Overall Equipment Effectiveness (OEE) serves as a foundational metric, measuring the efficiency of manufacturing processes by assessing availability, performance, and quality of equipment. This metric is crucial for identifying losses that hinder equipment efficiency, thereby facilitating improvements in production output. Labor efficiency is another critical KPI, as it evaluates how effectively labor resources are utilized within the manufacturing process. This metric can be enhanced through IoT technologies that provide real-time data, enabling better workforce management and productivity.

Cost reduction is also a vital KPI, reflecting the decrease in operational costs achieved through improved processes and efficiency. Finally, quality improvement metrics are essential, as they assess enhancements in product quality, which directly correlate with customer satisfaction and reduced defects. Together, these KPIs provide a comprehensive framework for assessing the impact of IoT on quality control and classification in manufacturing, driving continuous improvement and operational excellence.

In the rapidly evolving domain of manufacturing, the assurance of product quality is of utmost importance. Quality control mechanisms are integral in detecting and rectifying defects prior to the distribution of products to consumers. Through the examination of production data and the formulation of predictive models, manufacturers are able to preemptively identify potential challenges and implement corrective measures. Within the context of Industry 4.0, an awareness of sporadic manufacturing is essential, as it has the potential to jeopardize the quality of the processes involved. By accumulating data that facilitates enhancements in product and process service quality, various challenges can be effectively addressed (Illes et al. 2017). Although intelligent sensors are capable of recording and transmitting data, no substantial added value is derived from this unless the captured data is employed to inform decisions aimed at process improvement (Godina et al. 2018).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14833

Tercan (2022) posits that the predictive capabilities allow manufacturing enterprises to formulate data-driven assessments regarding product quality predicated upon process data, employing sensor data to facilitate the automation of quality inspections based on measurement data. The identification of infrequent quality occurrences has emerged as a matter of critical significance and presents an opportunity for manufacturing firms to elevate quality benchmarks. The process of defect identification is executed utilizing an l1-regularized logistic regression and elimination algorithm (Escobar & Morales 2018). A classification model was constructed, resulting in the development of a quality control system tailored for manufacturing. Images are acquired via mobile devices to identify defects on production items, followed by the extraction of features from these images through the application of a convolutional neural network (San et al. 2020).

Visual representations of production line dynamics facilitate the identification of pivotal stations and pathways that significantly influence product quality. The preliminary phase of data analysis is characterized by data exploration, wherein the objective is to comprehend the attributes of the dataset. This process entails a meticulous examination of the data to uncover patterns, irregularities, and interrelationships among variables. Methodologies such as summary statistics, data visualization, and clustering techniques are employed for the exploration of the dataset. Data exploration is instrumental in deriving insights from the data, pinpointing potential challenges, and informing subsequent analytical endeavors.

(Poler et al. 2021) introduced a comprehensive framework that assured sufficient levels of data accuracy, precision, and traceability, alongside data reliability, through the integration of sensing, communication, computational infrastructure, storage, analysis, and optimization methodologies. This framework enhanced both process quality and product quality within manufacturing, culminating in a zero-defect manufacturing paradigm. The methodologies employed for predictive quality are contingent upon the specific application in question. Drawing from the analyses presented by (Koksal et al. 2011) and (Rostami et al. 2015), the functions of machine learning and deep learning in the realm of predictive quality are enumerated as follows:

i. Quality description: The systematic identification, rigorous evaluation, and comprehensive interpretation of the interrelationships between process variables and the resultant product quality.
ii. Quality prediction: The model-based estimation of a quantitative quality variable derived from process variables, employed for the purposes of decision-making support or automation.
iii. Quality classification: A model-based estimation of qualitative quality variables.

Learning algorithms, such as support vector machines, random forests, artificial neural networks, and principal component analysis, may be employed to develop the quality monitoring model while taking into account the cumulative impacts of various manufacturing stages as well as the imbalanced and dynamic characteristics inherent in manufacturing processes (Ismail et al. 2022).

