ACADEMIC PLATFORM

Academic Platform Journal of Engineering and Science

journal homepage: http://apjes.com/

Application of Machine Learning Methods with Dimension Reduction Techniques for Fault Prediction in Molding Process†

‡1Fatma Demircan Keskin, 2İnanç Kabasakal

1 fatma.demircan.keskin@ege.edu.tr,

2 inanc.kabasakal@ege.edu.tr,

Research Paper | Arrival Date: 31.01.2020 | Accepted Date: 26.04.2020

Abstract

Significant advances in digital technology and advanced analytical tools have had a substantial impact on the production environment and laid the foundation for Industry 4.0 and intelligent production concepts. Predictive engineering is one of the key pillars of smart manufacturing that necessitates the collection and analysis of real-time data with an anticipatory point of view through advanced analytical techniques. In the literature, machine learning-based methods have received a great deal of attention to extract valuable information from process data for fault detection. In this study, fault prediction problem was addressed in a molding process that includes successive steps by applying machine learning methods with dimension reduction techniques. The techniques of Principal Component Analysis (PCA), and Isometric Feature Mapping (Isomap) were first utilized for dimension reduction. Then, the data was analyzed for fault prediction with several machine learning techniques, namely, Support Vector Machine (SVM), Neural Network (NN), and Logistic Regression (LR). The dataset for our analysis includes sensor data captured during the molding process of a wheel rim manufacturer. Several criteria, including accuracy, area under curve (AUC), Type I, and Type II error, were employed to assess the predictive performance of the methods applied, including and the model variants reinforced with PCA and Isomap. Our study demonstrates that all predictive model variants have performed with high accuracy, ranging between 92.16% (LR) and 98.04% (PCA-NN). PCA and Isomap improved the accuracy and Type-I error measures of all models; however, no such improvement was obtained on the Type-II error rates.

Keywords: Dimension Reduction, Fault Prediction, Industry 4.0, Machine Learning, Molding Process

1. INTRODUCTION

Industry 4.0 and smart manufacturing, which have originated based on technological advancements, have enabled more intelligent and interconnected production systems [1]. The intelligence in smart manufacturing stems from data. Collecting real-time, accurate data, analyzing it in a way to provide value, data-driven manufacturing, predictive engineering, and ensuring product quality based on evaluation of real-time product data are among the most critical points of smart manufacturing [2-5].

In the current interconnected and complex manufacturing environments, faults in processes might affect other processes and lead up to significant losses for manufacturers. The use of fault prediction techniques based on real-time processes and machinery data might help to prevent faults, avoid critical breakdowns, and provide insights that help to enhance the utilization of machinery, reduce machine breakdown times, and improve the process and product quality [6]. Real-time monitoring of manufacturing processes and fault prediction has received significant attention both from the practitioners and the researchers over the recent years, depending on the recognition of the magnitude of such possible losses and opportunities [7].

Machine learning methods, especially supervised machine learning methods have stood out in prior research within the context of the failure prediction problem based on the analysis of process data in manufacturing [8,9]. In some of those studies, it has been preferred to apply machine learning methods solely [10,11]. In contrast, some studies have introduced models with a dimension reduction step prior to machine learning methods, mostly due to the complexity of problems and/or availability of an excessive number of variables in process data [12,13].

[1]This paper is an extended version of the paper published in the Proceedings Book of 10th International Symposium on Intelligent Manufacturing and Service Systems, on 9-11 September 2019.

*[1]Corresponding Author: Ege University/Department of Business Administration, Izmir, Turkey, fatma.demircan.keskin@ege.edu.tr, 0(232) 311 22 68

Doi: 10.21541/apjes.682936

In this study, the fault prediction problem is addressed in a wheel-rim manufacturer's molding process with binary classification. In the previous study of Kabasakal et al. [14], the problem was handled with a pure LR model applied to the analysis of 29 process parameters. In this study, the same problem is readdressed by increasing the number of process parameters, expanding the training dataset size, and applying SVM and NN in addition to LR. Moreover, dimension reduction has been carried out considering the potential challenges in modeling due to the complex structure of the molding process that encompasses various sequential sub-processes. For dimension reduction, our approach involves PCA, one of the most frequently used dimension reduction techniques, and Isomap, a technique that has attracted considerable attention in machine learning recently. This study aims to compare the predictive performance of the machine learning models, both standalone and with dimension reduction beforehand. The supervised machine learning methods and dimension reduction techniques employed in our model are introduced in the next two sections. Subsequently, our review of the literature on the fault prediction problem is presented. Afterward, the characteristics of the problem, the initial dataset, and the data preparation steps are explained in detail in the case study section. The findings of the application and their analysis are presented in the application and findings section. Finally, the results are evaluated, and the potential use of machine learning models are discussed within the context of our case.

