Research Article
Journal of Intelligent Systems: Theory and Applications 7(2) (2024) 129-144
DOI: 10.38016/jista.1516453

Classification of Quality Defects using Multivariate Control Chart with Ensemble Machine Learning Model

Deniz Demircioğlu Diren[1*], Semra Boran[2]

[1] Distance Education Research and Application Center, Sakarya University, Sakarya, Türkiye

[2] Faculty of Engineering, Industrial Enginering Department, Sakarya University, Sakarya, Türkiye

ddemircioglu@sakarya.edu.tr, boran@sakarya.edu.tr

Abstract
Multivariate control charts enable to monitor processes affected by more than one variable. But, when the process is out of control, it cannot detect which variable is causing it. It is an important requirement to know which variables in the process need corrective actions. In this study, a machine learning-based model is proposed to predict the variable/s that make the process out of control. For this purpose, ensemble algorithms, which are known to have higher prediction performance than single algorithms, were preferred. Because it is aimed to determine the variable(s) that cause the process to be out of control in the most accurate way. It is thought that a classification model in which ensemble algorithms are used together can increase the prediction accuracy. The model, which has not been encountered before in a quality control problem, was applied to a real problem and 98.06% classification accuracy was achieved. Another benefit is that it can predict the variable/variables that make the process uncontrolled without the need for multivariate control charts.

Keywords: Multivariate control chart, Machine learning, Ensemble of ensemble algorithm, Hotelling T^2 chart, Mason-Young-Tracy method.

Çok Değişkenli Proses Kontrol Grafiği ve Topluluk Makine Öğrenme Modeli Kullanılarak Kalite Kusurlarının Sınıflandırılması

Öz
Çok değişkenli kontrol diyagramları birden fazla değişkenin etki ettiği süreçlerin izlenmesine olanak sağlamaktadır. Ancak süreç kontrol dışında olduğunda hangi değişkenin buna neden olduğunu tespit edilememektedir. Süreçteki hangi değişkenlerin düzeltici faaliyetlere ihtiyaç duyduğunu bilmek önemli bir gerekliliktir. Bu çalışmada süreci kontrolden çıkaran değişken/değişkenleri yüksek doğrulukla belirlenmesi tahmin etmek için makine öğrenmesi tabanlı bir model önerilmiştir. Bu amaçla tekli algoritmalara göre daha yüksek tahmin performansına sahip olduğu bilinen topluluk algoritmaları tercih edilmiştir. It is thought that a classification model in which ensemble algorithms are used together can increase the prediction accuracy. Daha önce bir kalite kontrol probleminde rastlanmayan model, gerçek bir probleme uygulanmış ve %98,06 sınıflandırma doğruluğu elde edilmiştir. Ayrıca bir diğer faydası da çok değişkenli kontrol grafiklerine ihtiyaç duymadan süreci kontrolden çıkaran değişken/değişkenleri tahmin edebilmesidir.

Anahtar Kelimeler: Çok değişkenli kontrol grafiği, Makine öğrenmesi, Topluluk algoritması topluluğu, Hotelling T^2 grafiği, Mason-Young-Tracy yöntemi.

1. Introduction

In order to produce quality products and ensure their sustainability, processes must be constantly monitored. The causes of out-of-control situations encountered while monitoring the processes should be determined as accurately and quickly as possible and corrective actions should be implemented. Since the products used today have a much more complex structure, the production processes should be evaluated according to their many features (Robert, 2002). While traditional control charts deal with a single measurable product feature (variable), multivariate control charts have the feature of being tools that can handle multiple variables simultaneously (Montgomery, 2009). Thus, time and cost savings are

* Corresponding Author.
E-mail: ddemircioglu@sakarya.edu.tr

Recieved : 15 Jul 2024
Revision : 7 Aug 2024
Accepted : 26 Aug 2024

achieved. In addition, these control charts also enable the evaluation of the relationship between the variables (Hotelling, 1947; Woodall, 1985; Lowry, 1992). In addition to this advantage, the most criticized feature of the charts is that it cannot detect which variable/s caused it in case of out-of-control signals (Aparasi, 2006). However, this causes very important problems because it is necessary to know which variable(s) corrective action should be applied in order to control the process. Because wrong estimation of variable lead to loss of time, increase in finances and worst of all, poor quality products. For this, traditional methods are not sufficient and new methods are needed. The leading of these is the Mason, Young, Tracy Decomposition (MYT) method, which has been specially developed for quality control charts. Principal component analysis and discriminant analysis are also used for similar purposes. (Jackson,1985; Rao et al., 2013; Pei et al., 2006; Hawkins, 1991; Mason et al., 1997; Das and Prakash, 2008; Li et al., 2008; Agog et al., 2014; Joshi and Patil, 2022). In addition to the mentioned statistical methods, it is seen that machine learning algorithms are frequently used in recent years (Aparasi, 2006; Niaki and Abbasi, 2005; Chen and Wang, 2004; Cheng and Cheng, 2008; Song et al., 2017; Shao and Lin, 2019; Du et al., 2012; Asadi and Farjami, 2019; Ahsan et al., 2020; Sabahno & Amiri, 2023). However, statistical methods have weaknesses such as not being able to make predictions for new data and not measuring the accuracy of the results with various criteria. Machine learning algorithms are more preferred because they have features to eliminate these weaknesses. In this study, a new ensemble machine learning model developed using the results obtained from Hotelling T^2 and MYT methods is presented in order to detect the variable(s) causing out-of-control situations. With this model, it is aimed to determine the variable(s) that cause out-of-control situations as accurately as possible. It is known that ensemble machine learning algorithms provide more accurate predictions than single algorithms (Jiang and Song, 2017; Asadi and Farjami, 2019). For this reason, the bagging and boosting ensemble algorithms in the classification model developed in the study were combined with the stacked generalization algorithm, which is another ensemble algorithm, and the ensemble structure ensemble was used. Thus, the variable(s) causing the out-of-control situation were determined in the most accurate way. Model data were obtained with Hotelling T^2, which is a multivariate control chart, and MYT method, which was specially developed for the chart. In order to determine the algorithm to be used in the model, Decision Trees (DT), Naive Bayes (NB), K-Nearest Neighbor (KNN), Multi Support Vector Machines (M-SVM) and Artificial Neural Networks (ANNs) were used, which are among the most basic single algorithms. Since the aim was to increase prediction accuracy, the algorithm that was most successful in single uses was chosen first. Then, ensemble models were developed with this algorithm using bagging and boosting.

The subsequent of the article is organized as follows. In Section 2, a literature review will be conducted. After explaining the methods in Section 3, the proposed model will be presented in Section 4. Then, the implementation will be carried out in Section 5 to carry out the experimental study of the model. The article concludes with Section 6, where discussion and conclusion is presented.

2. Literature Review

There are many studies in the literature about the determination of the variable(s) that cause the out-of-control situation, using statistical and machine learning methods.

The most frequently used method in the literature for multivariate control charts is the Mason Young Tracy (MYT) decomposition method (Robert, 2002). This method, which was developed by Mason et al., (1995), was designed specifically for the Hotelling T^2 control chart, based on principal component analysis (Mason et al., 1995; Özel, 2005). There are studies in many different areas where MYT is used (Çetin and Birgören, 2007; Parra and Loaiza, 2003; Ulen and Demir, 2013; Boullosa et al., 2017; Yilmaz, 2012).

Studies in which machine learning techniques are used to determine the variables that cause the out-of-control situation are examined in two classes as studies in which basic algorithms are used individually and as an ensemble.

Studies using single algorithms to detect variables that cause out-of-control situations have been encountered for many years. In two separate studies by Chen and Wang (2004) and Niaki and Abbasi (2005), an artificial neural network-based model was developed for the X^2 chart and presented by evaluating its successful performance. In the study performed by Aparisi et al. (2006), accuracy analysis of MYT method and neural network was performed in terms of classification. According to the results; It has been seen that the accuracy performance of the designed neural network is better than the accuracy performance of the MYT method (Aparisi et al., 2006). In the application by Cheng and Cheng (2008), which aims to detect variables with Artificial Neural Network (ANN) and Support Vector Machine (SVM), the performance of SVM was found to be similar to ANN. In addition, it has been stated that the ANN algorithm has weaknesses such as the large number of control parameters and the difficulty of applying steps. In another study, Li et al. (2013) compared the optimized SVM approach with the developed ANN for the estimation of the shift magnitude in the process. As a result, the best performance of the SVM approach has been demonstrated. Huda et al. (2014) developed an ANN-based model that does not need expert knowledge and requires little numerical computation. The results showed that the proposed approach is successful and easy to implement. Song et al. (2017) proposed a sample-based Navie Bayes (NB) method to interpret out-of-control situations. As a result of the performance comparisons, it was stated that the developed method outperformed other statistical techniques. In the study by Shao and Lin (2019), ANN-based classification model was developed in a multivariate process with variance shift. The performance of this model is compared with ANN, SVM and multivariate adaptive regression classifier. As a result, it was stated that the developed model was more successful. Bersimis et al. (2022) an ANN-based model was developed that uses the results of some analytical methods as input for the detection of uncontrolled variables. According to the results obtained, very successful results were obtained with the developed model. In another study conducted by Rakhmawan et al (2023), the Hotelling T^2 control chart was optimized with the decision tree model. It has been stated that this is a solution that can be used to obtain accurate predictions.