Table 1 Classifier Models

Learning Model	Technique	Pros	Cons
LibSVM	Maximizes the margin between the classes, making it robust to outliers.	High accuracy and ability to handle high-dimensional data efficiently.	Tuning the hyperparameters is challenging.
Radial Basis Function Support Vector Machine	Map the input data into a higher-dimensional space, uses a radial basis function as the kernel.	Effective in capturing complex relationships in the data and provide high accuracy in classification tasks.	Sensitive to the choice of hyperparameters.
Naive Bayes	Probabilistic classifier based on Bayes' theorem.	Computationally efficient and easy to implement.	Limit the performance on complex datasets with correlated features.
Random Forest	Constructs a multitude decision tree during training and output the class that is the mode of the classes or the	Robust against overfitting and can handle large datasets with high dimensionality.	Computationally expensive and may not perform well on imbalanced datasets.

DOI: 10.17148/IJARCCE.2025.14833

	mean prediction of the individual trees.		
Gradient Boosting	Builds multiple decision trees sequentially, with each tree correcting the errors of its predecessor.	High predictive accuracy and ability to handle complex datasets.	Tuning the hyperparameters is time-consuming, also requires careful handling of missing data and categorical features.

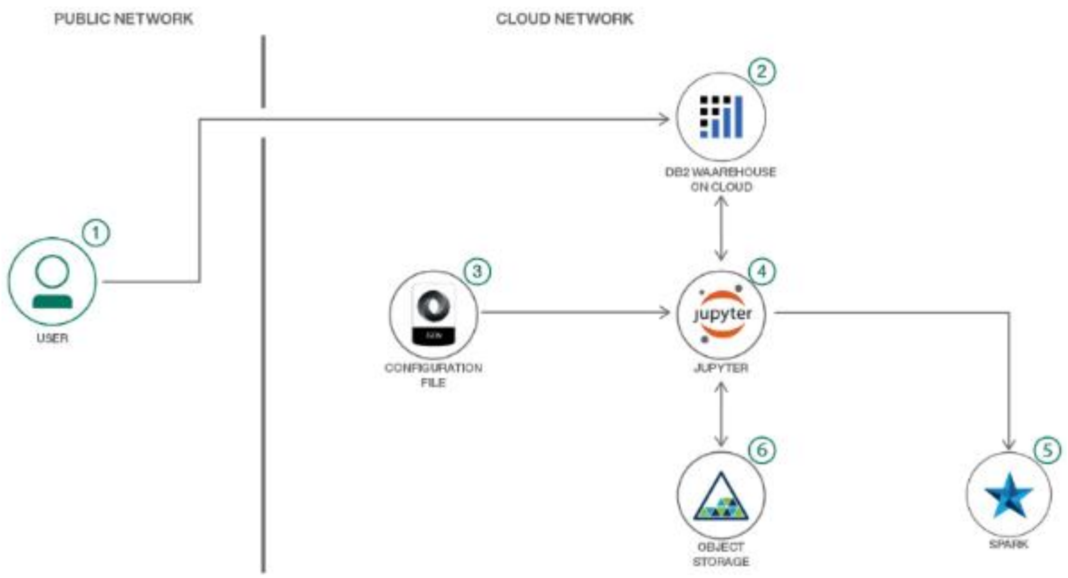
Feature engineering constitutes the systematic endeavor of generating novel features or altering pre-existing features within a dataset, with the objective of augmenting the efficacy of machine learning models. This undertaking encompasses the selection of pertinent features, the transformation of existing features, and the development of new features through methodologies such as one-hot encoding, normalization, and feature extraction. The principal goal of feature engineering is to furnish the machine learning model with more informative and discriminative features, thus enhancing its capacity to learn from and generalize across the data.

Classifier training entails the application of a machine learning algorithm to discern the relationship between the features and the target variable within the training dataset. This process necessitates the calibration of the model to the training data and the optimization of its parameters to minimize predictive error. Upon completion of the training phase, the model is subjected to evaluation using an independent testing dataset to gauge its performance. This evaluation process involves forecasting the target variable for the testing data and juxtaposing the predictions with the actual observed values.

Classifier training and testing represent iterative processes, wherein the model undergoes refinement and evaluation on multiple occasions to attain the desired level of performance. A variety of classification algorithms are scrutinized to ascertain the most effective in forecasting manufacturing failures. Table 1 encapsulates the classifiers evaluated in the development of a model that not only aims to accurately predict defective products but also seeks to elucidate the underlying factors that contribute to these failures. By facilitating this understanding, manufacturers can implement proactive strategies to enhance their production processes, mitigate waste, and improve product quality.

