Received 24 September 2023, accepted 6 October 2023, date of publication 12 October 2023, date of current version 18 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324192

RESEARCH ARTICLE

Double Ensemble Technique for Improving the Weight Defect Prediction of Injection Molding in Smart Factories

KWANGMO KOO[1], KEUNHO CHOI[2], AND DONGHEE YOO[3]
[1]Department of Management of Technology, Gyeongsang National University, Jinju 52828, South Korea
[2]Department of Business Administration, Hanbat National University, Daejeon 34158, South Korea
[3]Department of Management Information Systems (Business and Economics Research Institute), Gyeongsang National University, Jinju 52828, South Korea

Corresponding author: Donghee Yoo (dhyoo@gnu.ac.kr)

This work was supported in part by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea under Grant NRF-2020S1A5A2A01042169; and in part by the 'Program to train professionals in spreading convergence technology commercialization' of the Ministry of Trade, Industry and Energy.

ABSTRACT The growing move toward smart factories can leverage industrial big data to enhance productivity. In particular, research is being conducted on injection molding and utilizing machine learning techniques to analyze molding process data, discover optimal molding conditions, and predict and improve product quality. This study aims to identify the key factors influencing the weight defects of injection-molded products and demonstrate the potential use of the double ensemble technique for better prediction accuracy of weight defects. We obtain the key factors influencing weight defects prediction, barrel H2 temp real, metering time, and fill time using gain ratio analysis. Subsequently, we develop single models using machine learning algorithms, including decision tree, random forest, logistic regression, the Bayesian network, and the artificial neural network. Ensemble models, including bagging and boosting and double ensemble models are developed to compare their performance with that of single models. The findings indicate that ensemble models outperform the prediction accuracy of the single models. The double ensemble technique demonstrates the greatest improvements in prediction accuracy over the single models. These results showcase the potential of applying the double ensemble technique to other injection molding areas and suggest that adopting this technique will contribute to establishing other smart factories that will enhance both productivity and cost competitiveness.

INDEX TERMS Double ensemble, ensemble, machine learning, smart factory, injection molding, quality prediction, prediction accuracy.

I. INTRODUCTION

In recent years, the rise of Industry 4.0 and cyber-physical systems (CPS) has brought smart manufacturing to the forefront and led to transformative changes. Smart manufacturing utilizes vast amounts of industrial big data to enable flexible and fully connected factories known as smart factories [1], [2]. They can leverage continuous data streams in their operational and production systems. The adoption of technologies like the Internet of Things (IoT) and CPS has made it feasible to implement smart factories, thereby offering solutions for handling complexity and establishing intelligent products and production processes [3].

Smart factories are being utilized across various industries with the aim of enhancing competitiveness by making production systems intelligent, reducing production costs, and improving productivity. For example, by using IoT sensors to collect and analyze data during the production process, defects can be minimized, and the efficiency of production lines can be increased. This allows for the establishment of a process management system that enables fast and accurate decision-making based on data collected during the

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

production process. Many industries are moving towards digitization to ensure product quality and using new technologies such as machine learning (ML) to model and predict highly complex and nonlinear events when trained with actual experiences. ML is proving crucial in improving industrial productivity by utilizing data collected during product manufacturing, saving time, resources, and energy, or reducing waste to maintain the competitiveness of manufacturing companies [4], [5]. In the field of injection molding, research is being conducted on the optimization of molding conditions and the prediction of product quality using ML techniques based on data from injection molding machines in smart factories. Research is continuing to improve molding conditions and predict product quality using ML and manufacturing condition data from injection molding machines. Injection molding is a widely used plastic molding technique that enables rapid mass production by injecting molten plastic resin into a mold. Typically, the injection molding process involves six stages: mold clamping, injection, press holding, plasticization, cooling, and mold opening. Injection molding is crucial in industries like aerospace, defense, electronics, and telecommunications [6].

Injection molding involves injecting temperature-sensitive plastic resin into a mold under both high temperature and pressure. However, the process is prone to molding defects, such as incomplete filling, flash, warpage, deformation, and black spots, during packing and cooling. The quality of injection-molded products is influenced by various factors. These include raw materials, the machines, process conditions, mold design, and injection environment, all interact in complex ways. Traditionally, optimizing these conditions have relied heavily on the experience of on-site operators. However, individual judgment and skills have made it challenging to achieve consistent quality, productivity, and standardization. Despite the emergence of computer-aided engineering (CAE) and computer-aided design (CAD) for numerical analysis and process parameter optimization [7], [8], [9], relying solely on operator experience in injection molding still remains an important issue. In the field of injection molding, product quality is determined by various conditions, which is why relying solely on the experience and know-how of on-site operators to input and control injection molding conditions on injection molding machines for process optimization has several limitations. It is becoming increasingly important to utilize ML techniques and big data analysis based on production data to ensure process optimization and quality prediction.

To overcome these challenges and advance the production methods for injection molding, there is growing recognition of the significance of employing ML techniques and analyzing big data [10], [11]. ML techniques can identify the key variables that impact product quality by analyzing molding process data, including mold conditions, injection molding conditions, and environmental factors. Further, ML techniques can predict the quality of injection-molded products by analyzing the data generated throughout the production process. Previous research in injection molding has often involved developing single ML models for quality prediction or process optimization [12], [13], [14], [15], [16], [17], [18], [19], [20]. However, the current trend is toward utilizing ensemble model and, combining multiple models to enhance overall prediction accuracy [21], [22], [23], [24].

This study thus further improves the prediction accuracy of classification models in injection molding by applying the double ensemble technique that was introduced in our previous studies [25], [26]. This double ensemble model, applying the ensemble once more to the ensemble model, effectively addresses the overfitting problem by leveraging multiple models that learn from distinct datasets and incorporating diverse information and capturing generalizable characteristics. Although developing and training multiple models does require additional time, the double ensemble technique is expected to provide potential advantages for injection molding, including improved prediction accuracy, improved generalization, and more optimized injection molding processes.

The research methodology utilized involves identifying important variables that affect weight defects of products by analyzing molding process data. ML techniques, such as decision tree (DT), random forest (RF), logistic regression (LR), Bayesian network (BN), and artificial neural network (ANN)-based by multi-layer perceptron (MLP) will also be used to construct a single model that has optimal prediction accuracy. The performance of this model will then be compared to an ensemble model that applies bagging and boosting techniques. Finally, the prediction accuracy of the double ensemble model, applying both bagging and boosting sequentially, will be evaluated to determine whether it outperforms the single or ensemble models. By applying the optimal prediction accuracy derived from these approaches to the production of injection-molded products, it will be possible to reduce quality defects and contribute positively to cost competitiveness through improved productivity. This study thus addresses two research questions (RQs) as follows:

- RQ1: How can the significant variables impacting product weight quality and the evaluation of their prioritized order be identified?
- RQ2: Does the double ensemble model provide more accurate prediction results than the single or ensemble models?

The study is organized as follows: Section II presents a literature review on the use of ML in injection molding and introduces the dataset. Section III outlines the research methodology, including data preprocessing, model construction, and performance evaluation. In Section IV, the prediction accuracy of single, ensemble, and double ensemble models is compared. Finally, Section V concludes the study with a summary, implications of the research, its limitations and new opportunities for future research.

II. LITERATURE REVIEW

A. ML STUDIES ON INJECTION MOLDING

Predicting the quality of molded products and understanding the key influencing factors are a significant concern of injection molding. To address this challenge, extensive research that utilizes ML techniques for product quality prediction is being conducted. This effort involves analyzing the process data collected from equipment to optimize the molding conditions and also predict defects, leading to the development of smart injection molding methods [10], [11].

The study using the single model related to injection molding is thus summarized here. Changyu et al. [12] focused on modeling the complex relationship between process conditions and quality indices of injection-molded parts using the ANN method. Specifically, the researchers aimed to improve the quality index associated with volume shrinkage variation and demonstrate the effectiveness of their approach for optimizing the injection molding process.

Huang et al. [13] proposed a virtual measurement technique that combined real-time multi-quality prediction neural networks, autoencoder networks, and multilayer perceptron networks. This innovative approach enabled a simultaneous prediction of key characteristics, such as width, weight, and residual stress distribution, in injection-molded products.

Silva et al. [14] employed ML techniques, specifically ANN and support vector machine (SVM), to develop an intelligent system for classifying the quality of plastic injection parts. By employing five process variables, they successfully predicted and classified both good and defective parts, by evaluating the prediction accuracy and comparing the performance of the two models.

Ke and Huang [15] conducted injection molding experiments on IC trays, identifying four quality indices (holding pressure index, pressure integral index, residual pressure drop index, maximum pressure index) that highly correlated with the system and cavity pressure curves. They then used these indices as input data and built the MLP neural network model for various types of ANN, so as to predict the 'pass' or 'fail' criteria of finished products. The study revealed a strong correlation between the features extracted from the pressure curves and part quality.

Parizs et al. [16] conducted a study that compared the predictive effectiveness of different ML algorithms for the prediction of quality in porous injection molding. They utilized pressure-based quality metrics as input data and compared the prediction accuracy of four classification algorithms: the k-nearest neighborhood (KNN), naive bayes, linear discriminant analysis, and DT. Of these algorithms, they demonstrated that the DT algorithm was the most accurate in predicting injection molding quality. However, it was important to note that classification predictions may also vary depending on the specific input dataset being used.

Lee et al. [17] focused on injection molding and selected several input parameters, namely melting temperature, mold temperature, injection speed, packing pressure, packing time, and cooling time. They used these parameters to build an ANN model. The output variables considered were the mass, diameter, and height of the molded product. They compared the prediction performance of the ANN model to that of the linear regression and quadratic polynomial regression models.

Ozcelik and Erzurumlu [18] proposed an efficient optimization methodology for injection molding by utilizing ANN in combination with genetic algorithms. Their methods considered various process condition variables, such as mold temperature, melt temperature, packing pressure, packing time, runner type, gate location, and cooling time. They also performed an analysis of variance (ANOVA) study to analyze the influence of these particular variables on the warpage of plastic parts. Their findings revealed that packing pressure, mold temperature, melt temperature, packing time, and cooling time significantly affected warpage in descending order.