2. SUPERVISED MACHINE LEARNING

Supervised machine learning methods are widely applied in prediction problems based on the use of process data often obtained as labeled from manufacturing [15]. In this study, popular supervised machine learning methods SVM, NN, and LR have been utilized for fault prediction. In the next subsections, these applied methods have been introduced.

2.1. Support Vector Machine

SVM is a supervised machine learning method capable of handling nonlinear mapping [12] and dealing with complex and large-scale problems. The technique attempts to find an optimal hyperplane separating dependent variables' categories on each side of the plane with a structural risk minimization objective [16, 17]. Nonlinear classification models with SVM often utilize kernel functions. Radial basis function (RBF) is among the most preferred ones [18].

The classical unbiased Least Square SVM (LSSVM) involves optimization of the relationship between inputs Y and outputs Q in the space of feature $Q = \omega^T f(Y) + b$, where ω is the weight vector, and the nonlinear mapping function and the bias vector are denoted by f(.) and b, respectively. The objective function in LSSVM is [12, 19]:

$$\begin{cases} \min J(\omega, \xi) = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{i=1}^{n} \xi_i^2 \\ \quad s.t.\ Q_i = \omega f(y_i) + b + \xi_i \end{cases} \tag{1}$$

where ξ indicates the variance of the error, and $\gamma > 0$ is the penalty coefficient.

The optimal regression function to convert this model to a dual optimization problem and the kernel function F(.) is as follows [12,19]:

$$\begin{cases} Q = \sum_{i=1}^{n} \alpha_i F(Y, y_i) + b \\ F(Y, y_i) = \exp(-\frac{Y - y_i}{2\sigma^2}) \end{cases} \tag{2}$$

where α_i is the Lagrange multiplier.

2.2. Neural Network

NN, one of the popular supervised machine learning methods, is based on network structures that are connected via weighted links [20,21]. NNs are network structures mainly based on computing units called neurons and using activation values and a set of weighted inputs [22]. NNs include a series of interconnected inputs and outputs, in which the weights of the connections are adjusted by the network during the training stage to predict the correct class labels [23]. A multilayer neural network comprises a high number of units connected in a pattern. These units include input units where the information is received for processing, output units containing the processing findings, and hidden units between these two units [24].

The output of each neuron j in the hidden layer is calculated by using a function of activation f as follows [25]:

$$y_j = f(\sum w_{ji} x_i) \tag{3}$$

where w_{ji} denotes the connection weight between units j and i, and x_i represents the input node activation rule.

The resulting weighted sum value is transferred to the activation value of the hidden node using a proper transfer function [26].

2.3. Logistic Regression

LR, a relatively easy-to-apply approach with a wide range of applications, allows the prediction of dependent variables having two or more categories using categorical or continuous independent variables [27].

Since our case involves the prediction of a binary dependent variable, we adopt binary logistic regression in our model. The logit function for this technique is [28]:

$$logit = ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \tag{4}$$

where $\frac{\pi(x)}{1-\pi(x)}$ is odds that ranges from (0;+∞) and $\beta_{0,}\beta_1, \ldots, \beta_p$ are the regression coefficients.

Then the logistic function is obtained by the inverse of the logit function as follows [28]:

$$\pi(x) = \frac{\exp(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)} \quad (5)$$

where $\pi(x)$ ranges from (0;1).

3. DIMENSION REDUCTION

Dimension reduction is frequently applied as an initial step when creating a prediction model to capture patterns in complex data sets and to reduce the difficulties that may arise in modeling [29]. Accordingly, this section briefly introduces two dimension reduction techniques involved in our study, namely PCA and Isomap.