II. SYSTEM DESIGN

This Internet of Things (IoT) driven Quality Control mechanism encompasses several critical stages aimed at the proficient application of machine learning algorithms to enhance the accuracy of manufacturing defect predictions. The comprehensive architectural framework of this system is illustrated in Figure 1.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14833

Figure 1 Architecture of IoT Based Quality Control in Manufacturing

The dataset, which comprises numerical, categorical, and temporal features, undergoes preprocessing to address the issue of missing values and to guarantee alignment with the chosen algorithms. This process entails substituting missing values with appropriate placeholders and converting categorical variables into a format that is amenable to machine learning methodologies.
Subsequently, feature engineering methodologies are employed to identify pertinent features and to generate additional ones that augment the predictive efficacy of the models. This phase includes the selection of the most significant features, the transformation of existing features, and potentially the formulation of interaction terms to encapsulate intricate relationships present within the dataset.

The overall process flow of this system is as follows.

- Reading Internet of Things (IoT) sensor data from the database.
- A function is established to partition the dataset into training and testing subsets, construct Logistic Regression models, evaluate these models, and compute accuracy metrics such as the Confusion Matrix.
- The features and target variables are user-configurable for the purpose of predicting equipment failures, along with the training and testing datasets.
- The computation of essential statistics is conducted to facilitate the assessment of the predictive capabilities of the models.
- The experiment is replicated by modifying the configuration parameters through the re-execution of the models.

Module Description

For each task, it is imperative to identify the stations traversed and subsequently document the minimum duration spent at each. For every entry, the next station in the sequence is determined. For each identifier, it is essential to ascertain the initial node that is accessed. Before each commencement point, an edge is inserted from a common origin. For each identifier, the row is identified where there is an absence of a subsequent station, indicating the last station visited. The station is populated with the corresponding response value.

Following the preparation of the features, the dataset is divided into training and testing subsets to facilitate the training of classification algorithms. A variety of algorithms, including LibSVM, RBF SVM, Naive Bayes, Random Forest, and Gradient Boosting, are trained using the training subset and assessed on the testing subset employing metrics such as accuracy, precision, recall, and the Matthews correlation coefficient (MCC). In the concluding phase, the trained models are deployed within a production environment, enabling their application for the real-time prediction and classification of defective products. Continuous monitoring and enhancement of the models are critical to maintaining their efficacy and adaptability to evolving manufacturing environments.

Matthews Correlation Coefficient

- The Matthews correlation coefficient (MCC) serves as an indicator of the efficacy of binary classification models, particularly in the context of datasets characterized by class imbalance.
- This metric incorporates both true and false positives as well as true and false negatives, with values that span from -1 to +1.
- A MCC value of +1 signifies an impeccable prediction, a value of 0 denotes a random classification, while a value of -1 reflects a complete discordance between predicted outcomes and actual observations.

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- MCC is useful because it balances the dataset even if the classes are of very different sizes.

III. RESULTS & DISCUSSIONS

Different stations and lines exhibit a higher frequency of engagement with Faulty Items in comparison to their non-faulty counterparts. Therefore, this observation reinforces the hypothesis that certain stations play a pivotal role in the determination of product faults; these locations are characterized by a greater volume of recordings and more rigorous testing protocols aimed at evaluating manufacturing quality. The efficacy of the classification models was assessed utilizing the Matthews Correlation Coefficient (MCC) score to gauge their ability to predict manufacturing defects. The

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14833

LinearSVC model, employing a linear kernel, attained an MCC score of zero, which signifies a complete lack of predictive capability.