Ardestani et al. [19] analyzed eight process parameters (flow rate, melt temperature, mold temperature, holding pressure, runner diameter, gate diameter, gate angle, and included angle) for producing PVC bushings. They employed both the ANOVA and ANN techniques. The focus of their study was on predicting and then modeling blush defect areas. Their findings indicated that flow rate, melt temperature, and runner diameter had a powerful impact on blush defects during the production of PVC bushings.

Injection molding also posed a persistent challenge due to the need to manipulate numerous process parameters in real-time to ensure control over quality characteristics, meet process requirements, and achieve cost-effective production. Ramana [20] addressed this challenge by utilizing data mining models. Techniques, such as naive bayes, DT, ANN, and polynomial by binomial classification, were employed to predict product quality and identify the root causes of such defects as short shots and flashes in injection molding.

Next, the study using the ensemble model along with the single model is summarized here. Ensemble techniques have garnered significant attention in the recent research to try and enhance prediction accuracy. Polenta et al. [21] aimed to monitor the quality of injection molded products by tracking the process parameters in plastic injection molding. They collected data from the production of plastic road lenses and used them as process parameters to compare the predictive performance of six classifiers, namely, KNN, DT, RF, gradient boosting tree (GBT), SVM, and MLP. The research results showed that RF and GBT among the tested classifiers exhibited higher accuracy in predicting the quality of molded products, thereby confirming the potential of ML-based techniques. They also suggested that the utility of ensemble techniques based on bagging and boosting would achieve better results than single prediction models.

Jung et al. [22] tested and compared the performance of various ML algorithms for their quality prediction. They applied eight techniques: LR, SVM, RF, Gradient Boosting, XGBoost, CatBoost, LightGBM, and autoencoder. Of these,

TABLE 1. The different ML prediction studies on injection molding.

Subject	Model	Technique	References
Process optimization	Single	ANN	Changyu et al. (2006)
Quality prediction	Single	ANN (AE, MLP)	Huang et al. (2022)
	Single	ANN, SVM	Silva et al. (2021)
	Single	ANN (MLP)	Ke and Huang (2020)
	Single	KNN, Naive Bayes, Linear Discriminant Analysis, DT	Parizs et al. (2022)
	Single	ANN, LR, Polynomial Regression	Lee et al. (2022)
	Single	ANN, ANOVA	Ozcelik and Erzurumlu (2005)
	Single	FEA, ANOVA, ANN, ANN+PSO, ANN+GA	Ardestani et al. (2023)
	Single	Naive Bayes, DT, ANN, Polynomial by Binomial Classification	Ramana (2017)
	Single and Ensemble	KNN, DT, RF, GBT, SVM, MLP, Bagging, Boosting	Polenta et al. (2022)
	Single and Ensemble	LR, SVM, RF, Gradient Boosting, XGBoost, CatBoost, LightGBM, Autoencoder	Jung et al. (2021)
	Single and Ensemble	ANN, SVM, DT, KNN, Ensemble models based on DT and GPR	Struchtrupa et al. (2021)
	Ensemble	RF, Gradient Boosted Regression Tree	Ahmed et al. (2020)

the autoencoder yielded superior performance by comparing metrics, such as accuracy, precision, recall, and F1-Score. Further still, feature importance analysis identified temperature and time as significant factors that were influencing the quality.

Struchtrupa et al. [23] demonstrated that ensemble models outperform single models in terms of quality prediction for injection molding. They evaluated various single models, such as ANN, SVM, DT, and KNN, and compared them to ensemble models based on DT and gaussian process regression (GPR). The research findings highlighted the superior performance of the ensemble models in predicting quality compared to the single models. This finding underscored the effectiveness of ensemble techniques in the context of predicting of the quality of injection molding.

Ahmed et al. [24] developed prediction models for deformation in PVC parts by using ensemble algorithms, including RF and a gradient-boosted regression tree. This study found that of these two ensemble ML algorithms, the RF model exhibited more accurate deformation predictions for injection-molded parts. The study suggested that this model could be valuable for manufacturing engineers and production managers when controlling injection molding process parameters and minimizing deformation before actual production.

Table 1 summarizes the studies in the field of injection molding that are discussed here.

B. THE INJECTION MOLDING PRODUCT USED IN THIS STUDY

This study utilized a dataset obtained from Shinsung Delta Tech (www.gshinsung.com), a mid-sized manufacturing company located in South Korea. Shinsung Delta Tech specializes in the production of electronic products, including wireless vacuum cleaners, robot vacuum cleaners, skincare devices, and body dryers. They also manufacture key components and modules for household electronic products like washing machines, refrigerators, and air conditioners. Given their expertise in large-scale injection molding techniques, they have developed advanced capabilities for producing plastic parts. The specific focus of this study was on the production of fans used in the outdoor unit air conditioners, where weight balance plays a crucial role because fans are spinning. The researchers collected process data related to the injection molding of these fans and applied ML techniques for their analysis. As shown here in Fig. 1, the mold used for the outdoor air conditioner fan consisted of two cavities, which enabled the simultaneous production of two products during the injection molding process. Consequently, separate datasets were created for each cavity.

Temperature sensors and pressure sensors were installed in each cavity of the mold to collect data on the respective influences of each cavity. In addition, the barrel of the injection molding machine was divided into sections to collect temperature data as shown in Fig. 2. The barrel is typically a long cylindrical shape with heating elements and a screw inside, responsible for heating and melting the plastic raw material to inject it into the mold. The data collected from the barrel affects the entire mold, and this was utilized as common data. The temperature sensors T1, T2, T3, and T7, and the pressure sensors P1, P2, P3, and P7 exclusively influenced Cavity1, while the temperature sensors T4, T5, T6, and T8, and pressure sensors P4, P5, P6, and P8 exclusively influenced Cavity2 as shown in Fig. 1. Temperature data denoted

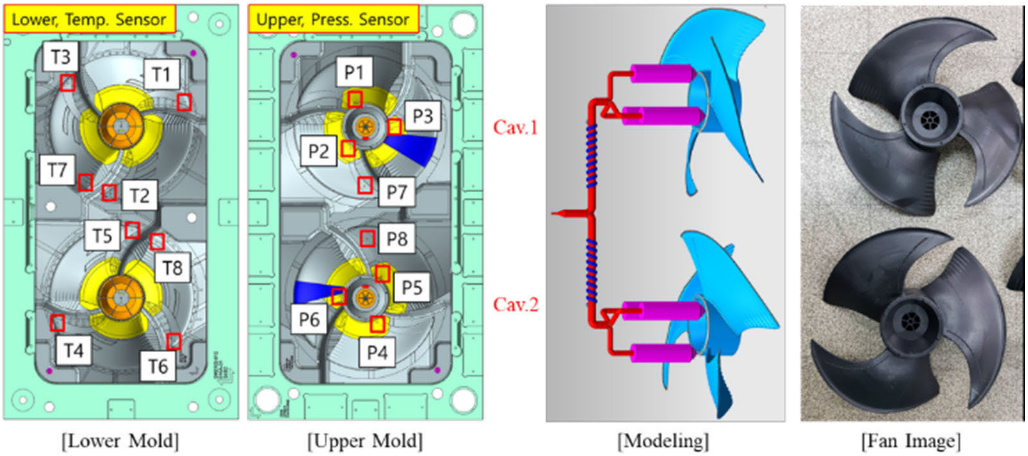

FIGURE 1. Locations of sensor in the mold and the fan image.

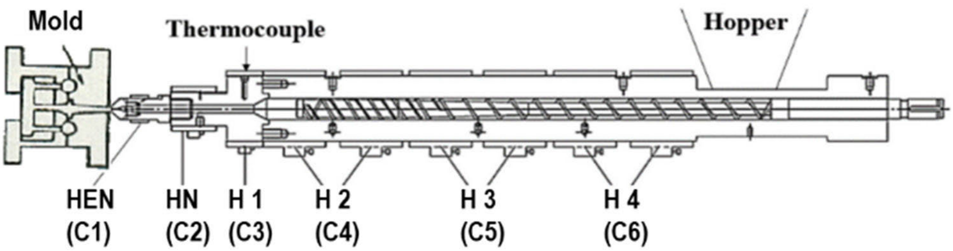

FIGURE 2. Locations of the temperature measurement for each section of the barrel in the injection molding machine.

as HEN, HN, H1, H2, H3, and H4 in the barrel, affected both Cavity1 and Cavity2 in the mold as shown in Fig. 2.

III. RESEARCH METHODOLOGY

A. RESEARCH DIRECTION

This study aims to develop predictive models for weight defects in injection-molded products and then assess their performance. The research process follows a systematic flow, including data collection, preprocessing, selection of the influential variables, constructing and evaluating the predictive model, and evaluating the ensemble and double ensemble algorithms as shown in Fig. 3.

First, we collected data specifically related to the injection molding conditions used in producing fans for an outdoor air conditioner. That dataset is comprised of information on the injection molding conditions employed for these fans. While the target weight range for the product is 1,260g±25, we narrowed the weight prediction management range for any defect prediction to 1,260g±20.

Second, the mold for the product consisted of two cavity structures, allowing for the simultaneous production of two items. The cavities may have subtle differences in structure and shape and different influential variables. Therefore, we differentiated the mold into Cavity1 and Cavity2. To address the issue of data imbalance for each cavity, we utilized the synthetic minority oversampling technique (SMOTE) as an oversampling technique to balance the data.

Third, through feature selection, we identified the important variables that affect each cavity into two types: common variables and individual variables. We analyzed the gain ratio for each cavity and listed the variables in their order of importance. For the data analysis, we used the data mining tool Weka version 3.8.6.

Fourth, to derive the optimal variable conditions for each model, we utilized the backward elimination method in conjunction with five different single prediction models. Starting with the least influential variables, we compared the accuracy of our performance prediction using a 10-fold cross validation mode and a mode where 66% of the data were used as training data; the remaining 34% was randomly allocated as test data (split 66%).

Fifth, the characteristics of a classification model are determined by the nature of the data. The methods for building classification models are categorized as parametric or non-parametric methods. Since it is not known which approach performs best on the data in this study, we selected the most commonly used algorithms from each approach. As a result, this study built five single models: DT, RF, LR, BN, and ANN (MLP).