There are studies where ensemble algorithms are used to detect variables that cause out-of-control situations. In the study by Guh and Shiue (2008), a simple and effective model obtained by sequentially combining the Decision Tree (DT) classification algorithm is proposed to detect the mean shifts in multivariate control charts. Experimental results show that the learning speed of the proposed model is much faster than an ANN-based model. For the same purpose, an ANN-based ensemble model was developed by Yu et al. (2009). The results of the study, which produced data according to 5 different shift sizes from the mean for each variable by simulation, are presented that the proposed model outperforms the use of single ANN in terms of average running length (ARL). In the study by Alfaro et al. (2009), ensemble trees have proven to be a very powerful tool for classification accuracy. Du et al. (2012) classified the causes of mean shifts in the multivariate process with the multiclass bagging ensemble SVM algorithm. The performance of the model evaluated according to the accuracy criterion with a real application has been proven to be effective. Similarly, the approach developed in the study by Cheng and Lee (2012) using the bagging ensemble SVM algorithm is compared with the traditional decomposition method and its performance is seen to be more successful. Yang (2015) concluded that the proposed artificial neural network ensemble model is a more effective approach in diagnosing out-of-control situations than other approaches in the literature. In the study by Jiang and Song (2017), which developed an ensemble model by combining decision trees in parallel, it was proven that the classification performance of the ensemble learning method was better. Another study in which decision trees were applied as an ensemble was carried out by Asadi and Farjami (2019). In the study, a structure with four classifiers in which decision trees are connected sequentially and a Monte Carlo simulation are used. The developed model ARL functions were compared according to accuracy, precision and precision criteria. The results showed better performance of the community DT construct. In the research conducted by Alfaro et al. (2020), the random forest method was used to detect out-of-control situations. This method has been compared with ANN and it has been stated that the random forest method is more successful when there is small and medium correlation between variables.

In this study, a ensemble algorithm is proposed in which bagging and boosting ensemble algorithms are combined. Based on the stacked generalization algorithm, this model was used to combine the power of other ensemble algorithms to detect variables that cause out-of-control in a multivariate process. The decision of the basic single algorithm to be used in the bagging and boosting ensemble algorithms was also made according to the high accuracy rate.

3. Methods

3.1. Hotelling T^2 Control Chart

Hotelling T^2 control chart was developed by Hotelling in 1947 to monitor the related p number of variables simultaneously (Montgomery, 2009). The chart is formed by scheduling the T^2 statistic, which is a statistical distance measure based on a multivariate normal distribution (Çetin and Birgören, 2007). In case the sample size is 1, the steps of the control chart are as follows. For each sample, the T^2 statistic is calculated with the help of Equation (1) according to p number of variables.

$$T^2 = (X - \bar{X})' S^{-1} (X - \overline{X}) \tag{1}$$

Where, X is variable, $\bar{X}$ is sample mean vector and S is the sample covariance. While the upper control limit (UCL) for the first phase of the multivariate control chart is calculated according to Equation (2), the lower control limit (LCL) is taken as the zero line as seen in Equation (3).

$$UCL = \frac{(m-1)^2}{m} \beta_{\alpha,p/2,(m-p-1)/2} \tag{2}$$

$$LCL = 0 \tag{3}$$

Where, m expresses the upper α percentage point of the beta distribution with the parameters $\beta_{\alpha,p/2,(m-p-1)/2}$ including the number of samples (Montgomery, 2009).

In order to use the Hotelling T^2 control chart, some assumptions must be met. These assumptions are conformity to multivariate normal distribution, linearity, absence of autocorrelation, variance covariance equality (homogeneity). If there are (s) not provided by the assumptions, the necessary conversion actions should be applied.

3.2. Mason Young Tracy (MYT) Decomposition Method

This method was developed by Mason, Young and Tracy in the 1990s to detect out-of-control variables by splitting the Hotelling T^2 statistic into two orthogonal parts, conditionally and unconditionally. In this method, firstly, the operated and operated are defined continuously and calculations are made. Then possible MYT decompositions are shown, and finally, similar values are calculated in periods and comments are made about the out-of-control variables (Mason et al., 1995).

T^2 statistic in Equation 1 is formed by combining conditional and unconditional terms as seen in Equation 4.

$$T^2 = T_{p-1}^2 + T_{p.1,\ldots,p-1}^2 \quad (4)$$

Here, the part shown in Equation (5) expresses the unconditional terms.

$$T_{p-1}^2 = \left(X_i^{(p-1)} - \bar{X}^{(p-1)}\right)' S_{XX}^{-1}\left(X_i^{(p-1)} - \bar{X}^{(p-1)}\right) \quad (5)$$

Where, $\bar{X}^{(p-1)}$ is the mean vector of n multivariate observation values of the first (p-1) variable. S_{XX} is the (p-1)*(p-1) basic submatrix of S.

The part shown in Equation (6-9) expresses the conditional terms.

$$T_{p.1,\ldots,p-1}^2 = \frac{x_{ip} - \bar{x}_{p.1,\ldots,p-1}}{s_{p.1,\ldots,p-1}^2} \quad (6)$$

$$\bar{X}_{p.1,\ldots,p-1} = \bar{X}_p + b_p' \left(X_i^{(p-1)} - \bar{X}^{(p-1)}\right) \quad (7)$$

Where $\bar{X}_p$ is the sample mean of n observation values of the pth variable.

$b_p = S_{XX}^{-1} s_{xX}$ is the dimensional vector that estimates the regression coefficients of the p-th variable in the first p-1 variable.

$$s_{p.1,\ldots,p-1}^2 = s_x^2 - s_{xX}' S_{XX}^{-1} s_{xX} \quad (8)$$

$$S = \begin{bmatrix} S_{XX} & s_{xX} \\ s_{xX}' & s_x^2 \end{bmatrix} \quad (9)$$

Where, s_{xX} is the vector of covariance between variables, s_x^2 is the variance of the variable p.

3.3. Machine Learning Algorithms

Machine learning is a technology developed to enable machines to be intelligent, enabling systems to learn directly from examples, data and experiences (The royal society, 2017). These technologies enable machines to make predictions, perform clustering, extract association rules or make decisions from a given data set (Mohammed et al., 2016). It is possible to examine algorithms in two classes, as single and ensemble, according to their usage structure.

3.3.1. Single machine learning algorithms

In single algorithms, only one algorithm is run and the results are obtained accordingly. In the study, DT, NB, ANN, SVM, KNN algorithms will be discussed.

A Decision Tree (DT): DT has a tree structure consisting of nodes. These nodes are called root, intermediate and leaf nodes according to their purpose (Maimon and Rokach, 2010). The working steps of the algorithm first start from the root. Then it continues by branching from the intermediate node to the leaf node. Classes in the tree are represented by leaves, and there is only one path to each leaf (Bilgin,2018; Maimon and Rokach, 2010; Han et al., 2012; Mitchell, 2014; Agrawal and Imielinsk, 1993; Utgoff et al., 1997). The samples are classified from the root of the tree to a leaf according to the result of the tests carried out along the way. These results can then be combined into a rule by taking the class estimate of the leaf as the class value (Maimon and Rokach, 2010). This structure, which can be re-represented with IF-THEN rule sets for easy understanding by the user (Mitchell,2014), can contain both nominal and numerical properties. Commonly used criteria for determining the root node feature are Information Gain, Gini index, Gain Ratio (Maimon and Rokach, 2010).