3.1. Principal Component Analysis

PCA, a linear based and multivariate reduction technique, transforms the variables in the initial data set into variables called principal components [30]. The steps of PCA are as follows [13]:

Step 1 Calculation the correlation coefficient matrix (R)

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad (6)$$

$$\mathrm{R} = \mathrm{Cor}(\mathrm{i},\mathrm{j}) = \frac{(\mathrm{n}-1).\mathrm{Cov}(\mathrm{i},\mathrm{j})}{\sqrt{\sum_{\mathrm{k}=1}^{\mathrm{n}}(\mathrm{x_i}(\mathrm{k})-\mu_\mathrm{i})^2 \sum_{\mathrm{k}=1}^{\mathrm{n}}\left(\mathrm{x_j}(\mathrm{k})-\mu_\mathrm{j}\right)^2}}$$

$$= (r_{ij})_{mxn} \quad (7)$$

$$Cov(i,j) = \frac{1}{(n-1)}(x_i-\mu_i)(x_j-\mu_j) \qquad i,j = 1,2,\dots,m \quad (8)$$

where X is a matrix including n rows (samples) and m columns (features), μ_i and μ_j denote X matrix's ith and jth rows averages, respectively.
Step 2 Computation of the eigenvectors (V_i) and eigenvalues (λ_i) of the matrix R.

$$AV_i = \lambda_i V_i \qquad i = 1,2,\dots,m \quad (9)$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, A is a mxn correlation matrix, and V_i is the vector $V_i = [V_{1i,}V_{2i,}\dots,V_{mi}]$.

Step 3 Calculation of a new set of the uncorrelated multivariate vector.

$$X_{new} = V^T.X \quad (10)$$

where X is the original correlated, and X_{new} is the new uncorrelated multivariate vector. V denotes the matrix of eigenvectors. Depending on the selected threshold variance value, all or some components of X_{new} are used as principal components [13].

3.2. Isometric Feature Mapping

Isomap is a nonlinear and noniterative dimension reduction method based on revealing the nonlinear degrees of freedom that lie behind the complex observations [31]. The emergence of machine learning studies that utilize this technique for dimension reduction is relatively more recent [12]. Given a manufacturing multi-parameter set $X = [x_{1,}x_{2,\dots,}x_n]$ with dimension n, a low-dimensional space with dimension k $Y = [y_{1,}y_{2,\dots,}y_k]$; the algorithm consists of the following steps [12, 31, 32]:

Step 1 Determination of the neighbor points on the manifold M. x_i *and* x_j can be considered as neighbors if the distance $d(x_i,x_j)$ between x_i *and* x_j is less than ε or x_i is one of the k nearest neighbors of x_j.

Step 2 Estimation of the geodesic distances between points and defining the graph G. If x_i *and* x_j are neighbors, $d_G(x_i,x_j) = d(x_i,x_j)$, otherwise $d_G(x_i,x_j) = \infty$.

Step 3 Constructing a k-dimensional embedding to the computed distances in Step 2 using multi-dimensional scaling. λ_p is the p-th eigenvalue of the matrix $\tau(D_G) = -HSH/2$. The squared geodesic distances matrix $S_{ij} = [d_G(x_i,x_j)]^2$, and $\mathrm{H} = \mathrm{I} - (1/\mathrm{N})\mathrm{EE}^T$ is the centering matrix. In this matrix, I indicates the N-dimensional unit matrix, and E the vector of all ones. The p-th component of the coordinate vector with k dimension is $y_i = \sqrt{\lambda_p v_p^i}$, where v_p^i denotes p-th eigenvector's i-th component.

4. LITERATURE REVIEW

Industry 4.0 and smart manufacturing have forced companies to an inevitable transformation process. In this transformation, one of the key points is to collect accurate, real-time process and machinery data through sensors and networks, and manage these data in a way to provide value with advanced analytical tools. The objective is to gain the ability to predict the faults or abnormal behaviors, and ultimately to prevent more significant losses [33]. Therefore, continuous monitoring of manufacturing processes and machinery, fault diagnosis and prediction, and application of data-driven methods have received a great deal of attention in the recent literature [34-36]. In the context of fault prediction, various studies have reported substantial results through the use of machine learning methods [37].

In the literature, there have been studies addressing the fault diagnosis and prediction problems by applying SVM, NN, and LR, with and without dimension reduction techniques. In one of the studies implementing machine learning methods solely, Kankar et al. [38] analyzed ball bearing vibration data and developed a classification model to predict faults with NN and SVM. The authors predicted five fault categories by analyzing fourteen features, including process parameters and computed statistical measures using vibration data. Kusiak and Li [39] proposed a machine-learning model with NN and SVM to predict whether faults will occur and their types, if any.

