Vol. 06, No. 1 (2024) 343-352, doi: 10.24874/PES.SI.24.02.017

Proceedings on Engineering Sciences

www.pesjournal.net

PROACTIVE QUALITY EVALUATION: A NOVEL STRATEGY-ASSISTED EARLY DETECTION IN MANUFACTURING

Swati Singh[1]
Raman Batra
Keerti Rai
S. Sujai

Received 29.11.2023.
Received in revised form 16.01.2024.
Accepted 24.01.2024.
UDC – 338.518

Keywords:

Manufacturing Industry, Quality, Modified Gravitational Search Algorithm-Based Decision Tree (EGSA-DT), Z-Score Normalization, Principal Component Analysis (PCA).

A B S T R A C T

The proactive exploration and avoidance of errors or variations from quality standards during the manufacturing process is referred to as "early quality detection" in the manufacturing industry. Post-production inspection, which can be expensive and time-consuming, is used in traditional quality control systems. To overcome this, we proposed a Modified gravitational search algorithm-based decision tree (MGSA-DT) to predict the quality of manufacturing processes at an early stage. We gathered sensors data in the manufacturing industry. In order to prepare the data for principal component analysis (PCA), Z-score normalization is used. Then, the essential features are extracted from the preprocessed data. To assess the effectiveness of the suggested approach in terms of accuracy (98.4%), precision (97.6%) and recall (97.2%), respectively. Implementing early quality detection techniques in manufacturing has demonstrated encouraging outcomes in enhancing the overall quality of products and decreasing production expenses.

1. INTRODUCTION

Early detection in manufacturing is the proactive detection of possible difficulties or flaws in the production process before they escalate into more substantial problems. Employing this approach is essential for guaranteeing the quality of the product, minimizing waste and maximizing overall efficiency (Scime, L., et al., 2020). To accomplish early detection, a range of technologies and approaches are used, such as sophisticated sensors, real-time tracking structures and predictive analysis (Ren, Z., et al., 2020).

An essential factor in early detection is the use of sensor technology throughout the entire manufacturing process. The sensors gather data on many factors like temperature, pressure, motion and other essential metrics (Iqbal, R., et al., 2019). Through the examination of this up-to-date data, producers can detect deviations from the standard, indicating possible problems in the production process (Westphal, E., & Seitz, H., 2021). For instance, a rapid increase in temperature could suggest a defective apparatus or an anticipated equipment breakage.

This enables producers to apply preventative actions before the occurrence of any issues, hence decreasing

[1] Corresponding author: Swati Singh
Email: swatisingh5444@gmail.com

the amount of time that production is stopped and lowering the probability of faulty items that are released into the market (Xu, K., et al., 2020). Predictive analytics could improve maintenance plans by determining the most advantageous periods for equipment care, therefore preventing needless delays in the manufacturing process (Ayvaz, S., & Alpay, K., (2021)). Timely identification not only enhances the quality of the product but also minimizes the total expenditure of production.

Manufacturers could save expensive repairs garbage and guarantee claims by recognizing and resolving defects at an early stage. Furthermore, it optimizes the overall efficacy of the manufacturing process by reducing periods of inactivity and guaranteeing that machinery functions at its maximum capacity (Saqlain, M., et al., 2019). Early identification in production is a preventive and data-driven strategy that uses sensor technology while predictive analysis to identify and resolve possible problems before they have a negative influence on product quality or interrupt the production process (Oduoza, C. F., 2020).

Proactive quality assessment is a strategic method used by organizations to examine and guarantee the quality of goods or services at different stages of the manufacturing or service-delivering process (Shcherbakov, M. V., et al., 2019). Proactive quality assessment differs from standard quality control approaches by emphasizing the prevention of problems before they arise rather than checking final items for flaws. Proactive quality evaluation involves the implementation of rigorous quality standards, extensive risk assessments and the use of modern monitoring tools and procedures by enterprises (Nguyen, G., et al., 2020).

Proactive quality evaluation entails the incorporation of modern technologies, real-time tracking systems and thorough data collection methods to detect and resolve any quality problems in the initial phases of production. The change towards a proactive strategy not only improves the quality of products but also reduces the chances of faults spreading downstream, leading to higher efficiency, lower costs and enhanced satisfaction with the product (Stelnicki, A. M., et al., 2021). The manufacturing industry serves as the foundation of worldwide economic progress, comprising a wide array of industries that manufacture essential products for daily existence. Manufacturing encompasses a wide range of industries, including automotive, technology, healthcare and consumer goods. It requires the complex utilization of new technologies and intricate procedures to convert raw materials into completed products (Saengthong, D., 2023).

(Suman, S., & Das, A., 2019) suggested diagnostic statistic captured the associate features' contribution, associate features were those with a strong association between the primary characteristics that constitute the leading cause of the identified issue. A multi-layered or multi-strata statistical process monitoring technique has been attempted to be developed. A latent factor score-based exponential weighted moving average (EWMA) graph was the monitoring statistic utilized for the early identification of the faulting condition. (Lee, W., & Seo, K., 2021) provided two iterations of feature extraction methods that used machine learning algorithms and the closest neighbor to detect equipment failures. The methods were based on multi-stream monitoring system information. It was crucial to identify machine failure to perform the required maintenance to avoid the system from breaking down unexpectedly. (Fan, S. K. S., et al., 2020) analyzed the significance of equipment sensor SVIDs using the random forest algorithm, filtered the key SVID using k-means and integrated several machine learning techniques to validate the key SVIDs that determine critical processing times along with stages. A research was done to test the suggested data-driven model for fault identification and diagnostic purposes.