Naive Bayes (NB): NB is used when there is leading knowledge and provides a probabilistic approach to logical inference. It aims to combine the value from the sample with the leading information. This algorithm ignores the relationships between the inputs and reduces a multivariate distribution to multiple univariate distributions, as seen in Equation (10) (Alpaydın, 2012).

$$p(x|C) = \prod_{j=1}^{d} p\left(x_j|C\right) \quad (10)$$

Here; P(X): Probability of X (independent), P(Y): Probability of Y (independent), P(X│Y): Probability of X occurring when Y has occurred and P(Y│X): Probability of Y occurring when X has occurred.

K-Nearest Neighbor (KNN): KNN is based on classification with the nearest neighbors approach (Han et al., 2012). The number of neighbors (k) is determined by the user. In order to find the location of the nearest neighbors of a sample, a distance function or criteria such as Euclidean, Manhattan and Minkowski Distance, which measure the similarity between two samples, are used (Bilgin, 2018). Euclidean Distance shown in Equation (11) was used in the study. Where p and q are two examples compared.

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^{k}(p_i - q_i)^2} \quad (11)$$

Here p and q are two examples compared.

Artificial Neural Network (ANN): ANN is an important classification method that includes parallel computation programs that work similar to the human brain. Multilayer Perceptron (MLP), which is the most commonly used artificial neural network model, consists of three layers: input, hidden and output. While the number of process elements in the input and output layer is determined according to the problem, the number of elements in the hidden layer is determined by trial and error in order to achieve the best performance. The weights showing the importance of the information are determined randomly at the beginning (Öztemel, 2003). Inputs are converted to output with the activation function (Yadav et al., 2015).

Multi-Class Vector Machine (M-SVM): The SVM method developed by Cores and Vapnik (1995) is used for two-group classification and prediction problems of both linear and non-linear data. Its working principle is based on transforming the size of the data, determining decision surfaces and dividing it into two classes in the most appropriate way. When the number of classes is more than two, multi-class vector machines should be used. There are three options for this algorithm.

Here w is the weight vector, x is the sample and r is the class of the data. The size of this interval is very important for the accuracy of classification. When $r^t = +1$ and $r^t = -1$.

1. When K>2, K two class problems are defined and K different separators distinguish each class from other classes; i=1,….,K support vector machine is trained. Here, while training the parser, the samples from the class C_i are classified as +1, and the samples from the class C_k k≠1 are classified as -1. All values are calculated and the largest one is selected.
2. The problem is divided into multiple linear subproblems. The algorithm for this is to train with K(K-1)/2 discriminant binary classifiers, similar to two-class SVM.
3. In this option, a single multi-class optimization problem that includes all classes is considered as seen in Equation (12).

$$\min \frac{1}{2} \sum_{i=1}^{K} \|w_i\|^2 + C \sum_i \sum_t \xi_i^t \quad (12)$$

where the constraints are as seen in Equation (13, 14).

$$w_{z^t} x^t + w_{z^t 0} \geq w_i x^t + w_{i0} + 2 - \xi_i^t, \forall i \neq z^t \quad (13)$$

$$\text{and } \xi_i^t \geq 0 \quad (14)$$

Although this option is a very good approach, it is less preferred than other options in terms of usage due to processing load and time.

3.3.2. Ensemble machine learning algorithms

Ensemble algorithms are predictive models created by combining multiple algorithms of the same or different types with various methods in different ways (Rokach, 2010). It is aimed to achieve higher prediction accuracy with ensemble algorithms than single algorithms.

Ensemble algorithms can be created as dependent /independent and homogeneous/heterogeneous. In the dependent method, the output of one classifier is used by the next classifier. Thus, it is possible to take advantage of the knowledge produced in previous iterations to guide learning in the next iterations. In independent methods, each classifier is created independently and its outputs are combined (Maimon and Rokach, 2010). In dependent methods, algorithms are connected in series with each other, while in independent methods, algorithms are connected in parallel. The basis of parallel ensemble methods is to use independence between single algorithms, since classification and prediction error can be significantly reduced by combining independent base learners (Zhou, 2012).

In addition to combining the algorithms dependently and independently, there are ensemble algorithms that are obtained homogeneously by using the same single algorithm and heterogeneously by using different single algorithms. The classification of ensemble algorithms according to the merging principles is given in the Figure 1 (Zhou, 2012; Gowda et al., 2018).

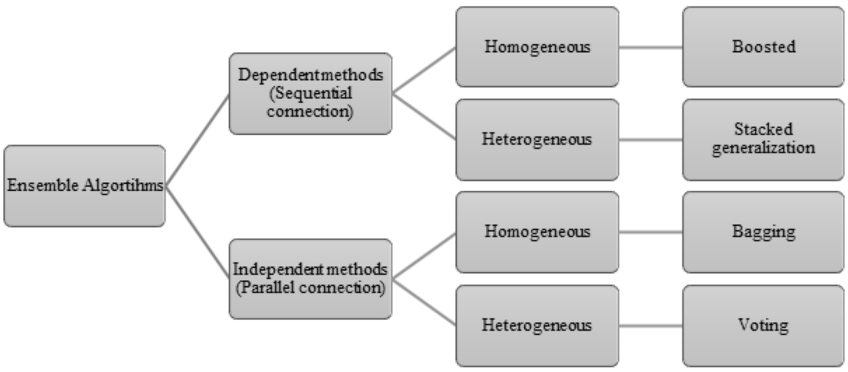

Figure 1. Types of ensemble algorithms

The bagging algorithm method was developed by Breiman (1996) and is the oldest and simplest ensemble algorithm. This method, which is based on combining basic learners in parallel, is a method that can be used with multiple classes (Zhou, 2012; Gowda et al, 2018; Zhang and Ma, 2012).

The boosting ensemble method is based on the principle that algorithms use the output of the previous algorithm as input, and the algorithms are connected in series. In this method, each classifier is affected by the performance of the previous algorithm and gives more importance to classification errors made by previously created classifiers (Rokach, 2010). When the number of classes is more than two, the AdaBoost method, which is the most preferred boosting method, is used (Zhou, 2012).

The stacked generalization method is a meta-learning based ensemble algorithm. Based on the predictions and correct answers of the basic learning algorithms, a meta-learner is trained (Onan, 2018). Here, the basic idea is to train the first-level learners using the original training dataset and then create a new dataset to train the second-level learner in which the outputs of the first-level learners are considered as the input features. First-level learners are often produced by applying different learning algorithms, and therefore stacked method are often heterogeneous (Zhou, 2012). The second-level metadata set consists of the predictions of all algorithms (Onan, 2018).

3.3.3. Performance criteria of machine learning models

In the study, the variable(s) that cause the out-of-control situation are determined by classification. For this reason, performance criteria such as accuracy, classification error, sensitivity and kappa statistics used in the classification problems of the learning performances of the developed models were evaluated. Performance criteria are as in Table 1 (Hossin and Sulaiman, 2015).

Where, g_{pi} is the true number of positives in class i, g_{ni} is the actual number of negatives in class i, y_{pi} is the number of false positives in class i, y_{ni} is the number of false negatives in class i, h_M is the macro mean of sensitivity, k_M is represents the macro average of precision.

Another criterion, the Kappa statistic, evaluates the classification accuracy by taking into account the chance factor in the probability of a correct guess. It is calculated as seen in Equation (15) (Lantz, 2013).

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{15}$$

Where, $\Pr(a)$ and $\Pr(e)$ represent the agreement ratio between the actual and expected values, the classifier and the actual values, respectively. Kappa values are commonly interpreted as; Bad estimate = less than 0.20, Acceptable estimate = 0.20 – 0.40, Intermediate estimate = 0.40 to 0.6, Good estimate = 0.60 to 0.80, Very good estimate = 0.80 – 1.00.

3.3.4. Handling Imbalanced Dataset

If the classes in the dataset are not approximately equally represented, the dataset can be eliminated the imbalanced. The performance of machine learning algorithms is often based on predictive accuracy. However, when data are unbalanced, often the majority class is predicted with little error, while the minority class(es) cannot be predicted. In this case, it can be said that using predictive accuracy would be misleading. Class imbalance in the data is addressed in two ways. The first is that it assigns different weights to the training examples. The other is to resample the original dataset by either oversampling the minority class and/or undersampling the majority class (Chawla et al., 2002). Synthetic Minority Oversampling (SMOTE) method, widely used for resampling, is a sampling technique that produces synthetic samples from the minority class. This method, which synthetically equates the number of data in the minority class to the number of data in the majority class, is used to obtain a training set with a balanced or nearly balanced class.

