IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 December 2022, accepted 7 December 2022, date of publication 12 December 2022, date of current version 15 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228241

RESEARCH ARTICLE

A Data-Driven Recipe Simulation for Synthetic Rubber Production

KIKUN PARK, HANBYEOUL PARK, AND HYERIM BAE, (Member, IEEE)
Department of Industrial Engineering, Pusan National University, Busan 46241, South Korea

Corresponding author: Hyerim Bae (hrbae@pusan.ac.kr)

This work was supported by the Ministry of Science and ICT (MSIT), South Korea, under the Grand Information Technology Research Center Support Program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2022-2020-0-01791.

ABSTRACT To manufacture synthetic rubber, rubber manufacturers require optimal recipes to ensure that it satisfies the required quality standards. Several experiments are required to create the optimal recipe, which adversely affects not only the cost and time required but also the health of workers. Suppose the experimental results can be predicted in advance at the recipe design stage before direct experimentation. In that case, the cost of the experiment can be reduced, and the workers' health can be significantly less impacted. For this purpose, a method called the prediction walk model using a machine learning model was developed to generate the temperature trajectory in a kneading machine. A cross-updating method to predict the quality of the kneading operation is also proposed. From the results of the experiment, it was confirmed that the performance of the proposed models was superior to that of the existing prediction models.

INDEX TERMS Synthetic rubber, rubber manufacturing, synthetic rubber recipe, prediction walk model.

I. INTRODUCTION

Synthetic rubber is a vital product in several manufacturing industries. Because it is used in various ways, the quality standards required vary depending on the product. It is difficult to design different optimal recipes to satisfy the quality requirements of various products. The synthetic rubber manufacturing process comprises nine steps: mastication, kneading, extrusion, calendaring, stamping, sealing, molding, vulcanization, and finishing. To design an optimal recipe for all these processes, the number of experiments required and the cost increase significantly [1]. Additionally, the synthetic rubber manufacturing environment adversely affects the health of workers [2]. Therefore, reducing the number of experiments required is important for company profitability, and a solution to this problem is considered in this study. A data-driven method was introduced that can generate results through simulations of the recipe design without conducting experiments. This study contributes to the reduction of the time and cost required for optimal recipe design by predicting the experimental results in advance. A simulation method that uses data from the kneading operation, which corresponds to the second stage of the synthetic rubber manufacturing process, is presented.

The kneading operation is the process of compounding rubber by mixing and dispersing ten or more types of solid or liquid compounding agents, such as cross-linking agents, fillers, and vulcanization accelerators. During the kneading process, the operator frequently checks the temperature of the material, and when a certain temperature is obtained, the material is placed in a Banbury mixer, which mixes the materials during the kneading operation. In the Banbury mixer, heat is generated as the rotor rotates and the temperature inside the channel increases. When the temperature exceeds the critical point, vulcanization of the rubber occurs, which deteriorates the quality of the synthetic rubber. Therefore, it is extremely important to design the process such that the temperature does not exceed the critical point in the optimal recipe's design stage. Hence, the proposed simulation method predicts the degree of temperature change to determine whether the temperature exceeds the critical point for the recipe design. A prediction walk model (PWM)

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

that predicts the temperature change by simultaneously predicting the kneading operation temperature and time is introduced. As previously mentioned, the temperature change during the kneading operation significantly affects the quality of the synthetic rubber. Based on this characteristic, we predicted the quality level of the synthetic rubber using the temperature change predicted by the PWM. To predict the results of the kneading operation, we introduce a cross-updating prediction, which refers to cross-predicting synthetic rubber qualities that are highly correlated with each other. It was confirmed that this method improves the prediction performance compared with the existing prediction methods.

Because the proposed data-driven method can accurately predict temperature changes and the results of the kneading operation, it has the advantage of simulating recipes for various combinations. It can also reduce the number of experiments required by filtering recipe candidates diagnosed as having poor quality levels. Thus, synthetic rubber manufacturers can experience a positive effect on operational costs and the health of workers. The remainder of this paper is organized as follows. Section II introduces related studies and explains the need for a new approach to time series data prediction. Section III describes the data used for the training, PWM, and cross-updating prediction. Section IV presents the experimental results to prove the performance of the proposed model, and finally, Section V contains the conclusions and limitations of the study.