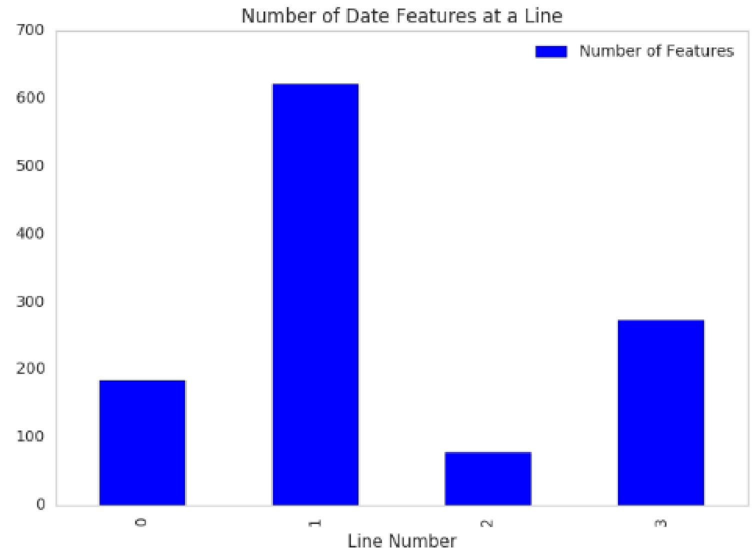


Figure 2 No. of Features in line

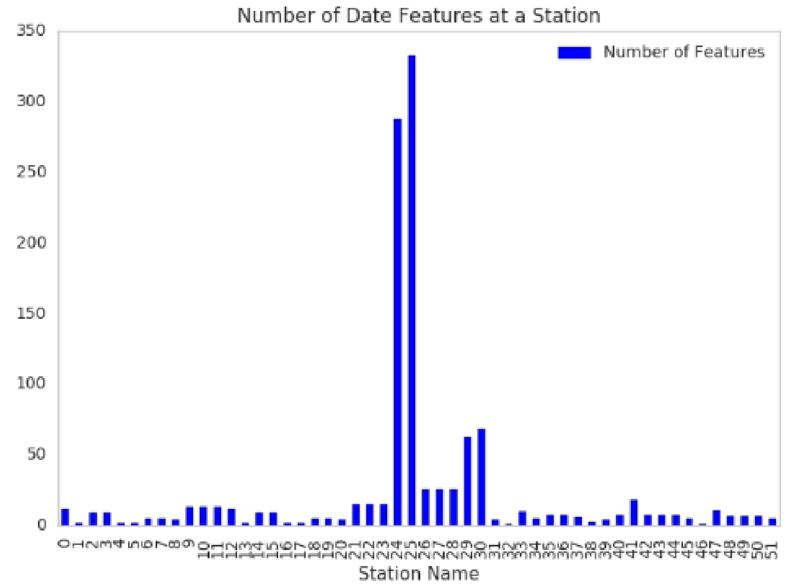


Figure 3 No. of Features at station

- Visualization of categorical features and frequency of features with respect to stations and lines. (Figure 4 & 5)
- Visualizing the first 1000 Defective IDs and first 1000 Non-Defective IDs and their movements across the production lines to develop useful insights about the data (Figure 6 & 7).
- Finding out the probabilities of seeing a defect after a defect.
- Figuring out the top most important features on which the classification algorithms depend using F-Score (Figure 8).
- Developing the Random Forest Classifier on the chosen subset of important features using Databricks Platform/Py-Spark.
- Developing different classification algorithms (LibSVM, RBF SVM, Naive Bayes, Random Forest, Gradient Boosting) on the chosen subset of important features.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 8, August 2025
DOI: 10.17148/IJARCCE.2025.14833