Table 1. Performance criteria based on confusion matrix for classification in multi-class problems

Performance Criteria	Formula	Description
Average Accuracy	$\frac{\sum_{i=1}^{l} \frac{g_{pi} + g_{ni}}{g_{pi} + y_{ni} + y_{pi} + g_{ni}}}{l}$	Average effectiveness of classes
Average Error Rate	$\frac{\sum_{i=1}^{l} \frac{y_{pi} + y_{ni}}{g_{pi} + y_{ni} + y_{pi} + g_{ni}}}{l}$	Average error rate of classes
Average Sensitivity	$\frac{\sum_{i=1}^{l} \frac{g_{pi}}{g_{pi} + y_{pi}}}{l}$	Average of precision per class
Average Precision	$\frac{\sum_{i=1}^{l} \frac{g_{pi}}{g_{pi} + y_{ni}}}{l}$	Precision average per class
Average F measurement	$\frac{2 * h_M * k_M}{h_M + k_M}$	F measurement per class

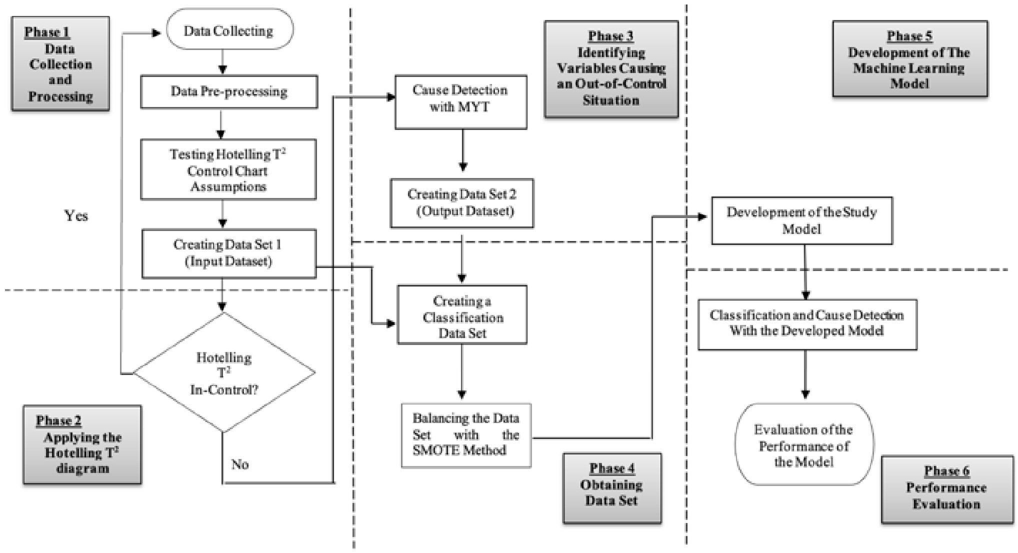

Figure 2. Architecture of the proposed model

SMOTE samples are linear combinations of two similar samples from the minority class and are obtained by Equation (16) (Blagus and Lusa, 2013).

$$s = x + u\,(x^R - x) \tag{16}$$

where x^R and x are two similar classes, x^R, is randomly selected from among the five closest minority classes of x. u is a random number between 0 and 1.

4. Proposed Model

The aim of the study is to develop an ensemble model to identify the causes of out-of-control situations in quality processes with the highest accuracy, is shown in Figure 2. The architecture consists of six phases. The phases can be stated as Data Collection and Processing, Applying the Hotelling T^2 chart, Identifying variables causing an out-of-control situation, Obtaining Data Set, Development of The Machine Learning Model and Performance evaluation.

The steps involved in the phases and the proposed model are described in detail below in Figure 2.

Phase 1. Data Collection and Processing: At this phase, data is collected about the examined properties of the manufacturing part. Before analyzing the data set, it should be checked whether it contains outlier, incomplete or inconsistent data, and if there are such cases, the data preprocessing process should be performed (Şişci et al., 2022).

Phase 2. Applying the Hotelling T^2 Chart: At this phase, it will be checked whether the T^2 statistic is suitable for linearity, normal distribution, autocorrelation and variance-covariance equality assumptions so that the data set can be used in the Hotelling T^2 control chart. Since the Hotelling T^2 control chart with a sample size of one is used in the study, the variance-covariance assumption is invalid and there is no need to check this assumption. With the linearity assumption, it is investigated whether there is a desired linear relationship between the two variables. For this, the Pearson correlation coefficients between the two variables should be calculated and compared with the level of significance. If this coefficient is greater than the significance level, there is a linearity relationship. According to the assumption of conformity to the multivariate normal distribution, each of the variables must be suitable for the normal distribution. With the Kolmogorov-Smirnov test, the conformity of the measurement values to the normal distribution is tested. After the suitability of all the variables to the normal distribution has been proven, the suitability of all the variables to the normal distribution should be evaluated with the Henze-Zirkler's test. It should be tested with the Box-Ljung statistic to determine whether there is autocorrelation between the autocorrelation assumption and the variables. After checking all assumptions, a Hotelling T^2 chart is created according to Equation (1).

Phase 3. Identifying Variables Causing an Out-Of-Control Situation: At this phase, the variable(s) that cause the samples outside the upper control limits determined by Hotelling T^2 to be out of control will be determined by MYT decomposition method.

Phase 4. Obtaining Data Set: At this phase, the inputs and outputs are brought together to obtain the data set. Variable measurement values constitute the input, and the variable classes belonging to the out-of-control situations obtained in the MYT results constitute the output. Inputs are obtained in the first phase, and outputs are obtained in the second and third phases. SMOTE was used to eliminate the imbalance caused by the difference in the data numbers of the classes in the data set. In the data set used in the proposed model, similar to other studies in the literature (Alfaro et al., 2009; Jiang and Song, 2019), only out-of-control situations are considered.

Phase 5. Development of the Machine Learning Model: After the data set to be used in the model is obtained, the algorithm to be used in the developed model will be selected. The model consists of the following steps:

1. By applying single machine learning algorithms, the most successful algorithm is selected according to the performance criteria.
2. Combining this selected single machine algorithm with bagging and boosting algorithms in parallel and sequentially.
3. Developing the two ensemble algorithms obtained in the second step by combining them with another ensemble algorithm, the stacked generalization method. The hybrid ensemble model, which is based on the combination of ensemble algorithms in order to increase the prediction performance, will be designed as seen in Figure 3.

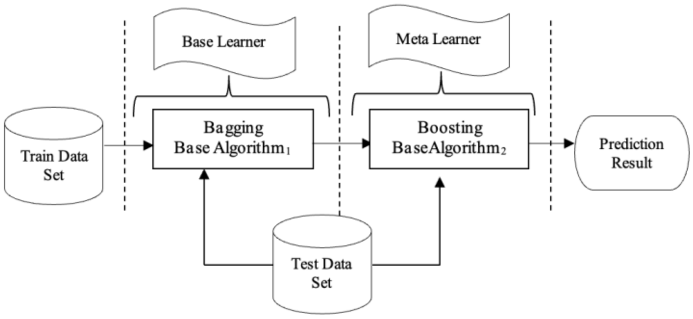

Figure 3. Architecture of the proposed model

Phase 6. Performance Evaluation: The proposed model will be trained with the dataset and its performance will be evaluated according to various criteria. If the evaluation results are found successful, the suitability of the model will be decided. For performance comparisons of classification algorithms, criteria such as accuracy, sensitivity, precision and kappa statistics were used.

5. Implementation of Proposed Model

In order to prove the validity of the proposed model, a real-life problem has been applied in the steel hydraulic pump cover production process of an automotive supplier operating in Turkey. The 3D view of the hydraulic pump cover part is shown in Figure 4. In addition, as seen Table 2, 8 variables that determine the quality of the part were determined by quality experts.

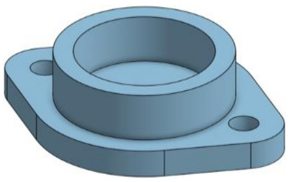

Figure 4. Hydraulic pump cover

Definition of variables, mean and specification values are as in Figure 4.

5.1. Data Collection and Processing

Data is obtained through measurements made during manufacturing. Measurements are made by taking a single sample per hour from the manufacturing process. 26700 measurement values were taken as basis in the study. Each measurement value constitutes a sample in the data set. Outlier, extreme or missing values in the data set were first examined in terms of variable and sample suitability through data pre-processing steps and it was determined that there were no data with undesirable characteristics. However, it was determined that the measurement results for some variables were missing in three samples. Therefore, these three samples were eliminated and quality evaluation was carried out on the remaining 26697 samples.