II. RELATED WORKS AND PROBLEM DESCRIPTION

A. PREDICTION OF MACHINE TEMPERATURE

Because machine temperature prediction is important in the manufacturing industry, several studies regarding it have been conducted in various fields for a long time. Furthermore, several of these studies were based on machine learning (ML) and neural networks (NNs). The temperature prediction algorithms using ML have exhibited satisfactory performance. Previous research has improved air conditioner operation efficiency through temperature prediction of data centers using regression analysis and support vector machines (SVM) [19]. Moreover, support vector regression (SVR) is applied to the hydration heat of mass concrete temperature prediction [28]. An artificial neural network (ANN) has been applied to predict the temperature change inside the tunnel [20]. A methodology to solve the problem of insufficient data for predicting the heat of a machine was also introduced. In this study, the problem was solved using an ensemble model [18]. Because several predictive models have been utilized for temperature prediction, a study comparing ML models and ANN has also been published [21], [22], [23]. According to previous research, convolution neural networks (CNNs), recurrent neural network (RNNs), and long short-term memory (LSTM) NNs exhibit good performance in predicting the temperature of transmission modes [24]. Similarly, a previous study showed that gated recurrent unit (GRU) outperformed SVR and ANN for air temperature prediction [29]. However, ML and DL models are inherently deterministic and often yield overconfident results [26]. Thus, a hybrid ML model was proposed to overcome this limitation [30]. In terms of predicting the temperature of the kneading operation, this disadvantage of ML and DL models leads to the issue of realistic predictions not being presented, such as an excessive predicted manufacturing time or temperature. Therefore, it is necessary to devise a prediction method that overcomes this problem. In this paper, we introduce a method for correcting the prediction values using the PWM.

B. TIME-SERIES FORECASTING MODEL

The temperature change problem during the kneading operation includes values that change with time. This problem is a type of time-series forecasting problem, but the temperature prediction of the kneading operation has slightly different properties. Traditional time-series models forecast the future by using past information. However, the temperature change in the kneading operation does not contain past information at the start of the experiment. Therefore, even if our problem handles temporal temperature changes, it must be estimated using past experimental results and recipe information. Furthermore, at the start of the kneading process, the start temperature must be reset, and the change in temperature for the process must be predicted using the features affecting the temperature change.

Because we do not use historical data as a time series, the volume of the feature data may not be sufficient to train our model. To obtain good ML predictive models, sufficient data must be applied for learning to converge to a global minimum. However, when the size of the collected data is limited, a particular type of ML model is required [4]. Therefore, it is necessary to consider a methodology that predicts the expected results of an experiment by only using limited information. When the temperature is predicted with time series characteristics, it is known that applying non-linear model results in better performance than linear models. [5]. Therefore, a method that employs a non-linear time series model is widely used to predict the future temperature. For applying non-linear predictive models, we must consider the model's reproducibility and reliability because the model's performance is affected by the environment and conditions [6]. The model must also be robust. Especially in studies of time series data analysis, such as the prediction of sea surface temperature, the robustness of the model needs to be considered when learning the ML model [7]. In other studies, temperature prediction algorithms that take advantage of stochastic properties have been introduced because the temperature is not deterministic [8]. Therefore, a new approach is required to compensate for the shortcomings of time-series prediction models and reflect the conditions of various recipes. The PWM presented in this study secured robustness while avoiding problems that may occur in time-series data prediction.

C. FEATURE EXTRACTION FOR ML

Appropriate features must be used for learning to train an effective ML predictive model. These input features should be selected by considering the correlation between the input variables and the target value. According to previous research, the commonly-used features for machine learning can be extracted from the industry domain [25]. In the kneading process, several features are collected, from which the features that can play an important role in prediction must be selected. The quality of synthetic rubber is sensitive to temperature changes during the kneading operation. Therefore, we define the temperature change predicted by the PWM as a trajectory to predict the quality of the synthetic rubber. Trajectory refers to information that records the movement path of objects over time and can be collected in various ways [10]. Because trajectory data are expressed in the form of a sequential matrix, they must be appropriately converted to a scalar value for use in ML prediction. Therefore, we convert them by measuring the similarity between two trajectories through methods such as dynamic time wrapping, Hausdorff distance, and Frechet inception distance [11], [12].