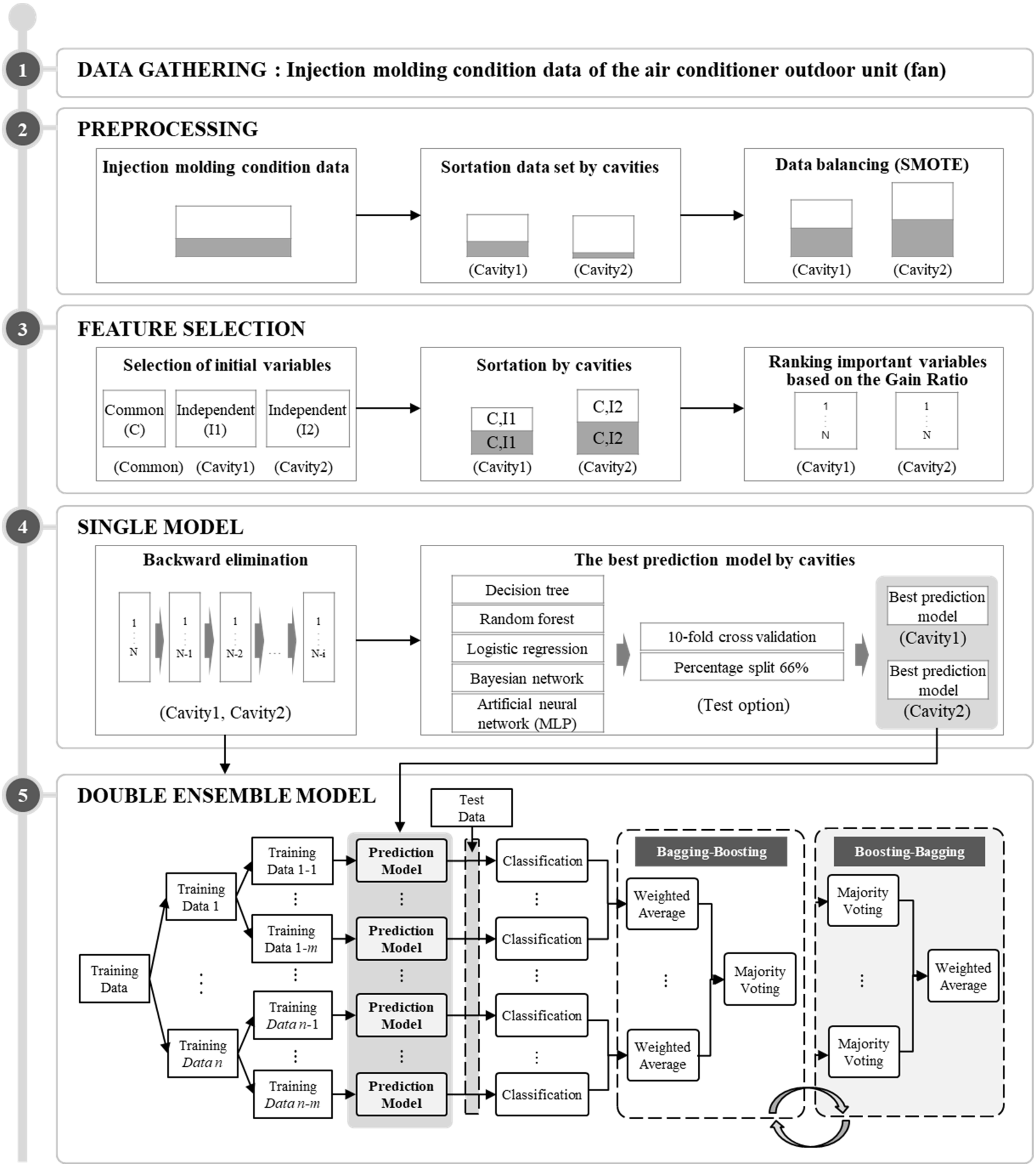

FIGURE 3. The research framework for single and double ensemble model.

Finally, we applied ensemble techniques, such as bagging, boosting, and double ensemble, including bagging-boosting and boosting-bagging, to the single optimal model that was derived earlier. The purpose and goal together were to evaluate whether these ensemble and double ensemble techniques improved the performance of the predictive model when compared to the single model.

B. DATA GATHERING & THE PREPROCESSING STEPS

The dataset for this study was collected during the mass production verification of a newly manufactured fan mold for the outdoor unit of an air conditioner. The data collection period ran from January 11, 2022 to April 19, 2022. The data consisted of two sets: Cavity1 and Cavity2. Fig. 4 shows the distribution of the weight data collected by each cavity.

TABLE 2. Data balancing results for equalization using SMOTE.

	Before SMOTE						After SMOTE					
	Good		Defect		Sum		Good		Defect		Sum	
Cavity1	1,562	56%	1,222	44%	2,784	100%	1,562	50%	1,562	50%	3,124	100%
Cavity2	2,426	96%	94	4%	2,520	100%	2,426	50%	2,426	50%	4,852	100%

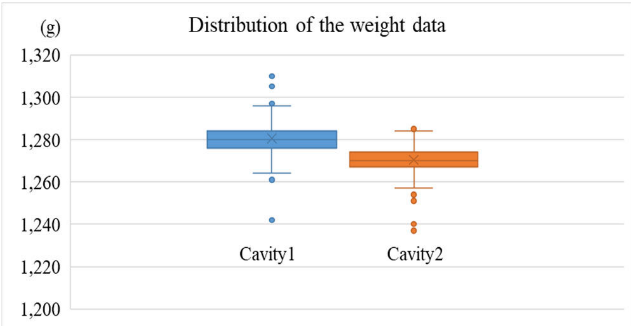

FIGURE 4. The distribution of the weight data for each cavity.

Cavity1 data were comprised 1,562 instances of good products and 1,222 instances of defective products, thus a total of 2,784 instances. Cavity2 data included 2,426 instances of good products and 94 instances of defective products, thus totaling 2,520 instances. Notably, Cavity2 exhibited a significantly lower defect rate, likely due to the higher precision of the mold and resulting in weights closer to the target value. Data imbalance occurs when there is a substantial difference in the number of observations between different categories, as observed in Cavity2. This circumstance can lead to biased prediction models that perform well on one side, but struggle on the other side.

To address this issue, sampling techniques can be employed to adjust the data and mitigate data imbalance. One such technique is under-sampling, which involves reducing the number of data points in the majority class to match the number of data points in the minority class. Under-sampling offers the advantage of faster computation time by retaining only meaningful data for training and analysis. However, it can also result in information loss and decreased classification accuracy due to data removal.

In contrast, oversampling is a method that increases the number of samples in the minority class to balance with the majority class and thus address the data imbalance by augmenting the representation of the minority class. Among the different oversampling techniques, SMOTE is commonly used. SMOTE selects random data points from the minority class and generates synthetic data by randomly selecting one of its k-nearest neighbors based on a defined k-value, using a synthetic formula. Oversampling techniques like SMOTE will help prevent data loss and often result in higher classification accuracy than under-sampling. However, these techniques may increase the computation time due to the larger volume of total data. In this study, SMOTE was specifically applied as an oversampling technique to address the data imbalance observed in Cavity2, thereby ensuring a balanced representation of the different categories in this dataset.

After applying the SMOTE oversampling technique, the dataset was rebalanced for Cavity1 and Cavity2. Cavity1, which initially had 1,562 instances of good products and 1,222 instances of defective products, now consisted of 3,124 instances. This rebalanced dataset included an equal number of 1,562 instances for both good and defective products. Similarly, Cavity2, which initially had 2,426 instances of good products and 94 instances of defective products, now comprised 4,852 instances. The rebalanced dataset for Cavity2 included an equal number of 2,426 instances for both good products and defective products. The impact of using SMOTE oversampling to balance the data can be observed in Table 2 here, which presents the results of this equalization process.

A detailed analysis of the mold structures was also conducted to account for the variability in the independent variables influencing the target variable (weight) between Cavity1 and Cavity2. This analysis identified a total of 51 independent variables, comprised of 26 common variables and 25 individual variables. These variables are listed in Table 3 below and provide a comprehensive overview of the independent variables for both Cavity1 and Cavity2. In this study, all independent variables are numerical and the dependent variable is nominal. The dependent variable is distinguished well from the defective by whether or not it meets the weight management range 1,260g±20. There are two groups of variables: those generated based on a time-series, and those generated simultaneously. The temperature values (T1 ∼ T8) and the pressure values (P1 ∼ P8) in the mold shown in Fig. 1 are collected simultaneously. The temperature values of the mold hot runner system (H/R1 ∼ H/R21) are also collected simultaneously. Fig. 2 shows that temperature data such as barrel H4, H3, H2, H1, HN, and HEN are collected sequentially. Therefore, variables are categorized based on the timing of their generation, including both time-series data and data generated simultaneously.

C. FEATURE SELECTION

We evaluated the importance of variables using two common methods gain ratio and correlation as shown in Table 4 and Table 5. The evaluation results, when normalized using min-max normalization and visualized in Fig. 5 and Fig. 6, show some differences in importance rankings depending on

TABLE 3. The variables for Cavity1 and Cavity2.