(Peres, R. S., et al., 2019) identified dimensional flaws in an actual automotive multiple-stage assembly process, a number of classification algorithms based on machine learning were developed and assessed using a variety of metrics. In complicated multistage production processes, where unknown faults could easily spread downstream, material dimensional variation plays a critical role in quality control. The line consists of two automated inspection stages separated by a number of manually controlled assembly and pre-alignment procedures. A quick surface defect detection technique for directed energy deposition (DED) was suggested (Chen, L., et al., 2021). Early detection of surface flaws in the AM process was necessary to prevent a further decrease in the component quality. The system used a combination of supervised and unsupervised machine learning approaches to identify and categorize surface flaws. An accuracy of 93.15% was obtained in surface defect detection using the verified indicated approach. (Imoto, K., et al., 2019) used a transferred learning-based Convolutional neural network to classify defects automatically. It focused on a defect analytics challenge that requires engineers to utilize the outcomes of defect categorization to determine the reasons behind productivity decrease. Using actual semiconductor manufacture information, they assessed the effectiveness of their suggested approaches by applying them to a defect identification task that included the use of an electron microscope image (Karatas, M., et al., 2022).

(Ghahramani, M., et al., 2020) used Deep Learning and Evolutionary Computing methods to make semiconductor production smart. They provided a dynamic approach for addressing a number of issues and gain important insights into semiconductor production processes. Their intention was to provide organizations access to efficient predictive technologies by offering an enhanced solution for managing production processes and gaining insight on several aspects. (Chacko, M., 2021) enhanced manufacturing effectiveness and adaptability in the context of Industry 4.0 activities, the Digital Twin-based Cyber-Physical Quality System (DT-CPQS) idea incorporates automated quality assessment, simulations and estimation of manufacturing procedures. Those decreased the inspection time and enable the operator to make necessary corrections before the resulting component experiences a quality failure. The production process would advance toward an independent quality platform for zero failure production in the future because of DT-CPQS. (Shahin, M., et al., 2023) provided updates on the investigation of over 20 defect detection models through some technologies. Manufacturers could identify possible fault conditions in operations to prevent interruptions caused by unanticipated tool usage or inadequate work piece quality by detecting prospective machine failure. The final findings showed that the gradient boosting and deep forest algorithms had extremely high average accuracy levels (over 90%) (Javaid, M., et al., 2021).

(Yu et al., 2019) provided a framework for big data that uses actual industrial big data collected from large-scale worldwide manufacturing to execute problem identification and diagnosis in advance maintenance. A collaborative firm has integrated the suggested detection method into its real-time industrial production system. Intelligent machines have been developed as a result of the digital age, transforming the industrial sector into smart manufacturing. (Ammar, M., et al., 2021) offered an in-depth examination of diverse Industry 4.0 technology designed to improve manufacturing processes and the quality of materials. They created an extensive inventory of the many technologies currently accessible, together with their distinct qualities and the advantages they offer for efficient administration in emerging industries. These technologies minimised disconnects in communication and assist in staying current with data. Utilizing Industry 4.0 technology, smart predictive modelling would effectively monitor machinery in industries. (Dalzochio, J., et al., (2020)) concentrated on systems that use logical thinking and "machine learning (ML)" to perform maintenance forecasts in the context of Industrial 4.0. The forecasting of failures incorporated the utilization of concepts such as an advanced system that predicts service outcomes as well as assists in making choices and creating plans. It was important to emphasize that maintenance prediction was a highly relevant subject in the framework of Industry 4.0. However, there were numerous obstacles that require more exploration in the domain of ML and the use of deduction.

(Orrù, P. F., et al., 2020) provided a fundamental and readily applicable machine learning (ML) algorithm for predicting faults in centrifugal pumps at an early stage in the gas and oil industry. Possible defects were identified and categorized, guaranteeing high precision in forecasting. The findings from their initial investigation demonstrated that the model identifies patterns of system departures from normal operation and generates alerts to predict faults. Their model served as a decision support system for maintenance employees, helping them prevent potential failures. (Liu, H., & Wang, L., 2020) presented a remote human-robot collaboration system that corresponds to the principles of cyber-physical systems. The implemented system has the capability to operate in four distinct modes, each suited for specific conditions. A remote robot management system and a model-based presentation system were developed by employing a teamwork robot and a manufacturing robot. The certain investigation revealed significant potential for implementing the created technology in a hazardous production environment. A proactive maintenance system based on data analysis was created (Ayvaz & Alpay 2021) for manufacturing production processes. The system's efficacy was evaluated by analyzing real-world manufacturing-related IoT data. The assessment findings demonstrated the efficacy of the forecasting maintenance system in detecting early indications of possible failures, therefore mitigating the occurrence of production interruptions. The most successful machine learning models identified in the current investigation have been included into the operational system in the factory. (Arinez et al., 2020) utilized a hierarchical organizational approach, employed in industrial facilities, their study analysed the interactions occurring at both the macroscopic systemic and the microcosmic level of individual materials entering processor flows. The study covered a diverse variety of subjects, including productivity and excellence, management control in human-robot cooperation, process tracking, diagnosis and prediction for improvements in technological advances materials for the purpose of attaining desired material qualities through simulation of processes and management. (Xu, D., et al., (2022)) developed an innovative method for predicting failures by combining a gated recurrent unit with an automatic encoder. Their approach intends to enhance the effectiveness of unbalanced learning. The development of efficient and accurate failure identification and prediction techniques was crucial for minimizing losses and an increasing number of algorithms depend on sophisticated machine learning technology. The failure predictions algorithm was implemented in an actual paper as well as fibber mill identified and predicted instances of sheet breaking during production.