PCA has been a frequently used technique in the studies aiming to predict faults based on process data; on the other hand, Isomap has been increasingly used in recent years. Benkedjouh et al. [32] applied Isomap and a non-linear version of PCA with Support Vector Regression to predict the remaining useful life of a machining tool. Zhang et al. [13] analyzed a dataset that contains equipment vibration signals with PCA and Artificial Neural Network (ANN) for maintenance-related decisions. Sakthivel et al. [40] dealt with the fault diagnosis problem based on vibration signals. They have applied various dimension reduction methods, including PCA, Isomap, Diffusion maps, Local linear embedding, and classified the features obtained from each dimension reduction method by using a decision tree, kNN, Bayes Net, and Naive Bayes. Jin et al. [41] applied dimension reduction methods, including PCA and Isomap, to classify machine faults based on the data sets of machine health conditions. Gao and Hou [34] dealt with the fault diagnosis problem using the Tennessee Eastman process data set; and utilized PCA for dimension reduction before SVM. Bai et al. [12] addressed the problem of quality prediction by applying PCA, ISOMAP, and Locally Linear Embedding with SVM. The authors reported that Isomap-SVM had the best predictive performance among the techniques employed.

In the literature, there have been also studies addressing the fault prediction problem with machine learning-based methods in molding processes that have similar features with the wheel-rim molding process analyzed in this study. Most of these studies have applications in the plastic injection molding process. Sadeghi [42] provided a backpropagation NN model to predict the injected parts' strength by using temperatures of mold and melt, the pressure of injection, and material grade as key variables. One of the other studies addressing the fault prediction in the same process, Ribeiro [43] used more than two fault categories such as unfilled parts, stains, burn marks, warped parts by considering explanatory variables including metering, injection, and cycle times, injection velocity, and cushion. He applied C-SVM and v-SVM classifiers and compared their performances with RBF NN. In other studies that address defect prediction in the plastic injection molding process; Kim et al. [44] proposed a recursive neural network model, Nasiri and Khosravani [45] applied a fuzzy case-based reasoning approach by considering 19 features under temperature, pressure, time, speed, and size categories. Taghizadeh et al. [46] provided an ANN-based prediction model for the warpage of molded parts by analyzing a variety of process parameters such as melt temperature, mold temperature, ejection temperature, thermal conductivity, and specific heat. Additionally, Kabasakal et al. [14] addressed the fault detection problem of a wheel-rim manufacturer's molding process with LR, and our study revisits this case by applying several machine learning techniques, occasionally reinforced with prior dimension reduction.

5. CASE STUDY

This study addresses a fault prediction problem through a dataset, including actual measurements captured during the molding process in a wheel rim manufacturer. The data recorded involves measurements regarding the faults and a list of parameters that can affect the process quality. The molding process consists of a series of sub-processes, including pouring the melted aluminum alloy into a tank that is kept at a certain temperature, pressurizing the tank, blowing out the melted alloy, pouring it into a mold and cooling the alloy in the mold.

The initial dataset involves 137 parameters for 594 products manufactured on the same casting machine by using a single mold for 5 days. The fault rate of the analyzed batch is 6.9% (41 units).

The faults defined by the company are in three forms: Start-up scrap, shrinkage, and visual scrap. In this study, binary classification is used, and a unit is classified as "faulty" or "normal". Most of the measurements in the dataset relate to the steps of the cooling phase. The company recorded both the ideal and actual temperatures/ pressures for many of the steps. As in [14], we consider the deviations from ideal values as potential causes for faults; thus, those differences are also included in the dataset for analysis.

Another step before our analyses was the removal of redundant columns. Several parameters within the dataset were excluded since they involved constant values in all records. As a result, our final data set involved 594 rows with 32 attributes that consist of 31 inputs and a class variable.