Common variables			Independent variables (Cavity1)			Independent variables (Cavity2)		
C1	Barrel HEN Temp Real	℃	I1	H/R 1 Real	℃	I1	H/R 7 Real	℃
C2	Barrel HN Temp Real	℃	I2	H/R 2 Real	℃	I2	H/R 8 Real	℃
C3	Barrel H1 Temp Real	℃	I3	H/R 3 Real	℃	I3	H/R 9 Real	℃
C4	Barrel H2 Temp Real	℃	I4	H/R 4 Real	℃	I4	H/R 10 Real	℃
C5	Barrel H3 Temp Real	℃	I5	H/R 5 Real	℃	I5	H/R 11 Real	℃
C6	Barrel H4 Temp Real	℃	I6	H/R 6 Real	℃	I6	H/R 12 Real	℃
C7	Cushion	mm	I7	H/R 13 Real	℃	I7	H/R 16 Real	℃
C8	Fill Time	sec	I8	H/R 14 Real	℃	I8	H/R 17 Real	℃
C9	Fill+Pack Time	sec	I9	H/R 15 Real	℃	I9	H/R 18 Real	℃
C10	Metering Time	sec	I10	iMOLD TempStart1	℃	I10	iMOLD TempStart4	℃
C11	Open Time	sec	I11	iMOLD TempStart2	℃	I11	iMOLD TempStart5	℃
C12	Close Time	sec	I12	iMOLD TempStart3	℃	I12	iMOLD TempStart6	℃
C13	H/R 19 Real	℃	I13	iMOLD TempStart7	℃	I13	iMOLD TempStart8	℃
C14	H/R 20 Real	℃	I14	iMOLD TempEnd1	℃	I14	iMOLD TempEnd4	℃
C15	H/R 21 Real	℃	I15	iMOLD TempEnd2	℃	I15	iMOLD TempEnd5	℃
C16	Cooling Temp 1 Real	℃	I16	iMOLD TempEnd3	℃	I16	iMOLD TempEnd6	℃
C17	Cooling Temp 2 Real	℃	I17	iMOLD TempEnd7	℃	I17	iMOLD TempEnd8	℃
C18	Cooling Temp 3 Real	℃	I18	iMOLD TempMax1	℃	I18	iMOLD TempMax4	℃
C19	Cooling Temp 4 Real	℃	I19	iMOLD TempMax2	℃	I19	iMOLD TempMax5	℃
C20	Cooling Temp 5 Real	℃	I20	iMOLD TempMax3	℃	I20	iMOLD TempMax6	℃
C21	Cooling Temp 6 Real	℃	I21	iMOLD TempMax7	℃	I21	iMOLD TempMax8	℃
C22	Oil Real	℃	I22	iMOLD PressMax1	bar	I22	iMOLD PressMax4	bar
C23	Cooling Water Low IN Real	℃	I23	iMOLD PressMax2	bar	I23	iMOLD PressMax5	bar
C24	Cooling Water High IN Real	℃	I24	iMOLD PressMax3	bar	I24	iMOLD PressMax6	bar
C25	Cooling Water Low OUT Real	℃	I25	iMOLD PressMax7	bar	I25	iMOLD PressMax8	bar
C26	Cooling Water High OUT Real	℃						

the evaluation method. We conducted preliminary tests by applying the priority rankings obtained from each method to evaluate the model's performance. The results shown in Table 6 indicated that gain ratio somewhat outperformed the other. Therefore, in this study, we intend to use gain ratio to evaluate variable importance.

A backward elimination process was employed to identify the combination of variables that produced the highest accuracy for the prediction models. In order to find the optimal combination of variables, this process involved sequentially removing variables with the least influence on the prediction models based on the low gain ratio values for each variable in each cavity. The GainRatioAttributeEval provided by Weka was used to evaluate the accuracy of the prediction models, as variables were eliminated. Thus, 51 variables were evaluated for each cavity, and their removal was carried out in descending order of gain ratio values. This process, which is known as variable selection using backward elimination, aimed to exclude the variables with minimal impact on the prediction models while still searching for the optimal combination. The performance evaluation results for each prediction model, as variables were eliminated, can be found in Table 4. By analyzing the independent variables influencing the target variable (weight) based on their gain ratio values, we could address RQ1. Variables with relatively larger gain ratio values had a greater impact on the target variable. From Table 4, it is evident that among the top 10 important variables for each cavity were 8 common variables for Cavity1 and 7 common variables for Cavity2. This notation indicates that common variables have a larger influence than individual variables do.

Further, among the common variables for Cavity1 and Cavity2, 3 variables stood out as having a high impact on weight: barrel H2 temp real (C4), metering time (C10), and fill time (C8). However, the individual variables were excluded from the analysis, as they only affected their respective cavity. It is important to note, and yet, that there may be other variables with significant impacts on weight during the injection molding process, and further, that the importance of variables can vary depending on the mold or the product. However, based on the dataset used in this study, the above three common variables were identified as the key factors that significantly influence the weight of the fan for the outdoor air conditioner unit.

D. A PREDICTION MODEL & EVALUATION

Using the preprocessed dataset, we conducted a weight defect prediction for injection molded products using five

TABLE 4. The ranking of variables importance with gain ratio.

Cavity1				Cavity2			
No.	**Type**	**Variables**	**Gain ratio**	**No.**	**Type**	**Variables**	**Gain ratio**
1	C25	Cooling Water Low OUT Real	0.2077	1	C4	Barrel H2 Temp Real	0.2148
2	C16	Cooling Temp 1 Real	0.1879	2	C9	Fill+Pack Time	0.2026
3	C14	H/R 20 Real	0.1761	3	C10	Metering Time	0.2003
4	C4	Barrel H2 Temp Real	0.1760	4	C5	Barrel H3 Temp Real	0.1990
5	C11	Open Time	0.1420	5	I10	iMOLD TempStart4	0.1875
6	C10	Metering Time	0.1379	6	C8	Fill Time	0.1802
7	C8	Fill Time	0.1198	7	C6	Barrel H4 Temp Real	0.1790
8	C13	H/R 19 Real	0.1109	8	I12	iMOLD TempStart6	0.1745
9	I8	H/R 14 Real	0.1098	9	I7	H/R 16 Real	0.1744
10	I3	H/R 3 Real	0.1086	10	C16	Cooling Temp 1 Real	0.1714
11	I5	H/R 5 Real	0.1086	11	C20	Cooling Temp 5 Real	0.1690
12	I7	H/R 13 Real	0.1086	12	C19	Cooling Temp 4 Real	0.1684
13	I1	H/R 1 Real	0.1073	13	C17	Cooling Temp 2 Real	0.1674
14	C23	Cooling Water Low IN Real	0.1072	14	C13	H/R 19 Real	0.1663
15	C9	Fill+Pack Time	0.1055	15	C25	Cooling Water Low OUT Real	0.1662
16	I15	iMOLD TempEnd2	0.0970	16	I8	H/R 17 Real	0.1650
17	I16	iMOLD TempEnd3	0.0948	17	C26	Cooling Water High OUT Real	0.1626
18	C12	Close Time	0.0902	18	C18	Cooling Temp 3 Real	0.1619
19	C7	Cushion	0.0880	19	C24	Cooling Water High IN Real	0.1614
20	C24	Cooling Water High IN Real	0.0826	20	C14	H/R 20 Real	0.1613
21	I19	iMOLD TempMax2	0.0792	21	C7	Cushion	0.1605
22	I17	iMOLD TempEnd7	0.0777	22	C21	Cooling Temp 6 Real	0.1592
23	I20	iMOLD TempMax3	0.0751	23	I15	iMOLD TempEnd5	0.1565
24	C5	Barrel H3 Temp Real	0.0718	24	I16	iMOLD TempEnd6	0.1564
25	I21	iMOLD TempMax7	0.0705	25	I14	iMOLD TempEnd4	0.1563
26	C26	Cooling Water High OUT Real	0.0697	26	I17	iMOLD TempEnd8	0.1549
27	C21	Cooling Temp 6 Real	0.0651	27	I21	iMOLD TempMax8	0.1532
28	I6	H/R 6 Real	0.0637	28	C11	Open Time	0.1530
29	I24	iMOLD PressMax3	0.0604	29	I9	H/R 18 Real	0.1525
30	I12	iMOLD TempStart3	0.0603	30	I18	iMOLD TempMax4	0.1524
31	I14	iMOLD TempEnd1	0.0590	31	I11	iMOLD TempStart5	0.1505
32	I25	iMOLD PressMax7	0.0586	32	I13	iMOLD TempStart8	0.1488
33	I11	iMOLD TempStart2	0.0582	33	C23	Cooling Water Low IN Real	0.1477
34	I13	iMOLD TempStart7	0.0568	34	I20	iMOLD TempMax6	0.1465
35	I22	iMOLD PressMax1	0.0557	35	I1	H/R 7 Real	0.1372
36	C22	Oil Real	0.0551	36	I3	H/R 9 Real	0.1340
37	I10	iMOLD TempStart1	0.0486	37	C12	Close Time	0.1320
38	C6	Barrel H4 Temp Real	0.0463	38	I5	H/R 11 Real	0.1251
39	I18	iMOLD TempMax1	0.0452	39	I19	iMOLD TempMax5	0.1248
40	C2	Barrel HN Temp Real	0.0424	40	I22	iMOLD PressMax4	0.1245
41	C1	Barrel HEN Temp Real	0.0303	41	C22	Oil Real	0.1244
42	I23	iMOLD PressMax2	0.0287	42	I6	H/R 12 Real	0.1219
43	C19	Cooling Temp 4 Real	0.0272	43	I24	iMOLD PressMax6	0.1197
44	C20	Cooling Temp 5 Real	0.0270	44	I2	H/R 8 Real	0.1190
45	C18	Cooling Temp 3 Real	0.0263	45	C1	Barrel HEN Temp Real	0.1143
46	C17	Cooling Temp 2 Real	0.0198	46	I4	H/R 10 Real	0.1133
47	C3	H/R 4 Real	0.0000	47	C3	Barrel H1 Temp Real	0.1122
48	C15	H/R 15 Real	0.0000	48	I23	iMOLD PressMax5	0.1092
49	I2	Barrel H1 Temp Real	0.0000	49	C2	Barrel HN Temp Real	0.1035
50	I4	H/R 21 Real	0.0000	50	I25	iMOLD PressMax8	0.1027
51	I9	H/R 2 Real	0.0000	51	C15	H/R 21 Real	0.0857

TABLE 5. The ranking of variables importance with correlation.