The study proposes the Modified Gravitational Search Algorithm based Decision Tree (MGSA-DT) as a method for predicting the quality of manufacturing processes in their initial phases.

1.1 Contribution

- We gathered sensors data in the manufacturing.
- Z score normalization is a technique used to standardize the values of different variables to a consistent scale.
- Principal component analysis (PCA) processed the raw data into processable numerical characteristics to maintain the information from the original data set unchanged.
- Modified gravitational search algorithm based decision tree (MGSA-DT) for predicting manufacturing process quality in advance.

The remaining study components might be classified: The approaches are discussed in section 2. The experiment's findings are presented in section 3. Discussions are presented in section 4. The last section of this paper, section 5, is the conclusion.

2. METHODOLOGY

This study suggests using a modified gravitational search algorithm-based decision tree (MGSA-DT) for predicting manufacturing process quality in advance. We collected information by locating sensors and data collection devices in the manufacturing sector. Principal component analysis (PCA) is used to extract the pertinent feature from the pre-processed data after Z-score normalization. Figure 1 illustrates the general flow.

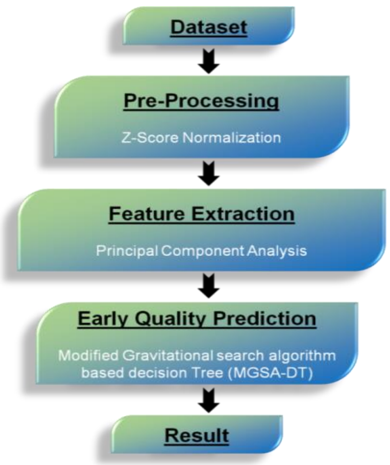

Figure 1. General flow

2.1 Dataset

We gathered sensors data in the manufacturing industry (Kao, H. A., et al., 2017). The collection contains 1567 samples, with each sample containing 590 manufacturing operation characteristics and 1 quality variable. From the entire dataset, only 104 examples indicate instances of failure. The operation variable information is obtained from a specific process control sensor in the electronic production machine and it is given an identification number corresponding to the sensor ID. In order to generate a balanced dataset with equal representation of positive and negative instances, the boosting technique is utilized. As a result, the number of unsuccessful instances in the dataset is increased to 1456. The testing and training information are partitioned from the initial dataset in a ratio of 3:1.

2.2 Data preprocessing

Z-score normalization is an essential data preprocessing step in early quality predictions for manufacturing. Through standardizing the variables, it guarantees a uniform scale throughout the dataset. Z-score normalization reduces the effect of differing feature scales by converting the information into a standard distribution with a mean of 0 and the standard deviation of 1.

2.2.1 Z-score normalization

The vector of every characteristic contained in the input data is normalized by using the mean and standard deviation of every characteristic over a sequence of learning data. We compute the mean and standard deviation for every characteristic. The following equation represents the equality utilised in the method, w' shows the normalised data, the input variablew_i, the average value of the input variableμ_i and the standard deviation of the inputs variableσ_i in Equation (1)

$$w' = \frac{w_i - \mu_i}{\sigma_i} \quad (1)$$

This approach standardizes every characteristic in the data set by setting its mean to zero and its normal deviation to one. As part of the method, the vectors of features in the information set are first subjected to normalization. The mean and standard deviation are computed for each feature using the training data and retained as weights for the final system design. Essentially, this approach is an initial processing step in the construction of an artificial neural network.

2.3 Feature extraction

Principal Component Analysis (PCA) is a crucial process for improving the efficiency and efficacy of early forecasting models for quality in the manufacturing business. It involves extracting relevant features. PCA offers an organized technique for reducing dimensionality in manufacturing datasets, which contain several variables, while preserving crucial information.

2.3.1 Principal component analysis (PCA)

PCA is a statistical method that uses data to create a model. It takes a group of variables that are related to each other and reduces them to a small set of independent additional factors. These new variables retain a significant amount of the details in the initial form. Let W be the input dataset, with an ordered collection of m-dimensional values represented by every column. Furthermore, it is important to note that every function in the collection of values has a zero average $(E(W) = 0)$. An information matrix in its original form consists of m samples and n variables, which can be represented as follows in Equations (2-5):

$$W = [w_1, w_2, \ldots, w_m]^S = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{pmatrix} \quad (2)$$

PCA allows for the conversion of environmental performance requirements and data variables inputs that preserves as much of the original data as is practical into a

new event space. Determining the instructions allows one to achieve this are concerned in the input information sets that have the highest amount of variation and projecting them into a new subspace that has the same space.
Consequently, an orthonormal conversion Y can be used in the manner that follows to relocate W to an alternate location S:

$$S = YW \tag{3}$$

A linear process yields the orthonormal variables that form the entire S-matrix of information mixture of components from the W-matrix. This combination describes the relationship between the samples. The Smatrix of covariance is defined as:

$$D_S = YD_wY^S \tag{4}$$

The variable D_w represents the matrix of covariance for the variable$W$.