Table 1. Descriptions of the discrete variables in the dataset

No	Variable Name	Categories
1	Shift	1-2-3
2	aluminium_transfer	0 – 1 – 5
3	phase_2_time	25-35-40
4	phase_4_time	23- 203-210

Table 2. Descriptions of the continuous variables in the dataset

No	Variable Name	Range
1	Hour (hr)	[0-23]
2	Humidity (hmd)	[21.8-34.2]
3	Temperature (tmp)	[29.4-52.2]
4	metal_temperature_diff (mtd)	[-11-4]
5	maximum_pressure_value (mpv)	[5.8-7.2]
6	cooling_temperature (wt)	[21-30]
7	cycle_time (ct)	[295-37009]
8	thermocouple_1_diff (t1d)	[-113-55]
9	thermocouple_2_diff (t2d)	[-192-29]

10	thermocouple_4_actual_value (t4a)	[468.79-509]
11	thermocouple_5_actual_value (t5a)	[445.5-526]
12	phase_1_pressure (ph1p)	[131.3-149.5]
13	phase_2_pressure (ph2p)	[301.3-516.9]
14	h1_flow_average_value (h1f):	[0-1205.2]
15	h2_flow_average_value (h2f)	[0-602.4]
16	h3_flow_average_value (h3f)	[0-624.8]
17	h7_flow_average_value (h7f)	[1200.6-1204.8]
18	h8_flow_average_value (h8f)	[785.2-798.4]
19	h9_flow_average_value (h9f)	[795.9-800.8]
20	h10_flow_average_value (h10f)	[795.8-800.6]
21	h11_flow_average_value (h11f)	[0-770.6]
22	h12_flow_average_value (h12f)	[0-604.8]
23	s1.1_side_core_cooling_flow_diff (s1.1sd)	[-16-0.4]
24	s1.2_side_core_cooling_flow_diff (s1.2sd)	[-16-0.1]
25	s1.3_side_core_cooling_flow_average_value (s1.3sa)	[0-15.4]
26	s1.4_side_core_cooling_flow_average_value (s1.4sa)	[0-16.3]
27	s2_bottom_core_cooling_flow_average_value (s2ba)	[0-17]

6. APPLICATION AND FINDINGS

In this study, fault prediction problem in molding process is addressed. A challenge in our problem was the high number of process parameters that complicate modeling for prediction. In such cases, dimension reduction is often involved as an initial step to construct prediction models over complex datasets [29]. In accordance, our approach involves the application of PCA and Isomap for dimension reduction prior to predictive modelling with LR, SVM, and NN techniques. The results are compared with the individual application of LR, SVM, and NN to evaluate the effects of these dimension reduction techniques on the predictive performance. As detailed in the previous section, each row in our dataset has 31 attributes that represent potential causes of faults, accompanied by a final attribute that denotes a class.

3/4 of the total dataset is randomly separated as the training set and the remaining part (1/4 of the total dataset) as the testing set to evaluate the considered methods' performances. Features of the training and testing sets are given in Table 3.

Table 3. Features of the training and testing sets

Training set size	441
Testing set size	153
Number (percentage) of faults in training set	30 (6.80%)
Number (percentage) of faults in testing set	11 (7.19%)

The performances of all model variants are presented in Tables 4 and 5. The Isomap has been run in R, all the other models in SPSS Modeler. The 31 variables considered in the prediction models have been reduced to 5 dimensions by the PCA and 9 dimensions by the Isomap. In the implementation of SVM, RBF is used as the kernel function.

Table 4. Comparison of the Prediction Performances of the Models

Model	**Performance**	
	Overall accuracy	**Area Under Curve**
SVM	96.08%	0.95
NN	95.43%	0.83
LR	92.16%	0.90
PCA-SVM	96.73%	0.85
PCA-NN	98.04%	0.96
PCA-LR	97.39%	0.88
ISOMAP-SVM	97.39%	0.90
ISOMAP-NN	97.39%	0.93
ISOMAP-LR	94.12%	0.93

Table 5. Comparison of False Results of the Models

Model	**Type I Error**	**Type II Error**	**Total false**
SVM	4/142 (2.82%)	2/11 (18.18%)	6/153 (3.92%)
NN	3/142 (2.11%)	4/11 (36.36%)	7/153 (4.58%)
LR	8/142 (5.63%)	4/11 (36.36%)	12/153 (7.84%)
PCA-SVM	1/142 (0.70%)	4/11 (36.36%)	5/153 (3.27%)
PCA-NN	1/142 (0.70%)	2/11 (18.18%)	3/153 (1.96%)
PCA-LR	0/142 (0.00%)	4/11 (36.36%)	4/153 (2.61%)
ISOMAP-SVM	0/142 (0.00%)	4/11 (36.36%)	4/153 (2.61%)
ISOMAP-NN	0/142 (0.00%)	4/11 (36.36%)	4/153 (2.61%)
ISOMAP-LR	5/142 (3.52%)	4/11 (36.36%)	9/153 (5.88%)