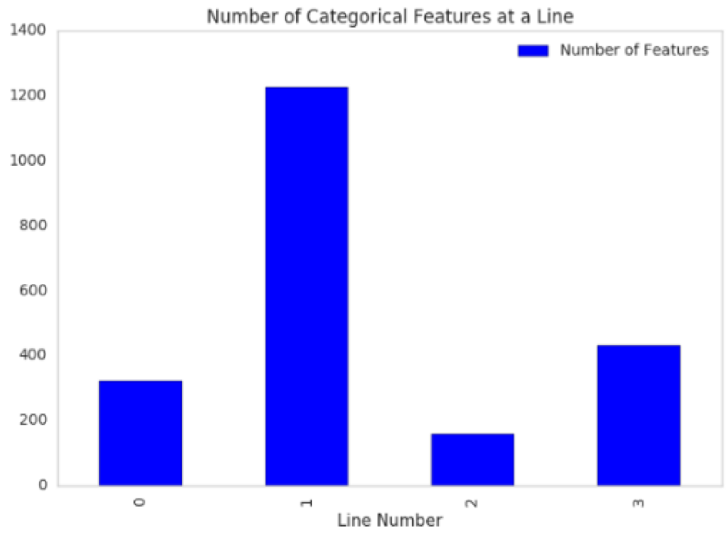


Figure 4 No. of categorical features at a line

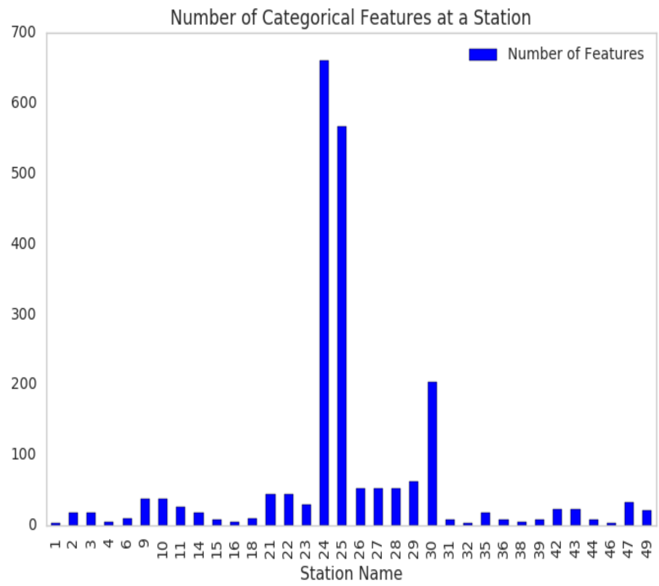


Figure 5 No. of categorical features at a station

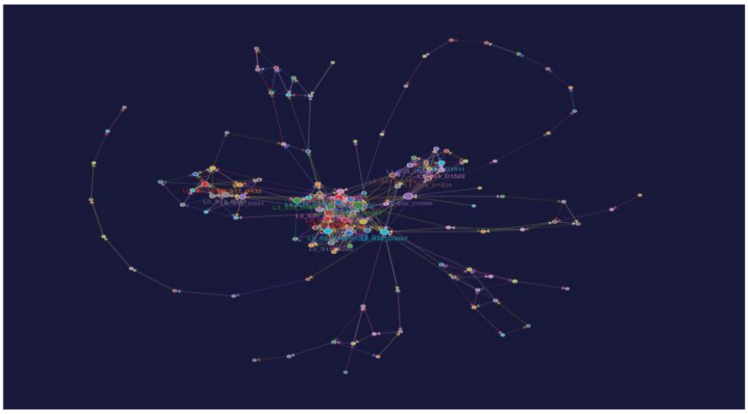


Figure 6 Visualizing the Defective & Non Defective IDs

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14833

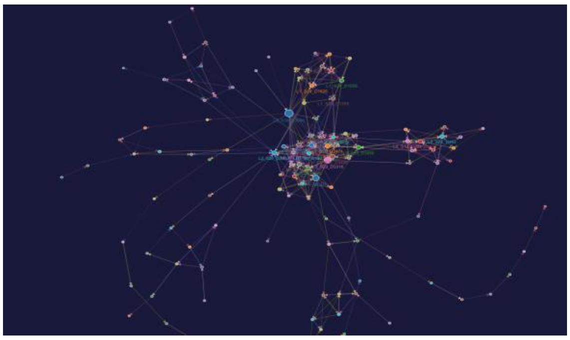


Figure 7 Visualizing the movement of Defective & Non Defective IDs

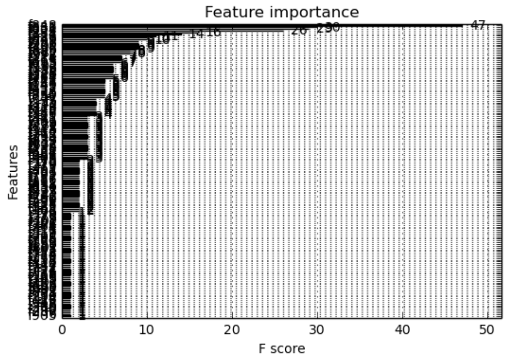


Figure 8 F-Score of Classification algorithms

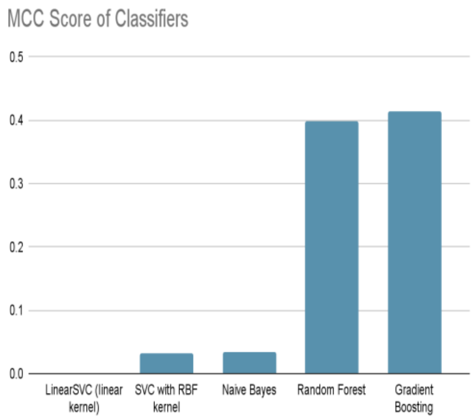


Figure 9 MCC Scores of Classifiers

The SVC with an RBF kernel and Naive Bayes models performed slightly better with MCC scores of 0.0323 and 0.0337, respectively. These models showed a limited ability to distinguish between defective and non-defective products. In contrast, the Random Forest and Gradient Boosting models demonstrated significantly higher MCC scores of 0.3977 and 0.4143 respectively. These models exhibited strong predictive capabilities and were able to effectively classify defective products. Overall, the Random Forest and Gradient Boosting models outperformed the other classifiers, highlighting their suitability for predicting manufacturing defects based on the selected subset of important features. MCC is a reliable measure for binary classification that considers all four outcomes of the classifier (Figure 9), making it particularly useful for imbalanced datasets.

IV. CONCLUSION

Through extensive data exploration, feature engineering, and the training of various classification algorithms, valuable insights from the dataset and the factors influencing product defects have been identified and visualized. The results indicate that the Random Forest and Gradient Boosting models outperform other classifiers, achieving higher MCC

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 8, August 2025
DOI: 10.17148/IJARCCE.2025.14833