The weighting matrices Y can be obtained by solving the eigenvalue equation:

$$(D_s - \lambda J)f_j = 0 \tag{5}$$

The covariance matrix contains the bilateral covariances between the several input variables.

The matrix of covariance's eigenvalues and eigenvectors are deconstructed (as shown in Equation (5)). The resulting eigenvectors represent the new orthogonal elements, referred to as "principal components", with the associated eigenvalues characterize their magnitudes. Following an ordered sequence of the eigenvalues and associated eigenvectors, the main elements will also be arranged in that same order. The first principal component will possess the highest variance, indicating the most significant information. The subsequent principal component will exhibit the second highest variance and so on. It is important to mention that the main components obtained are not associated with one another, regardless of the correlation between the input parameters. This is because the reduced eigenvector are orthogonal.

2.4 Modified gravitational search algorithm based decision tree (MGSA-DT)

A new technique that combines the decision-making ability of a decision tree with the optimization capabilities of the MGSA is the Decision Tree. Using the GSA, this methodology searches the ideal hyper parameter configuration, such as the decision tree's optimum depth, to improve the predictive reliability of the model.

2.4.1 Modified gravitational search algorithm

In the conventional GSA method, the adjustment of the search operator's location is identified by the gravitational impact of the operator. This impact influences the acceleration value, which in turn affects the searching step size of the operator. It establishes the algorithm's global search capability and convergence efficiency. Moreover, it is worth mentioning that in the GSA, the value in the method is a constant number. This characteristic leads to a rapid convergence of the algorithm, which could result in the algorithm that is trapped in a local answer when working with certain multi-peak objective measures. To increase the value ofα, the lower number is achieved early in the algorithm by increasing the search step duration of each iteration to expand the search domain, which improves the performance of the global search, later on, the lower value is obtained by reducing the search steps length to allow the fine search to identify the optimization solution. The distribution of the population diversity indicator (ED), which could reflect the dispersion of people throughout the population searching space, is presented at this stage. This indicates the degree to which the algorithm works. Its expression is generally as follows in Equations (6-9):

$$ED = \frac{\sum_{j=1}^{M} \sqrt{\sum_{i=1}^{C} \left(W_{ji}^{L} - \underline{W}_{i}^{L}\right)^2}}{M \times K} \tag{6}$$

The population size is denoted by$M$, the feature dimension by D, the inner population centre is indicated by$\underline{w}$, the number of cycles is L and K is half of the diagonal width of the solution space.

The diversity indicator ED has a value range of (1,0) and the population's diversity decreases with decreasing ED value not the other way around. The evolution degree indexing parameter DN is added and the following expression for the equation appears:

$$DN = \frac{fit^{ave}(s-1) - fit^{ave}(s)}{fit^{ave}(s)} \tag{7}$$

$fit^{ave}(s)$represents the mean of the total particle fitness at time s.

In the MGSA method, the attenuation coefficient α has the following value based on the information above:

$$\alpha(s) = \alpha(s-1) + D_2 \times (FC - D_B) + D_3 \times (D_A - DN) \tag{8}$$

Where D_2 and D_3 are the corrected coefficients of maturity and diversity, accordingly, D_B and D_A are the comparison of maturity and diversity, as well as $\alpha(s)$ is the coefficient of attenuation for the s times iteration, $\alpha(s-1)$ is the attenuated coefficient of the final iteration of$s$. In general, the coefficients D_B and D_A are considered to be $(7.0, 3.0)$.

The formula (9) yields the corrected value of the gravity constant and the correct value of an attenuation coefficient during iteration. An innovative individual speed updating approach is proposed in this investigation, using the following calculation formula:

$$u_j^c(s+1) = rand_1 \times u_j^c(s) + rand_2 \times b_j^c(s) + rand_3 \times d \times \left(oa_j^c(s) - w_j^c(s)\right) \quad (9)$$

2.4.2 Decision tree

Classifier generation systems are a popular data mining device. Classification algorithms in machine learning are able to process large amounts of data. It could be used to categorize newly available data, classify knowledge based on training sets and label categories, to create assumptions about the names of categorical classes. Figure 2 illustrates the decision tree structure.