D. ML MODELS

Several ML models can be applied to predict the temperature and quality of kneading operations. In this study, we apply seven ML models: linear regression (LR), general additive model (GAM), random forest (RF), support vector regressor (SVR), recurrent neural network (RNN), long-term memory (LSTM) and Seq2Seq. LR is a predictive method that finds weights that minimize the sum of squares of the residual [13]. GAM is a predictive model that considers non-linear relationships while increasing variance to improve the critical point where the linear model has a high bias [14]. RF is a methodology proposed to solve the overfitting problem by applying the ensemble method to the proposed decision tree [15]. SVR is a predictive model that has the advantage of encoding the non-linear relationship of input-output data by mapping the input data into a high-dimensional space [16]. The RNN is a natural generation of feedforward neural network to sequences [27]. The RNN is configured to transmit the information of the hidden layer to the next layer. Owing to these characteristics, the RNN is mainly used to predict time-series data. LSTM is a neural network that improves the vanishing gradient problem of the RNN and is also widely used for time-series data prediction [17]. The RNN exhibits the limitation that the output length should be fixed. Seq2Seq is known as a model that overcomes these shortcomings [27]. Seq2Seq is a neural network that offers the advantage of deriving a variable-length output value from the sequence data prediction problem. In this study, the PWM that was constructed exhibited excellent performance compared with LR, GAM, RF, and SVR in temperature change prediction, and its performance was also compared with that of RNN, LSTM and Seq2Seq.

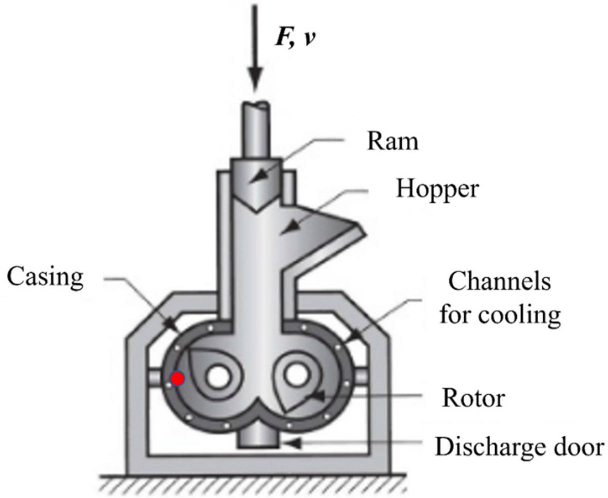


FIGURE 1. Banbury mixer.

III. METHODS AND MATERIALS

In this section, we introduce a new method of predicting the temperature change during the kneading operation and the quality of synthetic rubber. This method comprises two parts: first, an ML-based predictive model using historical experimental and recipe information, called PWM, was developed to predict the temperature changes appearing in the time series. PWM uses ML models to predict the temperature changes during the kneading operation. The temperature changes consist of five features: max temperature, max temperature time, minimum temperature, minimum temperature time, and end temperature. A temperature trajectory can be generated using these five features. Second, a cross-updating method for predicting the quality of the rubber product as the output of the kneading operation was introduced. The cross-updating method can improve the predictive performance of the ML model by continuously updating the predictions using each other's prediction values.

A. DATA DESCRIPTION

PWM uses historical information from the experiments and recipe information of the kneading operations as inputs. The kneading operation uses a Banbury mixer machine, which operates two rotors and mixes the rubber. Figure 1 shows a Banbury mixer and its primary components: the hopper, ram, rotor, and discharge door. In the Banbury mixer, the materials are first inserted through the hopper. If the machine's temperature is high because of previous work, it needs to be sufficiently cooled. Then, a worker lowers the ram to apply pressure and turns the rotor to start mixing. When the rotor begins to rotate, the temperature inside the channel rises with the mixing of rubber, and when the temperature is too high, the worker lifts the ram to cool down the system. Thereafter, the worker repeats the operation according to the recipe. During the kneading operation, the Banbury mixer records the temperature and voltage every second.