scores. These models exhibit strong predictive capabilities and can effectively classify defective products based on a subset of important features. Additionally, the visualization of data movements across production lines provided valuable insights into the manufacturing process. Stations and lines with higher frequencies of defective products suggest areas that may require further investigation or improvement. IoT enabled demonstrates the potential of machine learning in improving manufacturing processes by identifying and mitigating defects. Future work could focus on further refining the models, incorporating additional features, and integrating the predictive model into the production environment for real-time defect detection.

REFERENCES

[1]. Escobar CA, Morales-Menendez R. Machine learning techniques for quality control in high conformance manufacturing environment. Advances in Mechanical Engineering. 2018;10(2).
[2]. Godina, R., Matias, J.C.O., "Quality Control in the Context of Industry 4.0", Industrial Engineering and Operations Management II. IJCIEOM Springer Proceedings in Mathematics & Statistics, Springer, vol 281, 2018.
[3]. Illés, Béla, Péter Tamás, Péter Dobos, and Róbert Skapinyecz. "New Challenges for Quality Assurance of Manufacturing Processes in Industry 4.0." Solid State Phenomena 261 (August 2017): 481–86.
[4]. Ismail, M., Mostafa, N.A. & El-assal, A. Quality monitoring in multistage manufacturing systems by using machine learning techniques. J Intell Manuf 33, 2471–2486 (2022).
[5]. Koksal, G., Batmaz, İ, & Testik, M. C., A review of data mining applications for quality improvement in manufacturing industry. Expert Systems with Applications, 38(10), 2011, pp. 13448–13467.
[6]. R. Poler et al., "An IoT-based Reliable Industrial Data Services for Manufacturing Quality Control," 2021 IEEE International Conference on Engineering, Technology and Innovation ICE/ITMC), Cardiff, United Kingdom, 2021, pp. 1-8.
[7]. Rostami, H., Dantan, J.-Y., & Homri, L., Review of data mining applications for quality assessment in manufacturing industry: Support vector machines. International Journal of Metrology and Quality Engineering, 6(4), 401. 2015.
[8]. San-Payo, G., Ferreira, J.C., Santos, P. et al. Machine learning for quality control system. J Ambient Intell Human Comput 11, 4491–4500 (2020). https://doi.org/10.1007/s12652-019-01640-4.
[9]. Tercan, H., Meisen, T. Machine learning and deep learning based predictive quality in manufacturing: a systematic review. J Intell Manuf 33, 1879–1905 (2022).