The results indicate that applying PCA or Isomap before SVM, NN, and LR models have improved the prediction accuracy of these models' single applications. PCA-NN has the lowest total false rate (1.96%) (highest overall accuracy (98.04%)) among all considered models, followed by PCA-LR, ISOMAP-SVM, and ISOMAP-NN with 2.61% total false rate (97.39% overall accuracy). Also, LR model performance has increased compared to [14] when considering more process parameters and larger training dataset size (overall accuracy increased from 90.50% to 92.16%, and Type II error reduced from 46.20% to 36.36%). The occurrence probability of Type II error, which means incorrectly predicting a fault part to be a non-fault part is the lowest in PCA-NN and SVM, but still at a high rate of

18.18%. PCA-NN also has the highest performance in AUC criterion.

Additionally, the k-fold cross-validation, one of the most widely applied approaches to assessing models' prediction performance and validity [47], has been carried out to evaluate the validity of all model variants. By taking into consideration of the dataset size and number of faulty observations, it is preferred to divide the whole dataset into 3 groups, including an approximately equal number of faulty observations. In Table 6, 3-fold cross-validation results of all model variants are given.

Table 6. k-fold Cross Validation Test Results (k=3)

Model	Performance		
	Overall accuracy	**Type I Error**	**Type II Error**
SVM	95.45%	0.36%	61.09%
NN	94.44%	2.36%	47.46%
LR	92.93%	4.13%	45.83%
PCA-SVM	96.11%	0.18%	51.58%
PCA-NN	95.62%	1.45%	43.07%
PCA-LR	95.79%	0.36%	54.09%
ISOMAP-SVM	95.29%	1.27%	51.28%
ISOMAP-NN	94.95%	0.54%	65.26%
ISOMAP-LR	94.28%	1.99%	54.98%

3-fold cross-validation results imply that the overall accuracy of all model variants is between 92.93% (LR) and 96.11% (PCA-SVM). PCA has improved the performances of all model variants in terms of overall accuracy and Type I error rate. PCA has also decreased the Type II error rate in predictions with SVM and NN.

Our findings on the benefits of dimension reduction with Isomap were less apparent. Isomap has not improved the predictive performance of SVM in terms of overall accuracy and Type I error rate. However, the overall accuracy of NN and LR have been increased, with a reduction in Type I error rate. In contrast, Type II error rate has been found higher with Isomap dimension reduction. In fact, Type II error rate was quite high in all models, with a range of 43.07% (PCA-NN) to 65.26% (ISOMAP-NN).

7. CONCLUSION

Real-time collection and utilization of process data for predictive analysis is an essential prerequisite of the manufacturing concept in the era of Industry 4.0. This paper aims to compare both the individual application performances of machine learning methods for fault prediction in the molding process. Moreover, our study explores the benefits of linear and non-linear based dimension reduction in the predictive performance of those methods. Specifically, our study employs machine learning methods of SVM, NN, and LR; and dimension reduction techniques of PCA and Isomap.

The case examined in this study is the problem of a wheel-rim manufacturer that contains multiple process parameters and fault categories for 594 products obtained from the molding process. The dimensionality of the initial dataset was reduced to 31 parameters in data preprocessing. Three predefined fault types come out as the result of the molding process, namely, the start-up scrap, shrinkage, and visual scrap. However, our primary objective was limited to predict the occurrence of a fault, independent from its type. The techniques employed in the model have been trained and tested by randomly picking ¾ and ¼ of our dataset, respectively.

To compare the performances of the models applied in this study; prediction accuracy, AUC, Type I and Type II error measures are used. Prediction accuracy of the models is obtained within a high range of 92.16% (LR) and 98.04% (PCA-NN), where PCA and Isomap led to an increase in the accuracy of all predictive models. PCA and Isomap have also reduced the Type I error rate for all models, but they do not have the same performance at the models' Type II errors. 3-fold cross-validation test has also provided similar results. AUC, which is an important measure of the extent to which models can distinguish fault categories, is obtained in the range of 0.83 (NN) to 0.96 (PCA-NN). PCA was only able to increase the AUC of NN, while Isomap was able to improve the AUC of NN and LR.

The primary finding of our study demonstrates that all model variants evaluated in this study have had remarkable overall prediction accuracy. However, the Type II errors of their predictions are considerably high. Nevertheless, the models might further be tested on larger datasets that include sufficient fault observations from the same process, to reduce the Type II error.