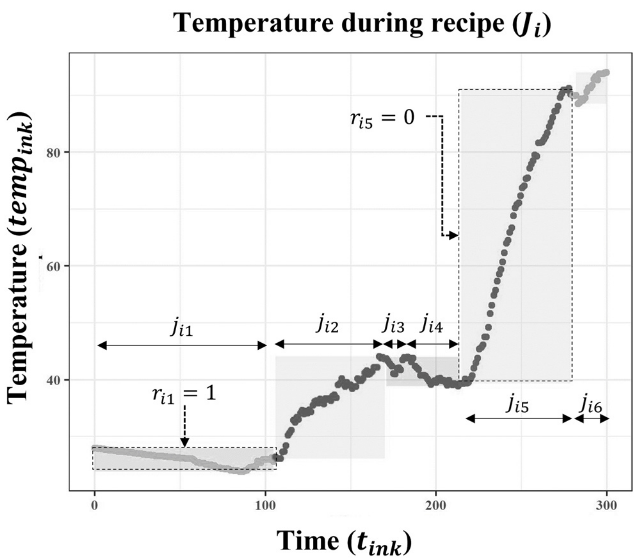


FIGURE 2. Temperature change of channels during kneading operations.

The features extracted from the kneading operation are summarized as follows:

J_i: A job sequence of the i-th recipe, $J_i = \{j_{in} | n = 1, \ldots, N\}$, j_{in} is the n-th job in J_i

t_{ink}: The k-th time during the execution of j_{in}, $k = 1, 2, \cdots, K$

tp_{ink}: Temperature recorded at t_{ink}

v_{ink}: Voltage recorded at t_{ink}

r_{in}: Binary variable representing whether ram is open during j_{in} ($r_{in} = 1$ when ram is open, otherwise $r_{in} = 0$)

δ_i: Tan-delta value of J_i

ht_i: High torque value of J_i

lt_i: Low torque value of J_i

hd_i: Hausdorff distance between the reference schedule and trajectory of J_i

dtw_i: Dynamic time-wrapping distance between reference schedule and trajectory of J_i

fre_i: Frechet inception distance between the reference schedule and trajectory of J_i

J_i refers to the recipe for the i-th kneading operation and it comprises several work schedules (j_{in}). For each schedule, the kneading operation time (t_{ink}), temperature (tp_{ink}), and voltage (v_{ink}) are recorded. r_{in} is also recorded to indicate whether ram was open or closed at the time. The kneading operation is performed for approximately 3–300 s, as shown in Figure 2. Figure 2 shows the data collected during the operation of the Banbury mixer. For J_i, the x-axis represents the kneading operation time (t_{ink}) and the y-axis represents the temperature (tp_{ink}), which is affected by the voltage and operation time of the rotor, and whether the ram is open or closed.

For example, in Figure 2, the temperature in the first work schedule (j_{i1}) decreases because the ram is opened for cooling before operating the rotor from 0–100 s. In this case, because ram is open, it is expressed as $r_{i1} = 1$. After the cooling is finished, the worker injects the material and turns on the rotor, and the heat rises (j_{i2}). In the 5^{th} work schedule (j_{i5}), the temperature rapidly increases because the ram is closed, and the rotor is operating. It should be noted that the temperature is affected not only by the ram being open but also by the voltage (v_{ink}).

To measure the quality level of synthetic rubber, three indicators are used. The first is tan-delta (δ_i), which is a value defined by the loss and storage moduli, and it numerically represents the elasticity and viscosity of rubber. The other two are low torque (lt_i) and high torque (ht_i). Low torque (lt_i) is defined as the torque value when tan delta (δ_i) is in the lower 10%, whereas high torque (ht_i) is defined as the torque value when tan delta (δ_i) is in the upper 10%. The quality of synthetic rubber is measured by the torque values, which represent the force required for rotation. From the torque values, we can conclude that the viscosity increases as the synthetic rubber is mixed.

One of our goals is to predict the quality of synthetic rubber products from the kneading process.

To predict quality measurements such as δ_i, lt_i, and ht_i, we use a reference trajectory and the similarity of each trajectory to the reference. Three additional indicators: Hausdorff distance (hd_i), Dynamic time wrapping (dtw_i) and Frechet inception (fre_i), were also used. These indicators are detailed in Section III.C.