5.2. Hotelling T^2 Control Chart Implementation

Before applying the Hotelling T^2 control chart, it was checked whether the data met the assumptions regarding the T^2 statistics.

Table 2. Definition of variables

Variables	Definition	Specification Value (mm)	Tolerance (mm)
x_1	1. Hole Diameter	30	±0,2
x_2	2. Hole Diameter	30	±0,2
x_3	Large Outside Diameter	210	±0,5
x_4	Distance Between Holes	230	±0,2
x_5	Cheek Height	21	±0,5
x_6	Cheek Outer Diameter	180	±0,1
x_7	Cheek Inner Diameter	140	±0,5
x_8	Cover Wall Thickness	27	±0,5

Table 3. Correlation matrix between variables (initial case)

		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1: 1.hole diameter	P.C	1	-0,75**	0,01	0,63**	0,048**	-0,004	0,062**	0,007
	Sig.		0,000	0,881	0,000	0,000	0,520	0,000	0,264
x_2: 2.hole diameter	P.C		1	-0,002	0,18**	-0,008	-0,001	-0,048**	0,004
	Sig.			0,725	0,003	0,195	0,852	0,000	0,562
x_3: large outer diameter	P.C			1	-0,010	-0,008	0,001	-0,001	0,008
	Sig.				0,104	0,184	0,874	0,881	0,183
x_4: distance between holes	P.C				1	-0,019**	-0,009	-0,038**	0,006
	Sig.					0,002	0,162	0,000	0,344
x_5: cheek height	P.C					1	-0,21**	0,037**	-0,004
	Sig.						0,001	0,000	0,508
x_6: cheek outer diameter	P.C						1	-0,011	0,001
	Sig.							0,065	0,907
x_7: cheek inner diameter	P.C							1	-0,003
	Sig.								0,606
x_8: cover wall thickness	P.C								1
	Sig.								

** Correlation significant at 0.01 level

- **Linearity:** Pearson coefficient was calculated for binary variables to test the linearity assumption. The evaluation result is summarized in Table 3. As can be seen from the table, it is understood that two of the variables (large outer diameter (x3) and cover wall thickness (x8)) have no relationship with any other variable.

For this reason, there was no need to evaluate it with a multivariate control chart. Since these variables are unrelated, they can be handled separately with univariate control charts. Large outer diameter (x_3) and cap wall thickness (x_8) variables were removed from the data set and the linearity assumption was repeated for six variables. The pearson correlation coefficient (P.C) values calculated to evaluate the relationships between six variables are shown in Table 4. When the significance levels of the remaining six variables are examined, they are generally seen to be significant, that is, there is a linear relationship.

- **Assumption of suitability for multivariate normal distribution:** The normal distribution suitability test results obtained for 6 quality variables are given in Table 5. It can be said that the p value for all variables is greater than 0.05 and therefore all variables individually comply with normal distribution.

Table 4. Correlation matrix between variables (final situation)

		x_1	x_2	x_3	x_4	x_5	x_6
x_1: 1.hole diameter	P.C	1	-0,75**	0,63**	0,048**	-0,004	0,062**
	Sig.		0,000	0,000	0,000	0,520	0,000
x_2: 2.hole diameter	P.C		1	0,18**	-0,008	-0,001	-0,048**
	Sig.			0,003	0,195	0,852	0,000
x_3: distance between holes	P.C			1	-0,019**	-0,009	-0,038**
	Sig.				0,002	0,162	0,000
x_4: cheek height	P.C				1	-0,21**	0,037**
	Sig.					0,001	0,000
x_5: cheek outer diameter	P.C					1	-0,011
	Sig.						0,065
x_6: cheek inner diameter	P.C						1
	Sig.						

** Correlation significant at 0.01 level

Table 5. Univariate normal distribution results

Test	Variable	KS value	P value
Kolmogorov-Smirnov	x_1	0,004	>0,150
Kolmogorov-Smirnov	x_2	0,004	>0,150
Kolmogorov-Smirnov	x_3	0,004	>0,150
Kolmogorov-Smirnov	x_4	0,005	>0,150
Kolmogorov-Smirnov	x_5	0,003	>0,150
Kolmogorov-Smirnov	x_6	0,002	>0,150

After proving the suitability of all variables for univariate normal distribution, multivariate normal distribution in which all variables were evaluated together was examined. Multivariate normal distribution results evaluated with Henze-Zirkler's test and the Q-Q chart are shown in Table 6.

Table 6. Multivariate normal distribution test results

Test	Variable	P value	Normality
Henze-Zirkler	$x_1 \ldots x_6$	0,4539	Yes

As seen in the table, six variables were found to be suitable for multivariate normal distribution.

- **No Autocorrelation Assumption:** This assumption was tested using the Box-Ljung statistic of the time independence of the variables. It was observed that there was no autocorrelation for all six variables. As a result, it has been determined that the Hotelling T^2 control chart is suitable for the assumptions.

For the measurement results of six variables on 26697 units, T^2 values were calculated using Equation (1) and UCL value was calculated using Equation (2). The Hotelling T^2 control chart created according to T^2 values is shown in Figure 5. The upper control limit of the control chart was found to be 30.1. It can be seen that the T^2 values of 25893 samples are between UCL and 0, while 804 samples are outside the UCL.

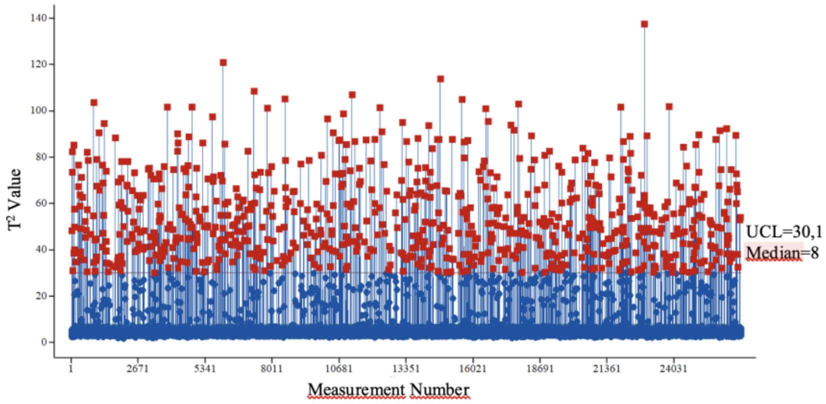

Figure 5. Hotelling T^2 Control Chart

5.3. Identification variables that cause uncontrolled situations with the MYT method

For each of the 804 samples that signaled that the process was out of control, the variable(s) causing the out-of-control situation were determined using Equations 4-6 of the MYT decomposition method. The results of MYT decomposition method implementation for six variables of 10 samples selected from 804 samples are as shown in Table 7.

"0" in the last column of the table represents variables that are under control, and "1" represents variables that cause an out-of-control situation. For example; variable class “100001”; It means that the out-of-control situation occurs due to the variables x_1 and x_6, while the other variables remain within the control limit according to the calculated threshold value and do not affect the out-of-control situation. Since there are 6 variables evaluated, there are $(2^6 - 1) = 63$ possible out-of-control situations (Niaki and Abbasi, 2005). The number of samples in which these possible out-of-control situations were observed in the examined data set is as shown in Table 8.