REFERENCES

[1] T. Niesen, C. Houy, P. Fettke, and P. Loos, "Towards an integrative big data analysis framework for data-driven risk management in industry 4.0," 49th Hawaii International Conference on System Sciences (HICSS), January 2016, 5065-5074. IEEE.
[2] Y. Oh, K. Ransikarbum, M. Busogi, D. Kwon, and N. Kim, "Adaptive SVM-based real-time quality assessment for primer-sealer dispensing process of sunroof assembly line," Reliability Engineering & System Safety, vol. 184, pp. 202–212, 2019.
[3] Q. Qi, and F. Tao, "Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison,", Ieee Access, vol. 6, pp. 3585-3593, 2018.
[4] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," Journal of Manufacturing Systems, vol. 48, pp. 157-169, 2018.
[5] A. Kusiak, "Smart manufacturing," International Journal of Production Research., vol. 56, no.1-2, pp. 508-517, 2018.

[6] H. N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," Enterprise Information Systems, pp. 1-25, 2019.
[7] Z. Li, Y. Wang, and K.-S. Wang, "Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario," Advances in Manufacturing, vol. 5, no. 4, pp. 377–387, 2017.
[8] S. Hou, and Y. Li, "Short-term fault prediction based on support vector machines with parameter optimization by evolution strategy," Expert Systems with Applications, vol. 36, no. 10, pp. 12383-12391, 2009.
[9] I. Santos, J. Nieves, Y. K. Penya, and P. G. Bringas, "Optimising Machine-Learning-Based Fault Prediction in Foundry Production," Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living Lecture Notes in Computer Science, pp. 554–561, 2009.
[10] Q. Zhu, Y. Jia, D. Peng, and Y. Xu, "Study and Application of Fault Prediction Methods with Improved Reservoir Neural Networks," Chinese Journal of Chemical Engineering, vol. 22, no. 7, pp. 812–819, 2014.
[11] H.-Q. Wang, Y.-N. Cai, G.-Y. Fu, M. Wu, and Z.-H. Wei, "Data-driven fault prediction and anomaly measurement for complex systems using support vector probability density estimation," Engineering Applications of Artificial Intelligence, vol. 67, pp. 1–13, 2018.
[12] Y. Bai, Z. Sun, B. Zeng, J. Long, L. Li, J. V. de Oliveira, and C. Li, "A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction," Journal of Intelligent Manufacturing, vol. 30, no. 5, pp. 2245-2256, 2019.
[13] Z. Zhang, Y. Wang, and K. Wang, "Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks," The international journal of advanced manufacturing technology, vol. 68, no.(1-4), pp.763-773, 2013.
[14] İ. Kabasakal, F. Demircan Keskin, A. Koçak, and H. Soyuer. "A Prediction Model for Fault Detection in Molding Process Based On Logistic Regression Technique," in Proceedings of the International Symposium for Production Research 2019, August 28-30, 2019. Wien, Austria.
[15] T. Wuest, D. Weimer, C. Irgens, and K. D. Thoben, "Machine learning in manufacturing: advantages, challenges, and applications," Production & Manufacturing Research, vol. 4, no. 1, pp. 23-45, 2016.
[16] R. Malhotra. "Comparative analysis of statistical and machine learning methods for predicting faulty modules," Applied Soft Computing, vol. 21, pp. 286-297, 2014.
[17] D. Thukaram, H. P. Khincha, and H. P. Vijaynarasimha, "Artificial neural network and support vector machine approach for locating faults in radial distribution systems," IEEE Transactions on Power Delivery, vol. 20, no. 2, pp. 710-721, 2005.
[18] J. Wang, X. Wu, and C. Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems," International Journal of Business Intelligence and Data Mining, vol. 1, no. 1, pp. 54-64, 2005.
[19] G. Rubio, H. Pomares, I. Rojas, and L. J. Herrera, "A heuristic method for parameter selection in LS-SVM: Application to time series prediction International Journal of Forecasting, vol. 27, no. 3, pp. 725-739, 2011.
[20] S. B. Alvi, R. Martin, and J. Gottschling, "Efficient Use of Hybrid Adaptive Neuro-Fuzzy Inference System Combined with Nonlinear Dimension Reduction Method in Production Processes", Computer Science & Information Technology (CS & IT), vol. 7, pp. 29-43, 2017.
[21] M. W. Craven, and J. W. Shavlik, Using neural networks for data mining," Future generation computer systems, vol.13, no. (2-3), pp. 211-229, 1997.
[22] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," Neural Networks, vol. 111, pp. 47-63, 2018.
[23] J. Han, J. Pei, and M. Kamber, "Data mining: concepts and techniques," Elsevier, 2011
[24] S. B. Kotsiantis, I. D. Zaharakis, P. E. Pintelas, Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, vol. 26, no. 3, pp. 159-190, 2006.
[25] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," Expert Systems with Applications, vol. 38, no. 10, pp. 13475-13481, 2011.
[26] S. N. Oğulata, C. Şahin, and R. Erol, "Neural network-based computer-aided diagnosis in classification of primary generalized epilepsy by EEG signals," Journal of medical systems, vol. 33, no. 2, pp. 107-112, 2009.
[27] J. E. King, "Binary logistic regression," in Best Practices in Quantitative Methods, J.W.Osborne, Ed., Sage Publications, USA, pp. 358-384, 2008.
[28] F. S. De Menezes, G. R. Liska, M. A. Cirillo, and M. J. Vivanco, "Data classification with binary response through the Boosting algorithm and logistic regression," Expert Systems with Applications, vol. 69, pp. 62-73, 2017.
[29] D. Lieber, M. Stolpe, B. Konrad, J. Deuse, and K. Morik, "Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning," Procedia CIRP, vol. 7, pp.193-198, 2013
[30] H. Abdi, and L. J. Williams, "Principal component analysis," Wiley interdisciplinary reviews: computational statistics, vol. 2, no. 4, pp. 433-459, 2010.
[31] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, vol. 290.5500, 2319-2323, 2000.
[32] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," Engineering Applications of Artificial Intelligence., 26(7), 1751-1760, 2013.
[33] A. Diez-Olivan, J. Del Ser, D. Galar and B. Sierra, "Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0," Information Fusion, vol. 50, no. 92-111, 2019.