Cavity1				Cavity2			
No.	Type	Variables	Correlation	No.	Type	Variables	Correlation
1	C4	Barrel H2 Temp Real	0.4352	1	C10	Metering Time	0.5613
2	C7	Cushion	0.4101	2	C4	Barrel H2 Temp Real	0.4870
3	C14	H/R 20 Real	0.3910	3	C7	Cushion	0.4705
4	C12	Close Time	0.3753	4	I22	iMOLD PressMax4	0.4034
5	I21	iMOLD TempMax7	0.3648	5	I23	iMOLD PressMax5	0.3733
6	I20	iMOLD TempMax3	0.3636	6	I24	iMOLD PressMax6	0.3578
7	I17	iMOLD TempEnd7	0.3573	7	C9	Fill+Pack Time	0.3267
8	C24	Cooling Water High IN Real	0.3522	8	C8	Fill Time	0.3263
9	I16	iMOLD TempEnd3	0.3476	9	C16	Cooling Temp 1 Real	0.2062
10	I14	iMOLD TempEnd1	0.3341	10	C26	Cooling Water High OUT Real	0.1415
11	C26	Cooling Water High OUT Real	0.3290	11	C24	Cooling Water High IN Real	0.1384
12	I15	iMOLD TempEnd2	0.3267	12	C14	H/R 20 Real	0.1351
13	I13	iMOLD TempStart7	0.3261	13	C20	Cooling Temp 5 Real	0.1314
14	I12	iMOLD TempStart3	0.3232	14	C22	Oil Real	0.1157
15	I19	iMOLD TempMax2	0.3171	15	C25	Cooling Water Low OUT Real	0.1083
16	I10	iMOLD TempStart1	0.3091	16	I3	H/R 9 Real	0.1072
17	C10	Metering Time	0.3070	17	C5	Barrel H3 Temp Real	0.0992
18	C9	Fill+Pack Time	0.3016	18	I10	iMOLD TempStart4	0.0983
19	C8	Fill Time	0.3013	19	I19	iMOLD TempMax5	0.0956
20	I11	iMOLD TempStart2	0.2945	20	C6	Barrel H4 Temp Real	0.0945
21	I18	iMOLD TempMax1	0.2894	21	I14	iMOLD TempEnd4	0.0944
22	I25	iMOLD PressMax7	0.2578	22	I1	H/R 7 Real	0.0943
23	C22	Oil Real	0.2430	23	I6	H/R 12 Real	0.0915
24	C16	Cooling Temp 1 Real	0.2237	24	I18	iMOLD TempMax4	0.0912
25	C21	Cooling Temp 6 Real	0.1326	25	C13	H/R 19 Real	0.0792
26	C11	Open Time	0.1215	26	I2	H/R 8 Real	0.0770
27	I22	iMOLD PressMax1	0.0943	27	C15	H/R 21 Real	0.0706
28	C20	Cooling Temp 5 Real	0.0836	28	I13	iMOLD TempStart8	0.0674
29	C6	Barrel H4 Temp Real	0.0790	29	I8	H/R 17 Real	0.0673
30	C1	Barrel HEN Temp Real	0.0734	30	C19	Cooling Temp 4 Real	0.0618
31	C13	H/R 19 Real	0.0613	31	I9	H/R 18 Real	0.0592
32	C23	Cooling Water Low IN Real	0.0600	32	C2	Barrel HN Temp Real	0.0569
33	C5	Barrel H3 Temp Real	0.0534	33	C3	Barrel H1 Temp Real	0.0565
34	C2	Barrel HN Temp Real	0.0512	34	C12	Close Time	0.0558
35	I7	H/R 13 Real	0.0451	35	C17	Cooling Temp 2 Real	0.0539
36	C15	H/R 15 Real	0.0444	36	C23	Cooling Water Low IN Real	0.0520
37	I3	H/R 3 Real	0.0388	37	I17	iMOLD TempEnd8	0.0418
38	C18	Cooling Temp 3 Real	0.0371	38	C11	Open Time	0.0414
39	I9	H/R 2 Real	0.0365	39	C21	Cooling Temp 6 Real	0.0408
40	I1	H/R 1 Real	0.0362	40	I4	H/R 10 Real	0.0327
41	I8	H/R 14 Real	0.0330	41	I11	iMOLD TempStart5	0.0320
42	I5	H/R 5 Real	0.0316	42	I7	H/R 16 Real	0.0291
43	I23	iMOLD PressMax2	0.0255	43	I5	H/R 11 Real	0.0276
44	C3	H/R 4 Real	0.0237	44	I25	iMOLD PressMax8	0.0230
45	I6	H/R 6 Real	0.0231	45	I15	iMOLD TempEnd5	0.0221
46	I4	H/R 21 Real	0.0200	46	I21	iMOLD TempMax8	0.0183
47	C25	Cooling Water Low OUT Real	0.0169	47	I16	iMOLD TempEnd6	0.0104
48	C17	Cooling Temp 2 Real	0.0163	48	C18	Cooling Temp 3 Real	0.0098
49	I24	iMOLD PressMax3	0.0115	49	I20	iMOLD TempMax6	0.0065
50	C19	Cooling Temp 4 Real	0.0077	50	I12	iMOLD TempStart6	0.0058
51	I2	Barrel H1 Temp Real	0.0031	51	C1	Barrel HEN Temp Real	0.0051

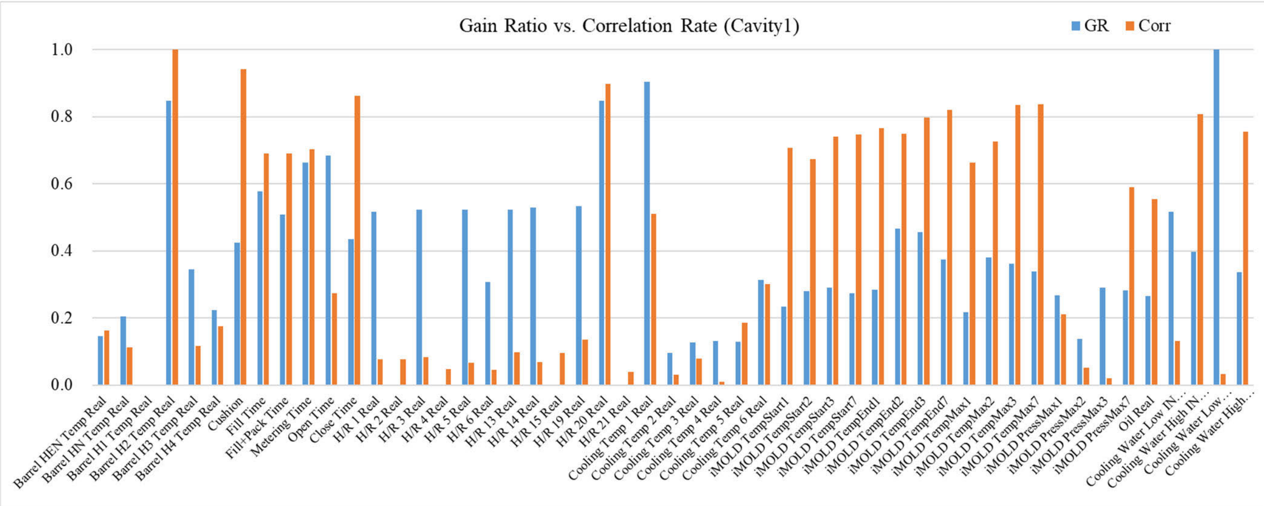

FIGURE 5. Min-max normalization of gain ratio vs. correlation (Cavity1).

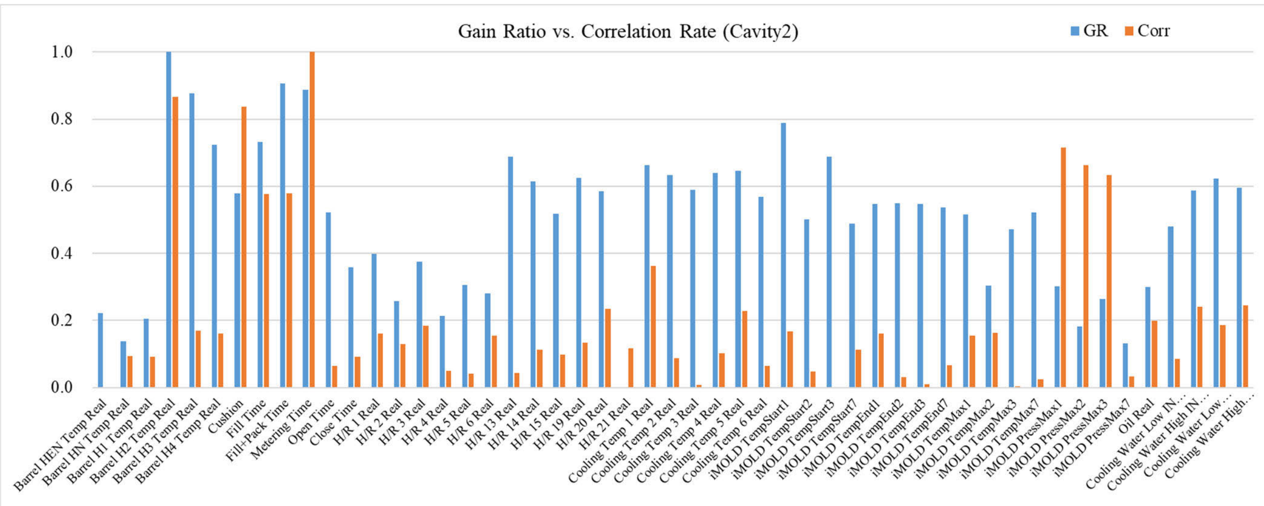

FIGURE 6. Min-max normalization of gain ratio vs. correlation (Cavity2).

TABLE 6. The prediction results for the single model with gain ratio vs. correlation (10-fold cross validation).

	Cavity1		Cavity2	
	Gain ratio	Correlation	Gain ratio	Correlation
DT	91.58%	91.26%	96.21%	96.13%
RF	93.82%	93.92%	97.98%	98.19%
LR	88.12%	88.19%	86.95%	87.02%
BN	87.16%	85.44%	97.71%	97.71%

single models: DT, RF, LR, BN, and ANN (MLP) algorithms for each cavity. We applied the variable combination for every single model that yielded the highest prediction accuracy. We then compared the performance of ensemble techniques (bagging, boosting) and double ensemble techniques (bagging-boosting, boosting-bagging) to assess the improvement in prediction accuracy. Ensemble techniques involve combining multiple single models to aggregate their results and make predictions and achieve a model that surpasses the performance of a single model in terms of accuracy and generalization. Ensemble techniques have shown a consistent tendency to improve prediction accuracy. Bagging and boosting are thus widely used as ensemble techniques.

Bagging, short for bootstrap aggregating, is a technique that enhances the predictive performance of a model by creating multiple single learning models via repeated random sampling using replacement, known as bootstrap sampling, from the training data. The results of these single models are then combined to generate the final prediction. As the sampling is done using replacement, some data points may be selected multiple times, while others may not be selected at all. In bagging, the final prediction is typically determined using majority voting based on the results of the single learning models. Boosting is a technique similar to bagging, but with its own distinct characteristics. It assigns higher weights to misclassified single learning models and combines them using weighted averaging to produce the final result. This process differs from bagging, where equal probabilities are assigned to the predictions of single learning models. Boosting thus sequentially increases the probability of selecting misclassified samples in subsequent sampling processes by assigning them higher weights. Therefore, in this study, we employed the Ada boosting technique.