Table 7. MYT unconditional part T^2 values and out-of-control situations

Sample No	T_1^2	T_2^2	T_3^2	T_4^2	T_5^2	T_6^2	Condition
19	9,818	0,298	0,560	0,272	2,978	35,053	100001
27	1,141	36,305	9,005	25,855	1,093	0,099	011100
38	24,641	6,589	11,685	0,704	8,655	31,180	111011
52	17,048	0,883	6,350	0,333	0,787	5,257	101001
92	1,094	6,014	27,093	0,760	2,775	0,421	011000
95	0,252	1,093	0,486	12,159	33,232	38,175	000111
109	0,575	0,509	15,589	15,671	11,380	14,084	001111
110	6,452	0,848	10,892	0,354	13,798	16,987	101011
115	4,532	0,640	29,088	0,894	3,802	1,148	101000
175	0,686	0,735	10,161	21,790	0,879	1,049	001100

Table 8. Out-of-control situations and number of samples encountered

Condition	Number of Sample	Condition	Number of Sample	Condition	Number of Sample
000001	14	010110	5	101011	5
000010	4	010111	12	101100	10
000011	7	011000	24	101101	10
000100	17	011001	20	101110	5
000101	21	011010	5	101111	7
000110	10	011011	5	110000	12
000111	13	011100	11	110001	16
001000	18	011101	10	110010	7
001001	30	011110	11	110011	2
001010	8	011111	3	110100	16
001011	14	100000	8	110101	13
001100	30	100001	15	110110	5
001101	23	100010	12	110111	3
001110	11	100011	11	111000	12
001111	10	100100	100	111001	13
010000	8	100101	25	111010	3
010001	23	100110	5	111011	7
010010	10	100111	4	111100	13
010011	5	101000	13	111101	9
010100	16	101001	15	111110	8
010101	14	101010	8	111111	0

There is no situation in which all variables have an impact on the out-of-control situation, expressed by the “111111” variable class. For this reason, 62 different out-of-control situations will be considered in the data set. Variables x_1 and x_4 (100100) cause 100 of the 804 out of control situations, which is the most common situation, to be out of control. The least common out-of-control situations belong to the variable classes 011111, 110111 and 111010, with 3 samples each. Out-of-control situations will be called classes in the following sections of the study.

5.4. Development and Implementation Proposed Model Based on Ensemble Algorithm

5.4.1. Create a dataset

While the input data set consists of measurement values of the samples collected from the process, the output data set is the classes that express the variables that cause out-of-control situations obtained as a result of the calculations made in the previous steps. The number of samples for 62 classes varies between 3 and 100. This situation creates an unbalanced data set in terms of sample numbers between classes. Since real data was used in order not to affect the classification accuracy, synthetic data was produced with the help of the SMOTE method, using the highest number of samples as 100, to complete 100 samples for all classes. Thus, we continued with 6200 data belonging to 62 uncontrolled classes.

5.4.2. Implementation of single machine learning algorithms

When basic machine learning algorithms are used single, the parameters that provide the best classification performance are estimated heuristic, taking into account the preliminary information of the data set. The models were redesigned and trained according to each parameter and the results were obtained. A comparison of the success rates obtained from the algorithms was made by determining the appropriate parameter values. Cross-validation method was used for the training phase of the models established with classification algorithms. Cross-validation is a statistical method used to evaluate and compare learning algorithms by dividing data into two parts, one used to learn or train a model and the other used to validate the model (Refaeilzahed et al., 2009). In k-fold cross validation, the data is first divided into k equal sized partitions. Then, a selected partition test set is considered as the remaining k-1 partition training set. In the next phase, a different section is selected for testing and the remaining ones form the training set. The cluster to be selected does not have a priority or importance, each section is of equal importance. This process is repeated k times, each time with a different subsection test set, so that each section is used for both testing and training. In order to ensure consistency of the study, all models were trained using the same parameters. For the number of folds, the value "10", which is frequently used in studies (Refaeilzadeh et al., 2009; Zhang et al., 2019; Jonathan et al., 2019; Karimi et al., 2015; Ramezan et al., 2019; Yu and Feng, 2014.), was taken. Additionally, the sampling type was selected automatically and folding sampling was used because the result values were nominal. Multi-class performance criteria were used to evaluate the classification success rates obtained using the parameter values determined for all algorithms. Rapidminer Studio 9.6 Program was used in all analyses.

***DT algorithm*:** Some of the parameters used for the DT algorithm are shown in Table 9. Similar to previous studies for the splitting process in the tree (Dreiseitl et al., 2001; Anwar et al., 2014), the criterion for selecting the attributes was determined as information gain, which calculates the entropy and selects the least valuable one as the splitting criterion. The maximum depth value was selected as 20 by trying 31 values between 0-30. The confidence level was selected by performing 11 trials in 0.1 step increments between 0 and 1. For the values of other parameters, the program was run with default values.

Table 9. DT parameters

Parameter	Value
Criterion	Information gain
Maximum depth	20
Confidence level	0,1
Min. earnings	0,1

K-NN algorithm: The number of nearest neighbors (k) used for classification was determined as 3, which gives the highest performance, by trying odd numbers between 1-13, as shown in Table 10. Since the accuracy rate remained constant until k=9 and then started to decrease, k=3 was taken as the first highest value among 7 trials.

Table 10. Performance values according to K-NN k parameters

k	Accuracy Rate
1	86,53%
3	**88,85**%
5	88,85%
7	88,85%
9	88,85%
11	88,74%
13	86,53%

The measurement type parameter used to detect the nearest neighbors was chosen as numerical measurements since the data set contains numerical values and Euclidean distance because it is the most frequently used distance type (Hu et al., 2016). The parameters used for K-NN are shown in Table 11.

Table 11. K-NN algorithm parameters

Parameter	Value
K	3
Measurement type	Numerical Measures
Mixed Measure	Euclidean Distance

NB algorithm: Classification is made based on only one parameter, Laplace correlations, there are no other parameters (Anwar et al., 2014).

Multi-class support vector machine algorithm: Since the process discussed in the study is multi-class, the M-SVM algorithm was used. For classification, a one-versus-one approach of the multi-class support vector was used, which has proven successful in the work of Du et al. (2012). The type of kernel function was determined as a radial basis function, taking into account past studies (Du et al., 2012; Farhan et al., 2014; Lu et al., 2011; Onel et al., 2019) and the data set structure. Other parameters were run with the program's default values. The parameters of the M-SVM algorithm are as shown in Table 12.

Table 12. M-SVM algorithm

Parameter	Value
SVM approach	One-versus-one
Kernel Type	Radial basis function

Artificial Neural Networks Algorithm: Feed-forward back-propagation multilayer perceptron neural network has been determined to be suitable from studies in the literature (Aparisi et al., 2006; Niaki and Abbasi, 2005; Salehi et al., 2012). In the network structure, there are input consisting of six variables, two hidden layers containing 100 neurons each, and 62 outputs consisting of classes. The parameters used for the neural network are shown in Table 13. As in classification and prediction studies, the activation function was used as sigmoid (Chen and Wang, 2004; Yu et al., 2009; Maleki and Amiri, 2015). In neural networks, the weight of each connection is updated to reduce the value of the error function. Using the training cycle parameter, the number of times this process should be repeated was tried 7 times, every 50 units in the range of 200-500, and was determined as 500. Learning rate and other parameters were used assuming default values (Shao and Lin, 2019).

Table 13. ANN algorithm parameters

Parameter	Value
Activation function	Sigmoid
Training Cycle	500
Learning rate	0,01

5.4.3. Performance evaluation of single machine learning algorithms

The performances of the five basic machine learning algorithms are shown in Figure 6. When the results are compared, it is seen that the DT algorithm is the most successful classification algorithm compared to the others. Thus, DT was determined as the basic classification algorithm.

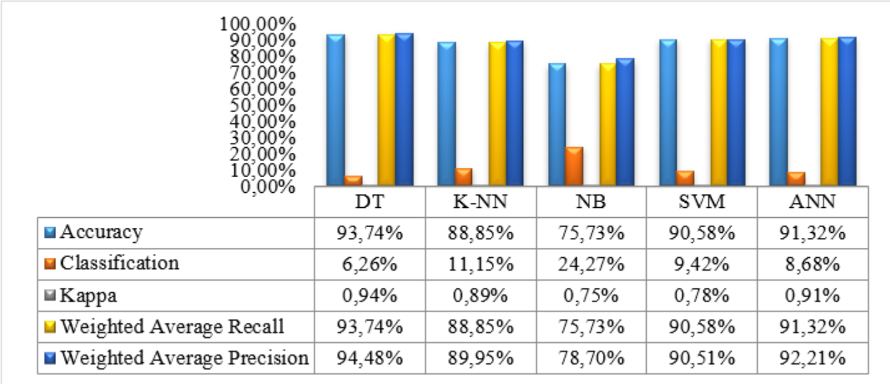

Figure 6. Performance comparison of Single Machine Learning Algorithms

5.4.4. Combining of the selected machine learning algorithm with ensemble methods

The decision tree algorithm, which was selected with the highest classification success among single algorithms, was combined with bagging and boosting methods using previously determined parameters.

Combining of with the bagging ensemble method: The results obtained by combining the DT algorithm in parallel with 10 repetitions are shown in Table 11.

Combining of with the boosting ensemble method: The results obtained by combining the DT algorithm sequentially (with the Adaboost method) in 10 iterations are shown in Table 14.