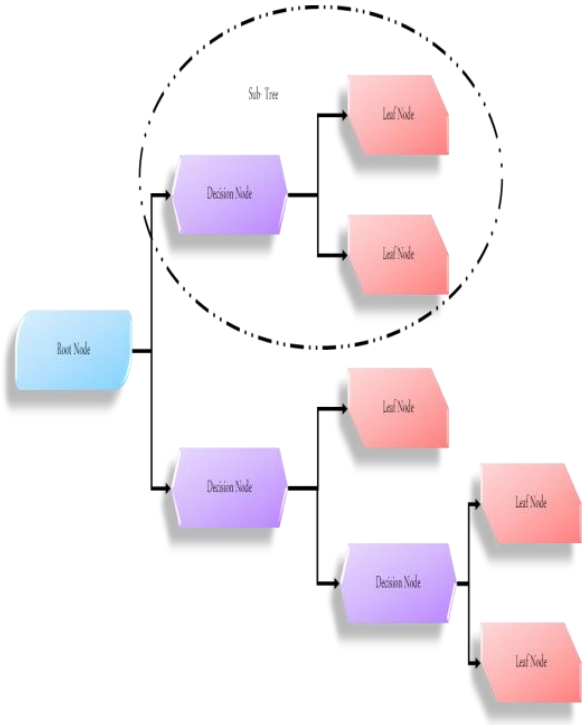

Figure 2. Decision tree structure

Decision trees (DT) are an effective technique used in multiple fields, such as recognition of patterns, processing images, and artificial intelligence. DT is a structured methodology that uses a series of basic tests to efficiently and methodically integrate results. Every test checks a numerical attribute up to a certain limit. Building the theoretical foundations is considerably easier than finding the exact values of the neural network's interconnections; DT is mostly used for the purpose of categorization. In addition, within the data analysis field, DT is a popular categorization model. Each tree is made of nodes and branches. Every node symbolizes characteristics in a certain category that need to be categorized and each subset indicates a possible value that the node could acquire. Due to its simple analysis and accurate handling of many data designs, decision trees have been widely used in several domains.

2.4.3 MGSA-DT

The MGSA-DT is a novel method that integrates the concepts of the GSA with decision tree building to improve optimization and decision-making in complicated problem environments. The GSA is derived from the principles of gravitation and the connections between the general populations yet it is optimized for efficient exploration of a possible space. The MGSA-DT framework utilizes the gravitational search method to maximize the decision tree parameters. This adaptation allows the model to learn from the intricate connections present in the manufacturing dataset. The algorithm navigates the solution space, adapting decision tree branches and nodes to identify characteristics that are predictive of favourable results.

This improved strategy not only speeds up the development of the method but also improves its capability to handle extensive and ever-changing production datasets. The combination of gravitational search and decision tree structures enhances the capabilities of MGSA-DT to make a substantial contribution to early quality prediction in production. This proactive tool enables the identification of possible faults before they become more serious, thereby optimizing the manufacturing process in general. Algorithm 1 illustrates (MGSA-DT).

Algorithm 1: (MGSA-DT)

```
def calculate_fitness(max_depth):
model.fit(X_train, y_train)
   return accuracy_score(y_test, model.predict(X_test))
population = np.random.rand(population_size, dimension)
for iteration in range(iterations):
fitness_values = np.array([calculate_fitness(int(position[0] * tree_max_depth)) for position in population])
   G = G0 * np.exp(-alpha * iteration / iterations)
   forces = np.zeros_like(population)
   for i in range(population_size):
      for j in range(population_size):
         if i != j:
            r = np.linalg.norm(population[i] - population[j])
            forces[i] += (G * fitness_values[i] * fitness_values[j] / (r + 1e-10)) * (population[j] - population[i]) / (r + 1e-10)
   population += forces
   population = np.clip(population, 0, 1)
best_solution = population[np.argmax(fitness_values)]
best_max_depth = int(best_solution[0] * tree_max_depth)
final_model = DecisionTreeClassifier(max_depth=best_max_depth, random_state=42).fit(X_train, y_train)
```

final_accuracy = *accuracy_score*(y_test, *final_model*.predict(*X_test*))
print("Final Decision Tree Accuracy:", *final_accuracy*)

3. EXPERIMENTAL RESULTS

The proposed solution is executed using Python version 3.10.1 on a laptop running Windows 10, equipped with 8 GB of RAM and an Intel i7 core CPU. Utilize libraries such as Scikit-Learn or Tensor Flow/Keras to train our suggested model on the training data. The suggested technique is (MGSA-DT) compared to existing methods such as Naive Bayes (Syafrudin, M., et al., 2018), Multi layer perception and XGBoost (Jung, H., et al., 2021). The performance of these methods is examined based on accuracy, precision and recall.

Figure 3 displays the accuracy and loss results for both the testing and training information of the suggested MGSA-DT approach.

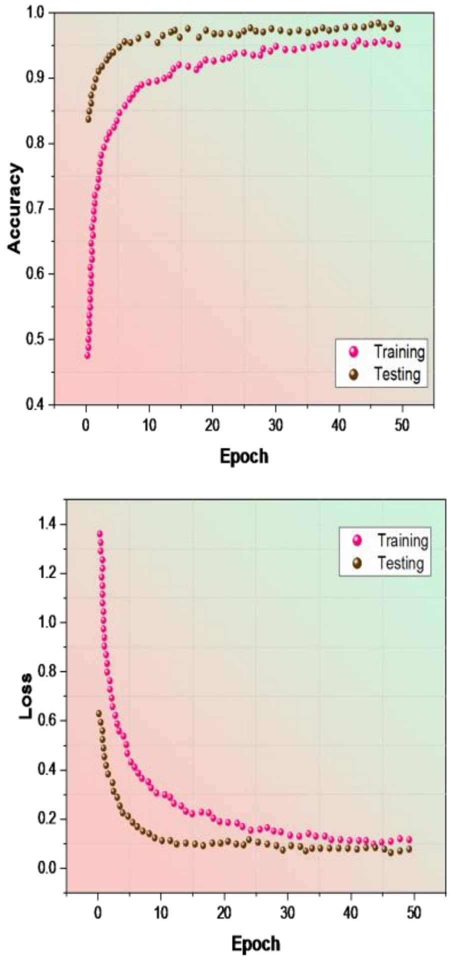

Figure 3. Accuracy and loss

The accuracy along with the dependability of the techniques are used to recognize and anticipate possible flaws or departures from quality requirements during the manufacture process are referred to as accuracy in the framework of early quality identification in manufacturing. It illustrates the extent to which the detecting system can determine a product's quality state earlier as shown in Equation (10).

$$Accuracy = \frac{TP+FP}{(TN+TP+FN+FP)} \tag{10}$$

Figure 4 and Table 1 display the accuracy performance. The proposed MGSA-DT system achieves an accuracy of 98.4%, outperforming the existing methods Naive Bayes, Multi layer perception and XGBoost, which have accuracies of 93.5 %, 96.7 % and 98 %, respectively. The suggested system is more accurate than the present methods used for early quality detection in the production process.