Finally, we applied the selected variable combinations from every single model to implement the double ensemble technique. We compared two approaches for predictive accuracy, the sequentially using of bagging-boosting and sequentially using boosting-bagging. These double ensemble techniques were employed to enhance the accuracy of weight defect predictions in injection molded products based on the selected variables, ultimately improving the overall performance of the prediction models.

IV. EXPERIMENTAL RESULTS OF PREDICTION MODEL

In this study, we focused on the dataset that represented the injection molding conditions for the outdoor unit fans of an air conditioner. The data were divided into Cavity1 and Cavity2. To predict weight defects, we utilized the widely used data mining tool, Weka version 3.8.6, which offers a range of algorithms for analysis and modeling. Specifically, we applied five single models: the DT, RF, LR, BN, and ANN (MLP) algorithms. Further, we sought to enhance the prediction performance by employing ensemble models. Ensemble models combine the predictions of multiple single models to produce a final prediction that is often more accurate and more robust. In this study, we thus evaluated the performance of ensemble models using bagging and boosting techniques.

Additionally, we explored the potential of double ensemble modeling, which combines bagging and boosting sequentially. By leveraging the strengths of both techniques, the double ensemble model aimed to improve the accuracy of weight defect prediction still further. Overall, the study employed various models and techniques, including the single models: DT, RF, LR, BN, and ANN (MLP), ensemble models (bagging and boosting), and the double ensemble models (which combined bagging and boosting). The objective was to evaluate and compare all their effectiveness in predicting weight defects in injection molded fans for outdoor units of an air conditioner. In this study, parameter settings for each model were configured using the features provided by Weka, and experiments were conducted as shown in Table 7. Specifically, ANN employed the most commonly used MLP. It consisted of one hidden layer with 27 neurons, configured using the predefined variable ‘a’, and the sigmoid function was used as the activation function.

A. THE SINGLE MODEL

1) CAVITY1 DATA ANALYSIS

For Cavity1 dataset, we performed a classification prediction evaluation using 51 variables, with weight defect as the target variable. This evaluation process involved training the models using the training data and conducting iterative training using a 10-fold cross validation. Each of the five single models: DT, RF, LR, BN, and ANN (MLP) was evaluated based on its accuracy using the confusion matrix. To improve the accuracy of the single models, we sequentially eliminated any variables with low gain ratios from the 51 variables. This process aimed to find the combination of variables that yielded the highest prediction accuracy for each model. The accuracy results obtained for each single model were as follows: DT 91.58%; RF 93.82%; LR 88.12%; BN 87.16%; and ANN (MLP) 91.54%. Among these models, RF achieved the highest prediction accuracy at 93.82%. During the variable elimination process, we identified the important variables that significantly impact the prediction accuracy of each model. The selected important variables for these models were 7, 36, 45, 7, and 37, respectively. The specific prediction results for each model are noted in Table 8. The comparison results of confusion matrix and correct classification rate (CCR) are shown in Table 9. Also, CCR means accuracy here. The optimal combination of variables obtained through a backward elimination process using the gain ratio is depicted in Fig. 7. These results clearly demonstrate the effectiveness of each model in accurately predicting weight defects in Cavity1 dataset. The selected variables were also crucial in improving the model's performance by providing valuable insights into the factors that were influencing weight defects in injection molded products.

2) CAVITY2 DATA ANALYSIS

For Cavity2 dataset, we performed a classification prediction evaluation using the same 51 variables with weight defect as the target variable and following the same procedure as for Cavity1. The models were trained using the training data and evaluated via 10-fold cross validation. The accuracy results obtained for each single model were as follows: DT 96.21%; RF 97.98%; LR 86.95%; BN 97.71%; and ANN (MLP) 94.91%. For these models, RF achieved the highest prediction accuracy at 97.98%. Similar to Cavity1, we conducted the variable elimination process to identify the important

TABLE 7. The parameter settings in Weka.

Algorithm	Classifier in Weka	Parameter	Setting
DT	J48	batchSize	100
		confidenceFactor	0.25
		minNumObj	2
		seed	1
RF	RandomForest	bagSizePercent	100
		batchSize	100
		numIterations	100
		seed	1
LR	Logistic	batchSize	100
		maxIts	-1
		numDecimalPlaces	4
		ridge	1.00E-08
BN	Bayesnet	batchSize	100
		SimpleEstimator	alpha 0.5
		numDecimalPlaces	2
		searchAlgorithm	scoretype BAYES
ANN (MLP)	MultilayerPerceptron	batchSize	100
		learningRate	0.3
		hiddenLayers	a
		trainingTime	500
		activationFunction	sigmoid
Bagging	Bagging	bagSizePercent	90
		batchSize	100
		numDecimalPlaces	2
		numIterations	11
		seed	1
Boosting	AdaBoostM1	batchSize	100
		numDecimalPlaces	2
		numIterations	11
		seed	1
		weightThreshold	100

variables that significantly impact the prediction accuracy of each model. The selected important variables for these models were 7, 27, 41, 13, and 50, respectively. These variables were crucial in improving the model's performance when predicting weight defects in Cavity2 dataset. The prediction results for each model can be seen in Table 8, which showcases the performance of the single models. Additionally, the optimal combination of the variables obtained via the backward elimination process using the gain ratio are depicted in Fig. 8.

These results highlight the effectiveness of the RF model in accurately predicting weight defects in Cavity2 dataset. The selected variables also provide valuable insights into the factors that influence weight defects in injection-molded products, thereby contributing to the overall understanding of that manufacturing process.

B. THE DOUBLE ENSEMBLE MODEL

1) CAVITY1 DATA ANALYSIS

In the case of Cavity1, we applied the important variable combinations derived from every single model to the ensemble models using the bagging and boosting technique. The predictive results obtained through a 10-fold cross validation for the bagging ensemble were as follows: DT 91.84%; RF 93.50%; LR 88.19%; BN 87.64%; and ANN (MLP) 92.67%. Similarly, the predictive results for the boosting ensemble were as follows: DT 91.17%; RF 93.60%; LR 88.12%; BN 89.69%; and ANN (MLP) 91.58%. Although there were slight variations in their predictive accuracy compared to the single models, overall, there were no significant changes in performance when using the ensemble techniques.

However, when applying the double ensemble technique to the order of bagging-boosting, the predictive results for

TABLE 8. The prediction results for the single model (10-fold cross validation).

Type	Result	DT	RF	LR	BN	ANN (MLP)	Average
Cavity1	Accuracy	91.58%	93.82%	88.12%	87.16%	91.54%	90.44%
	Important variables	7	36	45	7	37	26.4
Cavity2	Accuracy	96.21%	97.98%	86.95%	97.71%	94.91%	94.75%
	Important variables	7	27	41	13	50	27.6

TABLE 9. The confusion matrix and CCR (10-fold cross validation).

Type	Confusion matrix	DT	RF	LR	BN	ANN (MLP)
Cavity1	TP	1,393	1,438	1,369	1,385	1,429
	FP	169	124	193	177	133
	FN	94	69	178	224	131
	TN	1,468	1,493	1,384	1,338	1,431
	CCR	91.58%	93.82%	88.12%	87.16%	91.54%
Cavity2	TP	2,345	2,380	1,928	2,426	2,258
	FP	81	46	498	0	168
	FN	103	52	135	111	79
	TN	2,323	2,374	2,291	2,315	2,347
	CCR	96.21%	97.98%	86.95%	97.71%	94.91%

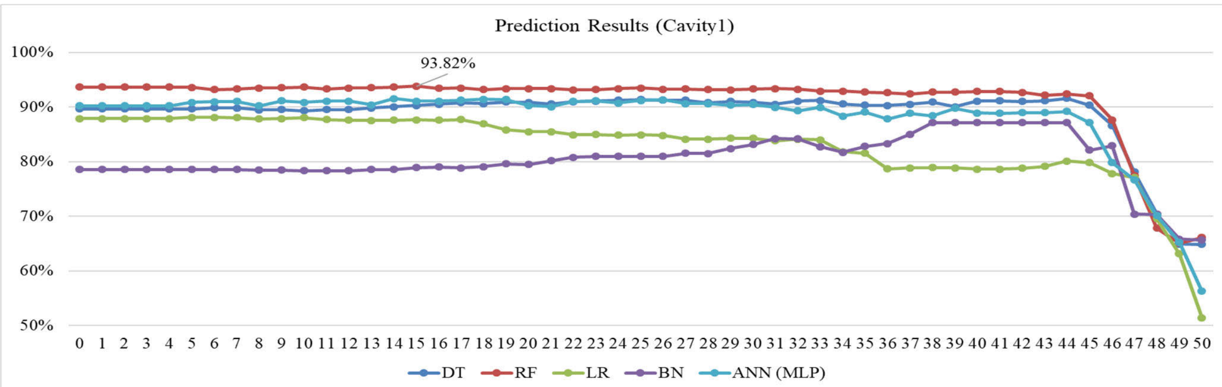

FIGURE 7. The backward elimination of variables based on the gain ratio for Cavity1 (10-fold cross validation).

the 10-fold cross validation were as follows: DT 92.25%; RF 93.25%; LR 88.09%; BN 90.46%; and ANN (MLP) 92.83%. On the other hand, when applying that technique to the order of boosting-bagging, the predictive results were: DT 91.17%; RF 92.83%; LR 88.12%; BN 89.24%; and ANN (MLP) 91.74%. Thus, applying the double ensemble technique showed an increase in accuracy of +0.67%p, −0.57%p, +0.00%p, +3.30%p, and +1.29%p, respectively, compared to the single models. Particularly, the accuracy of BN and ANN (MLP) improved significantly. Further still, the average predictive accuracy of the double ensemble model was 91.38% for bagging-boosting and 90.62% for boosting-bagging, an improvement of +0.94%p and +0.18%p, respectively, compared to the average predictive accuracy of the single models (90.44%), as shown in Table 10.

These results clearly indicate that the double ensemble technique, especially when using the bagging-boosting order, can improve the predictive accuracy of weight defect prediction in the Cavity1 dataset. They also highlight the advantage of combining different ensemble techniques to further enhance the performance of the single models.