B. DATA PRE-PROCESSING

To train the proposed prediction algorithm, the input data must be pre-processed. Noise and outliers were removed from the experimental data. To predict the temperature change and quality according to the recipe, raw data were changed to the form of a schedule. This process is illustrated in Figure 3. The table on the upper-left side of Figure 3 lists the raw data described in Section III.A. The raw data contains the job number (J_i), schedule number (J_{in}), machine operation time (t_{ink}), temperature of Banbury machine (tp_{ink}), input voltage (v_{ink}), and whether Ram is open or not (r_{in}). The table on the lower left side lists the result parameters of pre-processing the raw data. The temperature at the start, maximum, minimum, and endpoints of each work schedule of the recipe were prepared as the trained data of our learning algorithm. Next, the time to reach the maximum and minimum temperatures of each work plan was calculated, and the total working time and voltage were obtained. In the plot on the right in Figure 3, the red circles indicate the start, maximum, minimum, and end temperatures for each work schedule. The x-axis represents the time taken to reach the temperature and the total time. When the prediction was completed for all work schedules, a temperature trajectory was generated.

C. METHODOLOGY

The quality of synthetic rubber can be predicted by evaluating the temperature change during a process because this temperature substantially affects rubber quality. This study aims to predict the temperature change within the rubber

Raw Data	
Index	Features
1	Job No (J_i)
2	Schedule No (J_{in})
3	Time (t_{ink})
4	Temperature (tp_{ink})
5	Voltage (v_{ink})
6	RAM On/Off (r_{in})

Preprocessed Data	
Index	Features
1	Job No (J_i)
2	Schedule No (J_{in})
3	Total time (t_{ik})
4	Start Temperature ($tp_{i,n,\min t_{ink}}$)
5	Max Temperature ($\max tp_{ink}$)
6	Min Temperature ($\min tp_{ink}$)
7	End Temperature ($tp_{i,n,\max t_{ink}}$)
8	Start time ($\min t_{ink}$)
9	Max time ($\mathrm{argmax}_{t_{ink}} tp_{ink}$)
10	Min time ($\mathrm{argmin}_{t_{ink}} tp_{ink}$)
11	End time ($\max t_{ink}$)
12	Total voltage ($\mathrm{sum}\ v_{ink}$)
13	RAM On/Off (r_{in})

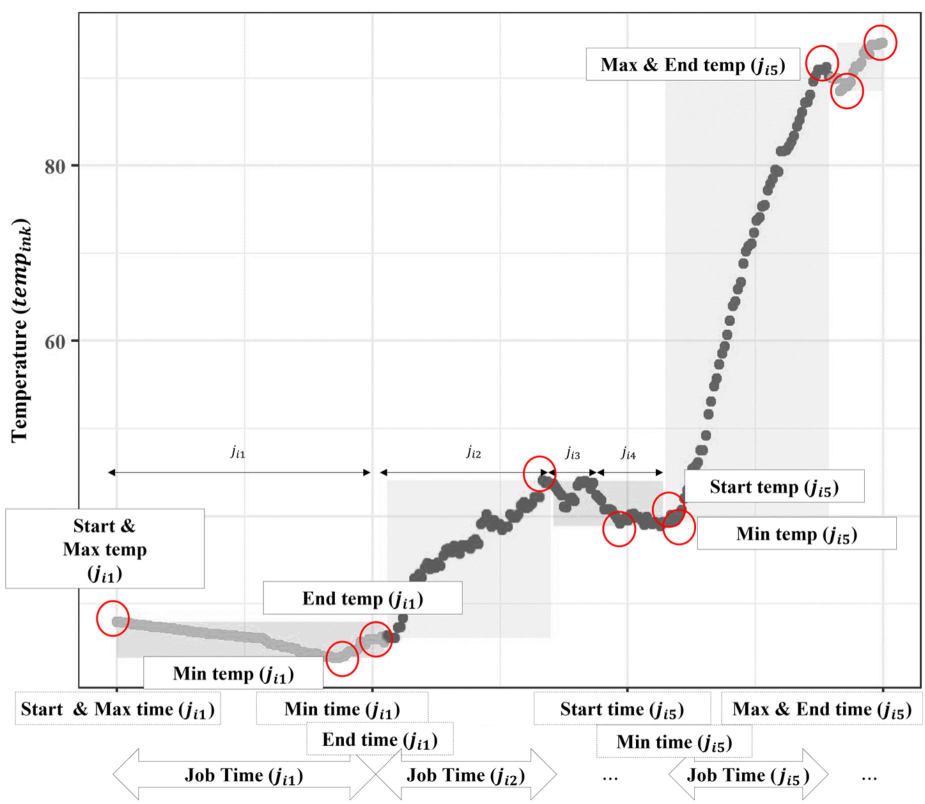
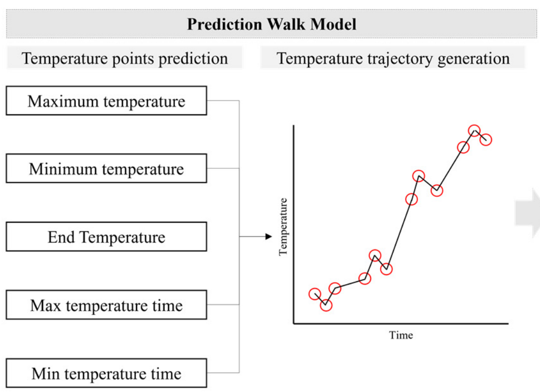