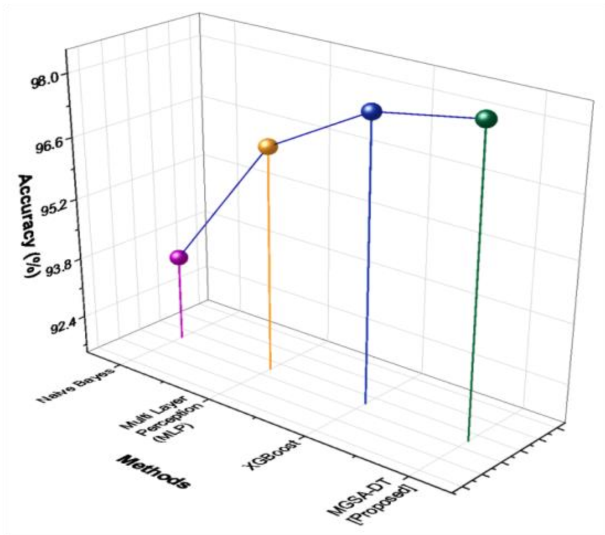

Figure 4. Accuracy performance

Table 1. Accuracy Values

Methods	Accuracy (%)
Naive Bayes	93.5
Multi-Layer Perception (MLP)	96.7
XGBoost	98
MGSA-DT [Proposed]	98.4

In the framework of early quality identification in manufacturing, precision refers to the system's accuracy and dependability in recognizing and predicting possible quality problems at the beginning of production. It calculates the percentage of true positive predictions, that is, the real cases of early quality issues that the system accurately identified among all the occurrences that have been designated as problematic as shown in Equation (11).

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

Figure 5 and Table 2 show the precision performance. The suggested system MGSA-DT has 97.6%, compared to the existing systems Naive Bayes, Multi layer perception and XGBoost, which are respectively 94.1 %, 96.8 % and 67.6 %. As a result, the proposed system is more precise than the current approaches of early quality detection in manufacturing process.

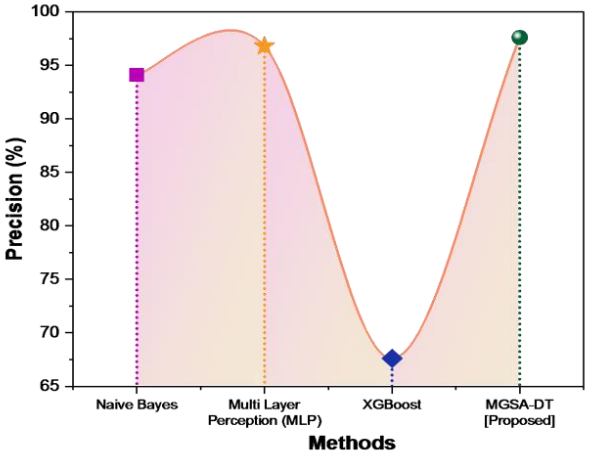

Figure 5. Precision performance

Table 2. Precision values

Methods	Precision (%)
Naive Bayes	94.1
Multi-Layer Perception (MLP)	96.8
XGBoost	67.6
MGSA-DT [Proposed]	97.6

The capacity of an approach or method to recognize and retrieve occurrences of faulty or inconsistent goods during the early phases of the production process is referred to as recall in the framework of early quality identification in manufacturing. It highlights the significance of reducing false negatives by showing the percentage of real damaged goods that the system properly detects in Equation (12)

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

Figure 6 and Table 3 shows the Recall performance. In comparison to the existing systems Naive Bayes, Multi layer perception and XGBoost, which have corresponding reliable rates of 93.6 %, 96.8 % and 60.52 %, the proposed system MGSA-DT obtained 97.2%. The suggested system is, therefore, more reliable than the present methods of early quality detection in the manufacturing process.

Table 3. Recall values

Methods	Recall (%)
Naive Bayes	93.6
Multi-Layer Perception (MLP)	96.8
XGBoost	60.52
MGSA-DT [Proposed]	97.2