2) CAVITY2 DATA ANALYSIS

For Cavity2, we applied the important variable combinations derived from every one of the single models to the ensemble models using bagging and boosting techniques. The predictive results obtained through 10-fold cross validation for the bagging ensemble were as follows: DT 96.46%; RF 97.47%; LR 87.10%; BN 97.71%; and ANN (MLP) 95.42%. Similarly, the predictive results for the boosting ensemble were:

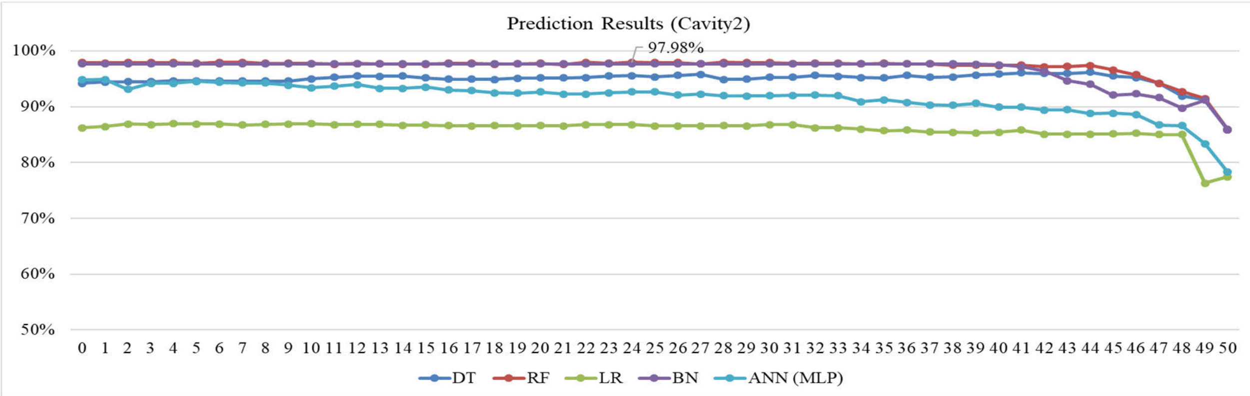

FIGURE 8. The backward elimination of variables based on the gain ratio for Cavity2 (10-fold cross validation).

TABLE 10. The prediction results for the ensemble and the double ensemble model for Cavity1 (10-fold cross validation).

Classifier		DT	RF	LR	BN	ANN (MLP)	Average
Ensemble	Bagging	91.84%	93.50%	88.19%	87.64%	92.67%	90.77%
	Boosting	91.17%	93.60%	88.12%	89.69%	91.58%	90.83%
Double Ensemble	Bagging-Boosting	92.25%	93.25%	88.09%	90.46%	92.83%	91.38%
	Boosting-Bagging	91.17%	92.83%	88.12%	89.24%	91.74%	90.62%

TABLE 11. The prediction results for the ensemble and the double ensemble model for Cavity2 (10-fold cross validation).

Classifier		DT	RF	LR	BN	ANN (MLP)	Average
Ensemble	Bagging	96.46%	97.47%	87.10%	97.71%	95.42%	94.83%
	Boosting	96.66%	97.51%	86.95%	97.67%	94.91%	94.74%
Double Ensemble	Bagging-Boosting	97.34%	97.05%	87.02%	98.10%	95.65%	95.03%
	Boosting-Bagging	97.16%	97.82%	87.10%	97.55%	96.89%	95.30%

TABLE 12. The prediction results for the single model (Split 66%).

Type	Result	DT	RF	LR	BN	ANN (MLP)	Average
Cavity1	Accuracy	92.00%	94.07%	88.23%	87.95%	92.18%	90.89%
	Important variables	9	28	46	7	43	26.6
Cavity2	Accuracy	96.24%	98.67%	88.06%	98.24%	95.58%	95.36%
	Important variables	24	19	42	12	49	29.2

DT 96.66%; RF 97.51%; LR 86.95%; BN 97.67%; and ANN (MLP) 94.91%. Similar to Cavity1, there were slight variations in predictive accuracy compared to the single models, but overall, there were no significant changes in performance when the ensemble techniques were used.

However, when applying the double ensemble technique to the order of bagging-boosting, the predictive results in 10-fold cross validation were: DT 97.34%; RF 97.05%; LR 87.02%; BN 98.10%; and ANN (MLP) 95.65%. On the other hand, when applying that same validation to the order of boosting-bagging, the predictive results were DT 97.16%; RF 97.82%; LR 87.10%; BN 97.55%; and ANN (MLP) 96.89%. The results from applying the double ensemble technique showed an increase in accuracy of +1.13%p; −0.16%p; +0.15%p; +0.39%p; and +1.98%p, respectively, compared to the single models. The accuracy of DT and ANN (MLP) showed notable improvement. Further still, the average predictive accuracy for the double ensemble model

TABLE 13. The prediction results for the ensemble and the double ensemble model for Cavity1 (Split 66%).

Classifier		DT	RF	LR	BN	ANN (MLP)	Average
Ensemble	Bagging	91.15%	93.97%	88.23%	88.04%	91.34%	90.55%
	Boosting	90.49%	93.13%	88.23%	89.83%	92.18%	90.77%
Double Ensemble	Bagging-Boosting	92.09%	93.41%	88.42%	91.05%	92.66%	91.53%
	Boosting-Bagging	91.62%	92.94%	87.85%	89.27%	92.18%	90.77%

TABLE 14. The prediction results for the ensemble and the double ensemble model for Cavity2 (Split 66%).

Classifier		DT	RF	LR	BN	ANN (MLP)	Average
Ensemble	Bagging	97.58%	98.37%	88.24%	98.24%	96.00%	95.69%
	Boosting	97.70%	98.42%	88.06%	98.24%	95.58%	95.60%
Double Ensemble	Bagging-Boosting	98.36%	98.06%	88.12%	98.61%	96.12%	95.85%
	Boosting-Bagging	98.24%	98.61%	87.88%	98.42%	97.52%	96.13%

TABLE 15. The prediction results for the DNN (Cavity1 & Cavity2).

Type	Learning rate	Batch size	Epochs	Accuracy
Cavity1	0.0001	32	500	90.22%
			1,000	91.82%
			1,500	92.57%
		64	500	88.05%
			1,000	89.46%
			1,500	92.47%
Cavity2	0.0001	32	500	95.64%
			1,000	96.00%
			1,500	96.30%
		64	500	95.33%
			1,000	96.00%
			1,500	96.24%

was 95.03% for bagging-boosting and 95.30% for boosting-bagging, which indicated an improvement of +0.28%p and +0.55%p, respectively, compared to the average predictive accuracy of the single models (94.75%), as shown in Table 11. These results suggest that the double ensemble technique, especially when using the boosting-bagging order, can improve the predictive accuracy of weight defect prediction in Cavity2 dataset. It also emphasizes the advantage of combining different ensemble techniques to further enhance the performance of the single models.

Based on these research findings, we obtained the answer to RQ2. In both Cavity1 and Cavity2, the RF algorithm demonstrated the highest accuracy among the single models. Additionally, for the other four ML techniques (DT, LR, BN, and ANN (MLP)), the double ensemble model showed higher predictive accuracy when compared to the single models. In Cavity1, there was a maximum improvement of +3.30%p, and in Cavity2, there was a maximum improvement of +1.98%p. On average, there was an improvement of +0.94%p and +0.55%p, respectively. However, it is also interesting to note that applying the double ensemble technique actually resulted in a decrease in accuracy when for the RF technique. This result can be attributed to the fact that the RF algorithm already incorporates ensemble characteristics through a DT ensemble, making the application of the double ensemble technique less effective at improving accuracy.

C. AN ADDITIONAL EXPERIMENTS

Following the same approach, we performed additional experiments on Cavity1 and Cavity2 datasets by randomly dividing them into training data (66%) and test data (34%). In additional experiments, we aimed to investigate the impact of different approaches for classifying the dataset into training data and testing data on the prediction accuracy of various models. The results exhibited a consistent pattern, with the double ensemble model consistently outperforming the single models for predictive accuracy. These findings are supported by the data presented in Table 12 to Table 14 here. These additional experiments also provide robust evidence that supports the conclusion that the double ensemble model offers superior predictive performance over the single models. This knowledge can helpfully inform future research and practical applications on weight defect prediction for injection molding processes.

In addition, we performed experiments on the Cavity1 and Cavity2 datasets using the state-of-the-art deep neural network (DNN) algorithm. We set the number of hidden layers to 4, the number of nodes in each hidden layer to 512, and the probability of drop-out to 0.5. Then we initialized the weights using 'glorot-uniform', and used 'relu' as an activation function. We experimented with adjusting the learning rate to 0.001, 0.0001, the batch size to 32, 64 and the number of epochs to 500, 1,000, 1,500 to find a model with better performance. As a result, the best performance was achieved with a batch size of 32 and 1,500 epochs in both Cavity1 and Cavity2 datasets, reaching 92.57%, and 96.30%, respectively, as shown in Table 15. However, double ensemble-based ANN (MLP) showed slightly higher prediction performance like Cavity1 92.83%, Cavity2 96.89%, in 10-fold cross validation, Cavity1 92.66%, Cavity2 97.52% in split 66%. The results show that the double ensemble model performs better than the DNN on this dataset.

V. CONCLUSION

This study sought to enhance the prediction performance of weight defects in injection-molded products using ML models. We constructed single models by applying various algorithms, including DT, RF, LR, BN, and ANN (MLP). Additionally, we evaluated prediction accuracy by using ensemble techniques, specifically bagging and boosting, after determining the optimal combinations of these variables in the single models. Finally, we compared the prediction accuracy between the double ensemble model, created by combining bagging and boosting in different orders and the single models. The results consistently demonstrated that the double ensemble model outperformed the single models for prediction accuracy. For the analysis of Cavity1 data, the double ensemble model based on BN and ANN (MLP) was used and exhibited improved prediction accuracy compared to the single models using BN and ANN (MLP). Similarly, for the Cavity2 data analysis, the double ensemble model based on DT and ANN (MLP) demonstrated enhanced prediction effectiveness when compared to the single models using DT and ANN (MLP). Moreover, when comparing the average prediction accuracy for these models, the double ensemble model consistently achieved higher accuracy than did the single models.