FIGURE 3. Data pre-processing and result of PWM.

Input data

Recipe information

Index	Features
1	Job No (J_i)
2	Schedule No (J_{in})
3	Total time ($t_{inK} - t_{in1}$)
4	Start time (t_{in1})
5	Start temperature (tp_{in1})
6	Total voltage ($\mathrm{sum}\ v_{ink}$)
7	RAM On/Off (r_{in})



Cross updating prediction

Feature extraction

Index	Features
1	Global max temperature ($\max tp_{ink}$)
2	Global min temperature ($\min tp_{ink}$)
3	Total Job Time (t_{iNK})
4	Hausdorff distance (hd_i)
5	Dynamic time wrapping (dtw_i)
6	Frechet inception distance (fre_i)

Quality prediction

Index	Features
1	Tan delta value (δ_i)
2	High torque value (ht_i)
3	Low torque value (ht_i)

FIGURE 4. Methodology.

manufacturing process. A single process of the synthetic process usually consists of multiple jobs. The first job starts when Ram opens and ends when it closes, and simultaneously, the next job starts. The following job lasts until it opens again. Therefore, each job consists of temperature changes between two Ram operations. We conduct predictive analysis for each job by forecasting the maximum temperature (y_1), minimum temperature (y_2), endpoint temperature (y_3), time at maximum temperature (y_4), and time at minimum temperature (y_5). If we connect these results of every job, we can generate a trajectory. We call this procedure PWM, and we apply cross-updating prediction methods to better predict the quality of synthetic rubber for kneading operations. Figure 4 shows the proposed methodology for predicting temperature change and quality according to the recipe of the kneading operation.

According to Figure 4, the quality prediction is made through two stages. The first is to predict the temperature

Algorithm 1 Prediction Walk Model for Generating Temperature Trajectory

Set a recipe J_i consisting of N-th jobs.
n $\leftarrow$ 1
$y_{1n}, y_{2n}, y_{3n,}, y_{4n}, y_{5n} \leftarrow 0$, for all n
1: **while** $n <= N$ **do**
2: **if** n = 1
3: $t_{in1} \leftarrow 1, tp_{in1} \leftarrow 20$
4: **else**
5: $t_{in1} \leftarrow t_{i(n-1)K} + 1, tp_{in1} \leftarrow tp_{i,n-1,K}$
6: **read** $t_{inK}, r_{in}, \sum v_{ink}$
7: $y_{1n} \leftarrow f_1(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink})$ # max temperature
8: $y_{2n} \leftarrow f_2(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink})$ # min temperature
9: $y_{3n} \leftarrow f_3(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink})$ # end temperature
10: $y_{4n} \leftarrow f_4(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink})$ # max temperature time
11: $y_{5n} \leftarrow f_5(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink})$ # min temperature time
12: **if** $y_{1n} < \max(y_{2n}, y_{3n})$; $y_{1n} \leftarrow \max(y_{2n}, y_{3n})$
13: **if** $y_{2n} < \min(y_{1n}, y_{3n})$; $y_{2n} \leftarrow \min(y_{1n}, y_{3n})$
14: **if** $y_{4n} > t_{inK}$; $y_{4n} \leftarrow t_{inK}$; $y_{1n}, y_{3n} \leftarrow mean\,(y_{1n}, y_{3n})$
15: **if** $y_{5n} > t_{inK}$; $y_{5n} \leftarrow t_{inK}$; $y_{2n}, y_{3n} \leftarrow mean(y_{2n}, y_{3n})$
16: $n \leftarrow n + 1$
17: **end while**
18: Generate the temperature trajectory