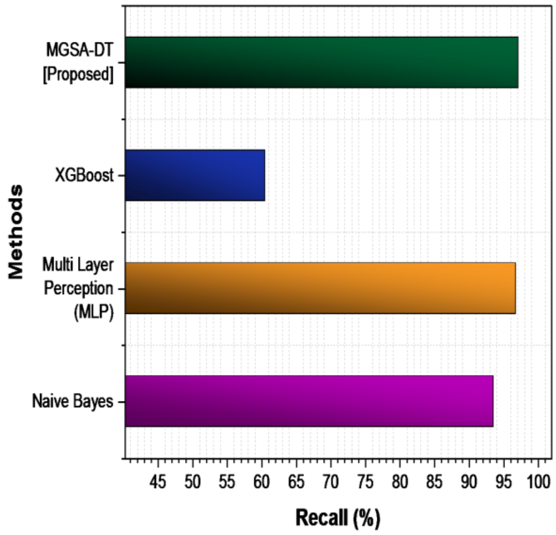

Figure 6. Recall performance

4. DISCUSSION

The existing methods are Naive Bayes, Multi layer perception and XG Boost. Naive Bayes assigns equal importance to data, thus unnecessary or redundant characteristics can generate noise and adversely affect predicting accuracy. In manufacturing applications, where data can contain several dimensions and comprise different characteristics, there is a greater chance to include irrelevant features in the model. This could decrease the usefulness of the algorithm in predicting quality earlier. The efficacy of MLP models is dependent upon the abundance and excellence of data accessible for training purposes. MLPs can have difficulties in capturing the underlying patterns when presented with limited or noisy data. The reliance on data can provide a significant disadvantage in manufacturing applications, when acquiring extensive and superior datasets could prove to be difficult. XGBoost produces intricate models that present difficulties in interpretation, preventing domain specialists and operators from understanding the underlying reasoning behind the framework's predictions. In the manufacturing industry, the ability to understand and explain decisions is crucial. The MGSA-DT facilitates early detection by offering a resilient framework for processing data in real-time. The incorporation of decision tree modeling enables rapid and precise detection of possible quality problems, hence enabling immediate action and minimizing the risk of faults spreading throughout the manufacturing process.

5. CONCLUSION

Early quality detection in the manufacturing company refers to the proactive identification and prevention of faults or deviations from quality requirements during the manufacturing process. The use of quality control measures in manufacturing is crucial for guaranteeing the creation of superior products, decreasing the occurrence of flaws and limiting the expenses associated with production. In this research, we proposed the MGSA-DT as a means for predicting the quality of manufacturing processes in their early phases. We gathered sensors data in the manufacturing industry. Z-score normalization is employed for data preprocessing and to extract the pertinent feature from the pre-processed data through principal component analysis (PCA). When compared to the existing method, the proposed method achieves accuracy (98.4%), precision (97.6%) and recall (97.2%), respectively. Applying the method in extensive manufacturing operations could involve substantial computational resources and exhaustive, high-calibre information. Considering the use of augmented reality (AR) and virtual reality (VR) technologies in quality prediction procedures has the potential to transform operator education and decision-making.

References:

Ammar, M., Haleem, A., Javaid, M., Walia, R., &Bahl, S. (2021). Improving material quality management and manufacturing organizations system through Industry 4.0 technologies. *Materials Today: Proceedings*, *45*, 5089-5096.DOI: 10.1016/j.matpr.2021.01.585

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., & Zhang, J. (2020). Artificial intelligence in advanced manufacturing: Current status and future outlook. *Journal of Manufacturing Science and Engineering*, *142*(11), 110804. DOI: 10.1115/1.4047855

Ayvaz, S., &Alpay, K. (2021). Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Systems with Applications*, *173*, 114598.DOI: 10.1016/j.eswa.2021.114598

Chacko, M. (2021). Cyber-Physical Quality Systems in Manufacturing. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(2), 2006-2018.

Chen, L., Yao, X., Xu, P., Moon, S. K., & Bi, G. (2021). Rapid surface defect identification for additive manufacturing with in-situ point cloud processing and machine learning. *Virtual and Physical Prototyping*, *16*(1), 50-67. DOI: 10.1080/17452759.2020.1832695

Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry*, *123*, 103298. DOI: 10.1016/j.compind.2020.103298

Fan, S. K. S., Hsu, C. Y., Tsai, D. M., He, F., & Cheng, C. C. (2020). Data-driven approach for fault detection and diagnostic in semiconductor manufacturing. *IEEE Transactions on Automation Science and Engineering*, *17*(4), 1925-1936. DOI: 10.1109/TASE.2020.2983061

Ghahramani, M., Qiao, Y., Zhou, M. C., O'Hagan, A., & Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. *IEEE/CAA Journal of AutomaticaSinica*, *7*(4), 1026-1037. DOI: 10.1109/JAS.2020.1003114

Imoto, K., Nakai, T., Ike, T., Haruki, K., & Sato, Y. (2019). A CNN-based transfer learning method for defect classification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, *32*(4), 455-459. DOI: 10.1109/TSM.2019.2941752

Iqbal, R., Maniak, T., Doctor, F., &Karyotis, C. (2019). Fault detection and isolation in industrial processes using deep learning approaches. *IEEE Transactions on Industrial Informatics*, *15*(5), 3077-3084. DOI: 10.1109/TII.2019.2902274

Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2021). Significance of Quality 4.0 towards comprehensive enhancement in manufacturing sector. *Sensors International*, *2*, 100109. DOI: 10.1016/j.sintl.2021.100109

Jung, H., Jeon, J., Choi, D., & Park, J. Y. (2021). Application of machine learning techniques in injection molding quality prediction: Implications on sustainable manufacturing industry. *Sustainability*, *13*(8), 4120. DOI: 10.3390/su13084120

Kao, H. A., Hsieh, Y. S., Chen, C. H., & Lee, J. (2017). Quality prediction modeling for multistage manufacturing based on classification and association rule mining. In *MATEC Web of Conferences* (Vol. 123, p. 00029). EDP Sciences. DOI: 10.1051/matecconf/201712300029

Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives. *Expert Systems with Applications*, *200*, 116912. DOI: 10.1016/j.eswa.2022.116912

Lee, W., &Seo, K. (2021). Early failure detection of paper manufacturing machinery using nearest neighbor-based feature extraction. *Engineering Reports*, *3*(2), e12291. DOI: 10.1002/eng2.12291

Liu, H., & Wang, L. (2020). Remote human–robot collaboration: A cyber–physical system application for hazard manufacturing environment. *Journal of manufacturing systems*, *54*, 24-34. DOI: 10.1016/j.jmsy.2019.11.001

Nguyen, G., Dlugolinsky, S., Tran, V., & García, Á. L. (2020). Deep learning for proactive network monitoring and security protection. *ieee Access*, *8*, 19696-19716. DOI: 10.1109/ACCESS.2020.2968718

Oduoza, C. F. (2020). Framework for sustainable risk management in the manufacturing sector. *Procedia Manufacturing*, *51*, 1290-1297. DOI: 10.1016/j.promfg.2020.10.180

Orrù, P. F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R., & Arena, S. (2020). Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability*, *12*(11), 4776. DOI: 10.3390/su12114776

Peres, R. S., Barata, J., Leitao, P., & Garcia, G. (2019). Multistage quality control using machine learning in the automotive industry. *IEEE Access*, *7*, 79908-79916. DOI: 10.1109/ACCESS.2019.2923405

Ren, Z., Fang, F., Yan, N., & Wu, Y. (2022). State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *9*(2), 661-691. DOI: 10.1007/s40684-021-00343-6

Saengthong, D. (2023). The development of a model school for proactive learning experiences at the early childhood level in elementary quality schools Under the Office of the Basic Education Commission..

Saqlain, M., Piao, M., Shim, Y., & Lee, J. Y. (2019). Framework of an IoT-based industrial data management for smart manufacturing. *Journal of Sensor and Actuator Networks*, *8*(2), 25. DOI: 10.3390/jsan8020025

Scime, L., Siddel, D., Baird, S., &Paquit, V. (2020). Layer-wise anomaly detection and classification for powder bed additive manufacturing processes: A machine-agnostic algorithm for real-time pixel-wise semantic segmentation. *Additive Manufacturing*, *36*, 101453. DOI: 10.1016/j.addma.2020.101453

Shahin, M., Chen, F. F., Hosseinzadeh, A., &Zand, N. (2023). Using Machine Learning and Deep Learning Algorithms for Downtime Minimization in Manufacturing Systems: An Early Failure Detection Diagnostic Service. DOI: 10.21203/rs.3.rs-2895149/v1

Shcherbakov, M. V., Glotov, A. V., &Cheremisinov, S. V. (2019). Proactive and predictive maintenance of cyber-physical systems. In *Cyber-Physical Systems: Advances in Design & Modelling* (pp. 263-278). Cham: Springer International Publishing. DOI: 10.1007/978-3-030-32579-4_21

Stelnicki, A. M., Jamshidi, L., Fletcher, A. J., & Carleton, R. N. (2021). Evaluation of before operational stress: a program to support mental health and proactive psychological protection in public safety personnel. *Frontiers in psychology*, *12*, 3218. DOI: 10.3389/fpsyg.2021.511755

Suman, S., & Das, A. (2019). Stratified statistical monitoring strategy for a multi-product manufacturing facility with early detection approach. *Computers & Industrial Engineering*, *130*, 551-564. DOI: 10.1016/j.cie.2019.03.018

Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, *18*(9), 2946. DOI: 10.3390/s18092946

Westphal, E., & Seitz, H. (2021). A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. *Additive Manufacturing*, *41*, 101965. Additive Manufacturing, 41, 101965. DOI: 10.1016/j.addma.2021.101965

Xu, D., Zhang, Z., & Shi, J. (2022, June). Failure prediction using gated recurrent unit and autoencoder in complex manufacturing process. In *2022 4th World Symposium on Artificial Intelligence (WSAI)* (pp. 68-73). IEEEDOI: 10.1109/WSAI55384.2022.9836412

Xu, K., Li, Y., Liu, C., Liu, X., Hao, X., Gao, J., &Maropoulos, P. G. (2020). Advanced data collection and analysis in data-driven manufacturing process. *Chinese Journal of Mechanical Engineering*, *33*(1), 1-21. DOI: 10.1186/s10033-020-00459-x

Yu, W., Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y. (2019). Yu, W., Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y. (2019). A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Transactions on Industrial Informatics*, *16*(1), 183-192. IEEE Transactions on Industrial Informatics, 16(1), 183-192. DOI: 10.1109/TII.2019.2915846

Swati Singh
Maharishi University of Information Technology, Uttar Pradesh, India
swatisingh5444@gmail.com
ORCID 0009-0007-3888-3258

Raman Batra
Noida Institute of Engineering & Technology, Greater Noida, Uttar Pradesh, India
ramanbatra@niet.co.in
ORCID 0009-0006-7359-7313

Keerti Rai
Arka Jain University, Jamshedpur, Jharkhand, India
dr.keerti@arkajainuniversity.ac.in
ORCID 0000-0003-1819-0722

S. Sujai
Jain (Deemed to be University), Bangalore, Karnataka, India
s.sujai@jainuniversity.ac.in
ORCID 0000-0003-0981-4273