Table 14. Ensemble methods performance values

Criterion	Adaboost	Bagging
Accuracy	95,08%	94,97%
Classification Error	4,92%	5,03%
Kappa	0,950%	0,949%
Weighted Average Sensitivity	95,08%	94,97%
Weighted Average Precision	95,56%	95,46%

When the results are examined, it is seen that combining the decision trees sequentially with the Adaboost method increases the accuracy.

5.4.5. Ensemble of Ensemble Model

In the stacked generalization method, which has a different working principle from the two methods, different types of classification algorithms are combined sequentially. The model of the study is formed by combining DT-Bagging and DT-Adaboost ensemble algorithms. Performance values are shown in Table 15.

Table 15. Stacked generalization performance values

Criterion	Value
Accuracy	98,06 %
Classification Error	1,94 %
Kappa	0,980 %
Weighted Average Recall	98,06 %
Weighted Average Precision	98,27 %

5.4.6. Performance Evaluation of the Proposed Model

The classification performances obtained by combining the DT algorithm single, the ensemble algorithms sequentially and in parallel, and the last combination of the ensemble algorithms are shown in Figure 7. It is seen that the merging process gradually increases the performances. While the classification accuracy was 93.74% when using the DT algorithm alone, DT-bagging was 94.97%, DT-boost was 95.08%, and the accuracy performance of the model created with the stacked generalization method, which was seen as the most successful, was 98.06%. Thus, it can be seen that the developed model has the ability to classify with higher accuracy.

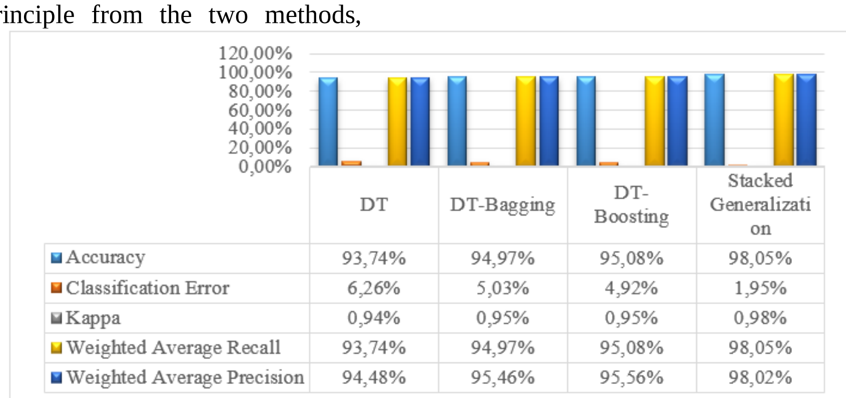

Figure 7. Comparison of the performance of the proposed model with other models

6. Discussion And Conclusion

In order for machine learning algorithms to make accurate predictions, their performance is required to be at the highest level. To achieve this, ensemble machine learning methods have been used. Ensemble algorithms are combined with the stacked generalization algorithm, which is an ensemble method that allows combining different algorithms. The single algorithm was improved by combining the Bagging and Boosting ensemble methods with the single algorithm, and then the two improved methods were combined. The intended target was achieved with the high success rates obtained as a result of the Implementation study carried out to determine the causes of uncontrolled situations in the casting process of the hydraulic pump cover. Thanks to the developed model, it will be possible to predict which variable is the cause in case the newly taken samples are out of control, without the need for multivariate control charts. Thus, faster and more accurate corrective measures can be taken. Great improvements in product quality can be achieved by applying corrective actions not on the product but during the production process.

The stacked generalization combination method used in the developed model has not been encountered before in the field of quality control or in a study on determining the causes of out-of-control situations. The limitation of the study is that only basic machine learning algorithms were used for single algorithm use. As future work, models will be enriched by using different single machine learning algorithms. It is thought that the algorithms will use an optimization technique instead of finding the parameters by trying them intuitively, and the model can be applied to different processes by changing the variables. In addition, accuracy will be evaluated by including feature selection in the study.

Acknowledgement: This study was produced from the doctoral thesis prepared by the first author under the supervision of the second author, which was carried out in the Department of Industrial Engineering, Institute of Natural Sciences, Sakarya University.

References

Agog, N. S., Dikko, H. G., Asiribo, O. E., 2014. Determining out-of-control variable(s) in a multivariate quality control chart. Sci. Africana, 13(2), 266–280.

Agrawal, R., Imielinski, T., 1993. Swami, A., mining association rules between sets of items in large databases. ACM SIGMOD, 1–10.

Ahsan, M., Mashuri, M., Lee, M. H., Kuswanto, H., Prastyo, D. D. 2020. Robust adaptive multivariate Hotelling's T^2 control chart based on kernel density estimation for intrusion detection system. Expert Systems with Applications, 145, 113105.

Alfaro, E., Alfaro, J.L., Gamez M., Garcia N., 2009. A boosting approach for understanding out-of-control signals in multivariate control charts. Int. J. Prod. Res., 47(24), 6821–6834.

Alpaydın, E., 2012. Yapay Öğrenme. 3. Edition. Boğaziçi University, 207-341.

Anwar, H., Qamar, U. Qureshi, A. W. M., 2014. Global optimization ensemble model for classification methods. Sci. World J., 1-9.

Aparisi, F., Avendaño, G., Sanz, J., 2006. Techniques to interpret T^2 control chart signals. IIE Trans., Institute Ind. Eng., 38(8), 647–657.

Asadi, A., Farjami Y., 2019. Online mean shift detection in multivariate quality control using boosted decision tree learning. J. Syst. Manag., vol. 2, 081–106.

Bersimis, S., Sgora, A., Psarakis, S. 2022. A robust meta-method for interpreting the out-of-control signal of multivariate control charts using artificial neural networks. Quality and Reliability Engineering International, 38(1), 30-63.

Bilgin, M., 2018. Veri Biliminde Makine Öğrenmesi Makine Öğrenmesi Teorisi ve Algoritmaları. 2. Edition Papatya Bilim, 31-138.

Blagus, R., Lusa, L., 2013. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 14(16), 1471–2103.

Boullosa, D., Larrabe, J. L., Lopez, A., Gomez M. A., 2017. Monitoring through T^2 Hotelling of cylinder lubrication process of marine diesel engine. Appl. Therm. Eng., 110, 32–38.

Breiman, L. 1996. Bagging predictors. Machine learning, 24, 123-140.

Çetin, S., Birgören B., 2007. Çok değişkenli kalite kontrol çizelgelerinin döküm sanayiinde uygulanmasi. Gazi Üniv. Müh. Mim. Fak. Der., 22(4), 809–818.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res., 16, 321-357.

Chen. L. H., Wang T. Y., 2004. Artificial neural networks to classify mean shifts from multivariate $\chi 2$ chart signals. Comput. Ind. Eng., 47(2–3), 195–205.

Cheng, C. S., Cheng, H. P., 2008. Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines. Expert Syst. Appl., 35(1–2),198–206.

Cheng, C.S., Lee H.T., 2012. Identifying the out-of-control variables of multivariate control chart using ensemble SVM classifiers. J. Chinese Inst. Ind. Eng., 29(5), 314–323.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn., 20(3), 273–297.

Das, N., Prakash, V., 2008. Interpreting the out-of-control signal in multivariate control chart — a comparative study. Int. J. Adv. Manuf. Technol., 37, 966–979.

Dreiseitl, S., Machado, O, L., Kittler, H., Vinterbo, S., Billhardt, H., Binder, M., 2001. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. J. Biomed. Inform., 34(1), 28-36.

Du, S., Lv, J., Xi, L., 2012. On-line classifying process mean shifts in multivariate control charts based on multiclass support vector machines. Int. J. Prod. Res., 50(22), 6288–6310.

Farhan, S., Fahiem, M. A., Tauseef, H., 2014. An ensemble-of-classifiers based approach for early diagnosis of alzheimer’s disease: Classification using structural features of brain images. Comput. Math., Methods Med., 2014.

Gowda, S., Kumar, H., Imran, M., 2018. Ensemble based learning with stacking. Boosting and Bagging for Unimodal Biometric Identification System, 30-36.

Guh, R. S., Shiue Y. R., 2008. An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts. Comput. Ind. Eng., 55(2), 475–493.

Han, J., Kamber, M., Pei, J., 2012. Data mining. concepts and techniques. The Morgan Kaufmann Series in Data Management Systems, 3. Edition.

Hawkins, D. M., 1991. Multivariate quality control based on regression-adiusted variables. Technometrics, 33(1), 61–75.