[34] X. Gao and J. Hou, "An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process," Neurocomputing, vol. 174, pp. 906–911, 2016.
[35] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," Neurocomputing, vol. 174, pp. 643–650, 2016.
[36] H. Lahdhiri, M. Said, K. B. Abdellafou, O. Taouali and M. F. Harkat, "Supervised process monitoring and fault diagnosis based on machine learning methods," The International Journal of Advanced Manufacturing Technology, vol. 102, no. (5-8), pp. 2321-2337, 2019.
[37] W. Mao, L. He, Y. Yan, and J. Wang, "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine," Mechanical Systems and Signal Processing, vol. 83, pp. 450–473, 2017.
[38] P. K. Kankar, S. C. Sharma and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," Expert Systems with applications, vol. 38, no. 3, pp. 1876-1886, 2011.
[39] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," Renewable Energy, vol. 36, no. 1, 16-23, 2011.
[40] N. R. Sakthivel, B. B. Nair, M. Elangovan, V. Sugumaran, and S. Saravanmurugan, "Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals, Engineering Science and Technology, an International Journal, vol. 17, no.1, pp. 30-38, 2014.
X. Jin, F. Yuan, T. W. Chow, and M. Zhao, "Weighted local and global regressive mapping: A new manifold learning method for machine fault classification," Engineering Applications of Artificial Intelligence, vol. 30, pp.118-128, 2014.
[41] B. H. M. Sadeghi. "A BP-neural network predictor model for plastic injection molding process," Journal of materials processing technology, vol. 103, no. 3, pp. 411-416, 2000.
[42] B. Ribeiro, "Support vector machines for quality monitoring in a plastic injection molding process. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 3 no. 3, pp. 401-410, 2005.
[43] S. Kim, and S. Kim, K. R. Ryu, "Deep Learning Experiments with Skewed Data for Defect Prediction in Plastic Injection Molding," IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), October 2018, 1-2. IEEE.
[44] S. Nasiri, and M. R. Khosravani, "Faults and failures prediction in injection molding process", The International Journal of Advanced Manufacturing Technology, vol. 103, no. (5-8), pp. 2469-2484, 2019.
[45] S. Taghizadeh, A. Özdemir, and O. Uluer, "Warpage prediction in plastic injection molded part using artificial neural network" IJSTM, 37(M2), 2013.
[46] T. T. Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition, vol. 48, no. 9, pp. 2839-2846, 2015.