These research findings also revealed the importance of specific variables in influencing the weight of injection-molded products. Through the evaluation of the gain ratio, barrel H2 temp real, metering time, and fill time were identified as three significant variables. It was also observed that by managing the weight below the control target value of 1,260g (with the average weight of Cavity1 products being 1,276g), a material saving effect of 1.2% (16g) can be achieved. This result emphasizes the practical significance of identifying key variables that can impact product weight in injection molding. By effectively managing the product weight to meet or stay below the control target value material savings, quality improvement, and productivity enhancement are benefits that can be realized. Another key finding of the research is the effectiveness of the double ensemble model in improving prediction accuracy in the field of injection molding. However, it is also important to note that the applicability of the double ensemble technique may vary depending on the specific problem at hand. It is thus crucial to carefully select the most suitable technique and adjust it accordingly whenever developing prediction models.

Based on our research findings, it was determined that adopting the double ensemble model in the field of injection molding can improve the prediction accuracy of product weight and enhance the prediction accuracy of various aspects related to quality. This approach enables the early detection of defects based on production data-driven equipment operation, thereby preventing continuous defects and reducing the cost of quality failures and deriving optimal production conditions to ultimately establish a smart factory with better cost competitiveness and improvements in productivity. This particular approach thus goes beyond having to rely solely on the expertise of field workers.

In this study, we have collected data focusing on temperature and pressure. More various injection conditions have to be considered for prediction accuracy. In our future research, we plan to enhance this prediction accuracy by considering additional variables, such as melt index, hopper temperature, and injection speed. Additionally, we will explore optimizing the parameters in the double ensemble model to construct a more refined prediction model. Through these endeavors, we hope to achieve even higher prediction accuracy and thereby improve the overall effectiveness of our new approaches.

REFERENCES

[1] X. Xu and Q. Hua, "Industrial big data analysis in smart factory: Current status and research strategies," *IEEE Access*, vol. 5, pp. 17543–17551, 2017.

[2] T. M. Fernández-Caramés and P. Fraga-Lamas, "A review on human-centered IoT-connected smart labels for the Industry 4.0," *IEEE Access*, vol. 6, pp. 25939–25957, 2018.

[3] J. Wan, M. Sxia, J. Hong, Z. Pang, B. Jayaraman, and F. Shen, "IEEE access special section editorial: Key technologies for smart factory of Industry 4.0," *IEEE Access*, vol. 7, pp. 17969–17974, 2019.

[4] A. J. Román, S. Qin, V. M. Zavala, and T. A. Osswald, ''Neural network feature and architecture optimization for injection molding surface defect prediction of model polypropylene,'' *Polym. Eng. Sci.*, vol. 61, no. 9, pp. 2376–2387, Sep. 2021.

[5] P. S. Harsha, J. S. Chohan, and H. S. Gill, ''Machine learning based comparative analysis and prediction for sustainable additive manufacturing,'' in *Proc. 7th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Nov. 2022, pp. 558–563, doi: 10.1109/PDGC56933.2022.10053152.

[6] Z. Ren, Y. Li, Z. Wu, and S. Xie, ''Deep learning-based predictive control of injection velocity in injection molding machines,'' *Adv. Polym. Technol.*, vol. 2022, Nov. 2022, Art. no. 7662264, doi: 10.1155/2022/7662264.

[7] Y.-M. Deng, Y. C. Lam, S. B. Tor, and G. A. Britton, ''A CAD-CAE integrated injection molding design system,'' *Eng. Comput.*, vol. 18, no. 1, pp. 80–92, Apr. 2002.

[8] H.-S. Peng, P.-W. Huang, K.-F. Liew, P.-C. Chien, and W.-J. Su, ''Study of CAE predictive analytics on structure design, molding and improved strength of plastic injection molded parts,'' in *Proc. Int. Conf. Appl. Syst. Innov. (ICASI)*, May 2017, pp. 546–549, doi: 10.1109/ICASI.2017.7988479.

[9] F. Hentati, I. Hadriche, N. Masmoudi, and C. Bradai, ''Optimization of the injection molding process for the PC/ABS parts by integrating Taguchi approach and CAE simulation,'' *Int. J. Adv. Manuf. Technol.*, vol. 104, nos. 9–12, pp. 4353–4363, Aug. 2019.

[10] H.-H. Tsou, K.-D. Lee, Z.-H. Wang, and C.-C. Huang, ''The feature extraction modeling of product analysis in injection molding for intelligent manufacturing,'' in *Proc. Int. Comput. Symp. (ICS)*, Dec. 2020, pp. 325–329, doi: 10.1109/ICS51289.2020.00071.

[11] H. Lee, K. Ryu, and Y. Cho, ''A framework of a smart injection molding system based on real-time data,'' *Proc. Manuf.*, vol. 11, pp. 1004–1011, Jun. 2017.

[12] C. Shen, L. Wang, and Q. Li, ''Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method,'' *J. Mater. Process. Technol.*, vol. 183, nos. 2–3, pp. 412–418, Mar. 2007.

[13] M.-S. Huang, K. C. Ke, and P. W. Wu, ''A machine learning method for predicting part weight, dimensions, and residual stress during injection molding,'' in *Proc. 25th Int. Conf. Mechatronics Technol. (ICMT)*, Nov. 2022, pp. 1–14, doi: 10.1109/ICMT56556.2022.9997777.

[14] B. Silva, J. Sousa, and G. Alenya, ''Machine learning methods for quality prediction in thermoplastics injection molding,'' in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Dec. 2021, pp. 1–6, doi: 10.1109/ICECET52533.2021.9698455.

[15] K.-C. Ke and M.-S. Huang, ''Quality prediction for injection molding by using a multilayer perceptron neural network,'' *Polymers*, vol. 12, no. 8, p. 1812, Aug. 2020.

[16] R. D. Párizs, D. Török, T. Ageyeva, and J. G. Kovács, ''Machine learning in injection molding: An Industry 4.0 method of quality prediction,'' *Sensors*, vol. 22, no. 7, p. 2704, Apr. 2022.

[17] J. Lee, D. Yang, K. Yoon, and J. Kim, ''Effects of input parameter range on the accuracy of artificial neural network prediction for the injection molding process,'' *Polymers*, vol. 14, no. 9, p. 1724, Apr. 2022.

[18] B. Ozcelik and T. Erzurumlu, ''Comparison of the warpage optimization in the plastic injection molding using ANOVA, neural network model and genetic algorithm,'' *J. Mater. Process. Technol.*, vol. 171, no. 3, pp. 437–445, Feb. 2006.

[19] A. Mollaei Ardestani, G. Azamirad, Y. Shokrollahi, M. Calaon, J. H. Hattel, M. Kulahci, R. Soltani, and G. Tosello, ''Application of machine learning for prediction and process optimization—Case study of blush defect in plastic injection molding,'' *Appl. Sci.*, vol. 13, no. 4, p. 2617, Feb. 2023.

[20] V. Ramana, ''Data mining based approach for quality prediction of injection molding process,'' *Int. J. Eng. Technol.*, vol. 9, no. 3, pp. 2220–2224, Jun. 2017.

[21] A. Polenta, S. Tomassini, N. Falcionelli, P. Contardo, A. F. Dragoni, and P. Sernani, ''A comparison of machine learning techniques for the quality classification of molded products,'' *Information*, vol. 13, no. 6, p. 272, May 2022.

[22] H. Jung, J. Jeon, D. Choi, and J.-Y. Park, ''Application of machine learning techniques in injection molding quality prediction: Implications on sustainable manufacturing industry,'' *Sustainability*, vol. 13, no. 8, p. 4120, Apr. 2021.

[23] A. S. Struchtrup, D. Kvaktun, and R. Schiffers, ''Adaptive quality prediction in injection molding based on ensemble learning,'' *Proc. CIRP*, vol. 99, pp. 301–306, May 2021.

[24] T. Ahmed, P. Sharma, C. L. Karmaker, and S. Nasir, ''Warpage prediction of injection-molded PVC part using ensemble machine learning algorithm,'' *Mater. Today, Proc.*, vol. 50, pp. 565–569, Jan. 2022.

[25] J. Kwon, K. Choi, and Y. Suh, ''Double ensemble approaches to predicting firms' credit rating,'' in *Proc. Pacific Asia Conf. Inf. Syst. (PACIS)*, Jeju Island, (South) Korea, 2013, p. 158.

[26] S. Lee and D. Yoo, ''Using double ensemble techniques to predict the insolvency of SMEs based on technological feasibility assessment information,'' *J. SME Finance*, vol. 42, no. 2, pp. 3–28, 2022, doi: 10.33219/jsmef.2022.42.2.001.

KWANGMO KOO received the B.S. degree from the School of Electronic and Electrical Engineering, Hanyang University, Seoul, South Korea, in 1999. He is currently pursuing the M.S. degree in management of technology (MOT) with Gyeongsang National University, Jinju, South Korea. He was an Engineer with the DRAM Product Engineering Department, Semiconductor Division, Samsung Electronics, South Korea, from 1999 to 2001. Since 2008, he has been a Leader in various teams, including the Advanced Development Team, Smart Factory Team, Shinsung Delta Tech, South Korea. His research interests include machine learning, data analysis, and smart factory.

KEUNHO CHOI received the Ph.D. degree in MIS from the Korea University Business School (KUBS), in 2013. He was a Research Fellow with the Labor Welfare Research Institute, Korea Workers' Compensation and Welfare Service (COMWEL), from 2014 to 2018, and performed various research projects on the machine learning and big data analysis. He is currently an Associate Professor with the Department of Business Administration, Hanbat National University (HBNU), Daejeon, South Korea. His work has appeared in *Knowledge-Based Systems*, *Expert Systems With Applications*, *Journal of Information Science*, *Electronic Commerce Research and Applications*, *The Data Base for Advances in Information Systems*, *Data Technologies and Applications*, *Information Systems and e-Business Management*, and other journals. His research interests include data mining, artificial intelligence, recommendation systems, and big data analysis.

DONGHEE YOO received the Ph.D. degree in MS/IS from the Korea University Business School, in 2009. He was an Assistant Professor with the Department of Electronics Engineering and Information Science, Korea Military Academy (KMA), from 2009 to 2013. He was a Research Professor with Yonsei University. He is currently a Professor with the Department of Management Information Systems (MIS), Gyeongsang National University, Jinju, South Korea. His work has appeared in *Knowledge-Based Systems*, *Expert Systems With Applications*, *Journal of Information Science*, *Electronic Commerce Research and Applications*, and other journals. His research interests include artificial intelligence, deep learning, text mining, knowledge graph, ontology, and big data analysis.