Hossin, M, Sulaiman, M., N, 2015. A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process, 5(2), 01–11.

Hotelling H., Multivariable quality control—illustrated by the air testing of sample bombsight, McGraw Hill, 111-184, 1947.

Hu, L. Y., Huang, M. W., Ke, S. W., Tsai, C. F., 2016. The distance function effect on k-nearest neighbor classification for medical datasets. Springerplus, 5(1).

Huda, S., Abdollahian, M., Mammadov, M., Yearwood, J., Ahmed S., Sultan I., 2014. A hybrid wrapper-filter approach to detect the source(s) of out-of-control signals in multivariate manufacturing process. Eur. J. Oper. Res., 237(3), 857–870.

Jackson, J. E., 1985. Multivariate quality control. Commun. Stat. Theory Methods, 14(11), 2657–2688.

Jiang, J., Song, H.-M., 2017. Diagnosis of out-of-control signals in multivariate statistical process control based on bagging and decision tree. Asian Bus. Res., 2(2).

Jonathan, O., Omoregbe, N., Misra, S., 2019. Empirical comparison of cross-validation and test data on internet traffic classification methods. Journal of Physics: Conference Series, 1299(1), 1-9.

Joshi, K., Patil, B. 2022. Multivariate statistical process monitoring and control of machining process using principal component-based Hotelling T^2 charts: A machine vision approach. International Journal of Productivity and Quality Management, 35(1), 40-56.

Karimi, S., Yin, J., Baum, J., 2015. Evaluation methods for statistically dependent text. Comput. Linguist., 41(3), 539–548.

Lantz, B., 2013. Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications. Birmingham: Packt Publishing Ltd; 66-343.

Li, J., Jin, J., Shi, J., 2008. Causation-based T^2 decomposition for multivariate process monitoring and diagnosis. J. Qual. Technol., 40 (1), 46–58.

Li, T., Hu, S., Wei, Z., Liao, Z., 2013. A framework for diagnosing the out-of-control signals in multivariate process using optimized support vector machines. Math. Probl. Eng., 2013(2), 1–9.

Lowry, C. A., Woodall, W. H., Champ, C. W., Rigdon, S. E., A multivariate exponentially weighted moving average control chart, Technometrics, 34(1), 46–53, 1992.

Lu, C. J., Shao, Y. E., Li, P. H., 2011. Mixture control chart patterns recognition using independent component analysis and support vector machine. Neurocomputing, 74(11), 1908-1914.

Maimon, L., Rokach, O., 2010. Data mining and knowledge discovery handbook. 2. Edition. Springer London, 165-174.

Maleki, M. R., Amiri, A., 2015. Simultaneous monitoring of multivariate-attribute process mean and variability using artificial neural networks. J. Qual. Eng. Prod. Optim., 1(1), 43–54.

Mason, R. L., Champ, C. W., Tracy, N. D., Wierda, S. J., & Young, J. C. (1997). Assessment of multivariate process control techniques. Journal of quality technology, 29(2), 140-143.

Mason, R. L., Tracy, N. D., Young, J. C., 1995. Decomposition of T^2 for multivariate control chart interpretation. J. Qual. Technol., 27(2), 99–108.

Mitchell, T. M., 2014. Machine learning. McGraw-Hill Science, 52-155.

Mohammed, M., Khan, M. B., Bashier, E. B. M., 2016. Machine learning: Algorithms and applications. 1. Edition. CRC Press, 5-11.

Montgomery D. C., 2009. Introduction to statistical quality control. 6. Edition. John Wiley & Sons, 499-507.

Niaki, S. T. A., Abbasi. B., 2005. Fault diagnosis in multivariate control charts using artificial neural networks. Qual. Reliab. Eng. Int., 21(8), 825–840.

Onan, A., 2018. Particle swarm optimization based stacking method with an application to text classification. Acad. Platf. J. Eng. Sci., 6(2), 134–141.

Onel, M., Kieslich, C. A., Pistikopoulos, E. N., 2019. A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process. AIChE J., 65(3), 992–1005.

Özel, S. 2005. Çok değişkenli kalite kontrolün döküm sanayiinde uygulanması, Master’s Thesis, Kırıkkale University, YOK Thesis Center.

Öztemel E., 2003. Yapay Sinir Ağları. İstanbul, Papatya Yayınları, 7.

Parra, M. G., P. Loaiza, R., 2003. Application of the multivariate T^2 control chart and the Mason Tracy Young decomposition procedure to the study of the consistency of ımpurity profiles of drug substances. Qual. Eng., 16(1), 127–142.

Pei, X., Yamashita, Y., Yoshida, Matsumoto, M., S., 2006. Discriminant analysis and control chart for the fault detection and identification. Comput. Aided Chem. Eng.,21, 1281-1286.

Rakhmawan, S. A., Omar, M. H., Riaz, M., Abbas, N. 2023. Hotelling T^2 control chart for detecting changes in mortality models based on machine-learning decision tree. Mathematics, 11(3), 566.

Ramezan, C. A., Warner, T. A., Maxwell, A. E., 2019. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sens., 11(185), 1-22.

Rao, O. R. M., Subbaiah, K.V., Rao, K. N., Rao T. S., 2013. Application of multivariate control chart for improvement in quality of hotmeal-a case study. Int. J. Qual. Res., 7(4), 623–640.

Refaeilzadeh, P., Tang, L., Liu, H., 2009. C Cross-validation. Springer, Boston, 1-3.

Robert J. C. Y., Mason L., 2002. Multivariate statistical process control with industrial applications. Society for Industrial and Applied Mathematics, 5-17.

Rokach, L., 2010. Ensemble-based classifiers. Artif. Intell. Rev., 33(1–2), 1–39.

Sabahno, H., Amiri, A. 2023. New statistical and machine learning based control charts with variable parameters for monitoring generalized linear model profiles. Computers & Industrial Engineering, 184, 109562.

Salehi, M., Kazemzadeh, R. B., Salmasnia, A., 2012. On line detection of mean and variance shift using neural networks and support vector machine in multivariate processes. Appl. Soft Comput. J., 12(9), 2973–2984.

Shao, Y. E., Lin, S. C., 2019. Using a time delay neural network approach to diagnose the out-of-control signals for a multivariate normal process with variance shifts. Mathematics, 7(10).

Şişci, M., Torkul, Y. E., Selvi, İ. H. 2022. Machine learning as a tool for achieving digital transformation. Knowledge Management and Digital Transformation Power, 55.

Song, H., Xu, Q., Yang, H., Fang, J., 2017. Interpreting out-of-control signals using instance-based bayesian classifier in multivariate statistical process control. Commun. Stat. Simul. Comput., 46(1).

The Royal Society, 2017. Machine learning: the power and promise of computers that learn by example, 5-6.

Ulen, M., Demir, I., 2013. Application of multivariate statistical quality control in pharmaceutical industry. Balk. J. Math.,1, 93–105.

Utgoff, P. E. Berkman, N. C., Clouse, J. A., 1997. Decision Tree Induction Based on Efficient Tree Restructuring. Kluwer Academic Publishers, 29, 5-44.

Woodall W. H., Ncube M. M., Multivariate CUSUM quality-control procedures, technometrics, 27(3), 285–292, 1985.

Yadav, M., Yadav, A., Kumar N., 2015. An introduction to neural network methods for differential equations. Springer.

Yang, W. A., 2015. Monitoring and diagnosing of mean shifts in multivariate manufacturing processes using two-level selective ensemble of learning vector quantization neural networks. J. Intell. Manuf., 26(4), 769–783.

Yılmaz, H., 2012. Çok değişkenli istatistiksel süreç kontrolü: Bir hastane uygulaması, Master’s Thesis, İstanbul Teknik University, YOK Thesis Center.

Yu, J. Bo., Xi, L. Feng., 2009. A neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes. Expert Syst. Appl., 36(1), 909–921.

Yu, Y., Feng, Y., 2014. Modified cross-validation for penalized high-dimensional linear regression models. J. Comput. Graph. Stat., 23(4), 1009-1027.

Zhang, Y., Li, M., Han, S., Ren, Q., Shi, J., 2019. Intelligent identification for rock-mineral microscopic images using ensemble machine learning algorithms. Sensors, 19(9), 1-14.

Zhang, Y., Ma, C., 2012. Ensemble machine learning. Springer US.

Zhou, Z. H., 2012. Ensemble methods: foundations and algorithms Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Taylor & Francis.