change of the Banbury mixer using the PWM. PWM is an algorithm that generates temperature change trajectory after training job schedule. Before the kneading operation, operation environments must be set, which include the start temperature (t_{in1}), total time ($t_{inK} - t_{in1}$), total voltage ($sumv_{ink}$), and whether to open or close Ram (r_{in}). Using these features, the PWM predicts the five output variables: maximum temperature (y_1), minimum temperature (y_2), end-point temperature (y_3), maximum temperature achievement time (max time, y_4), and minimum temperature achievement time (min time, y_5) of each work schedule. ML models are used to predict the five values (y_1, y_2, y_3, y_4, y_5), and the model with the best performance is selected. After the prediction is completed for all work schedules, a temperature trajectory is generated by connecting the endpoint of the previous job to the start point of the next job. After generating the trajectory, ML models are used again to predict the quality of rubber using cross updating method. From the generated trajectory, three descriptive features can be obtained: maximum temperature ($\max tp_{ink}$), minimum temperature ($\min tp_{ink}$), and total job time (t_{inK}). Thereafter, three similar measures between the two trajectories are prepared: Hausdorff distance (hd_i), dynamic time wrapping (dtw_i), and Frechet inception distance (fre_i). These six features are used for quality prediction. The objective values

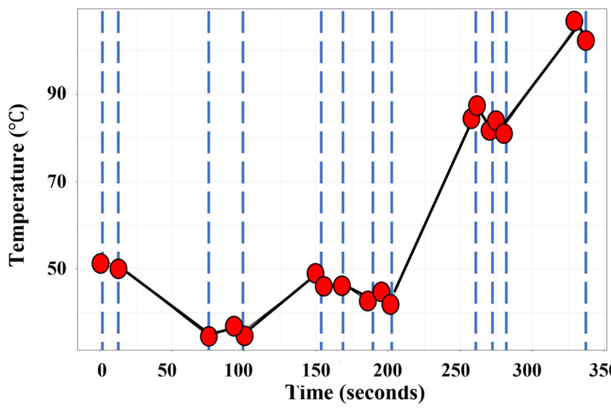


FIGURE 5. Temperature trajectory generated by PWM.

for quality prediction of synthetic rubber are tan-delta (δ_i), low torque (lt_i), and high torque (ht_i), which are described in Section III.A. It was discovered that these three values have a high correlation with each other; therefore, using this, we introduced a method to improve performance by cross-updating their prediction results.

D. TEMPERATURE TRAJECTORY GENERATION USING THE PWM

As shown in Figure 4, the PWM predicts five values to generate a trajectory representing the temperature change. Additionally, as explained in Figure 4, because one recipe consists of n work schedules, the PWM predicts for all n schedules. The features used for prediction utilize the total time ($t_{inK} - t_{in1}$), starting temperature (tp_{in1}), ram state (r_{in}), and total voltage ($\sum v_{ink}$).

$$y_1 = \max tp_{ink} = f_1\left(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink}\right) \tag{1}$$

$$y_2 = \min tp_{ink} = f_2\left(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink}\right) \tag{2}$$

$$y_3 = tp_{inK} = f_3\left(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink}\right) \tag{3}$$

$$y_4 = \underset{t_{ink}}{\operatorname{argmax}}\, tp_{ink} = f_4\left(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink}\right) \tag{4}$$

$$y_5 = \underset{t_{ink}}{\operatorname{argmin}}\, tp_{ink} = f_1\left(t_{inK} - t_{in1}, tp_{in1}, r_{in}, \sum v_{ink}\right) \tag{5}$$

Equations (1)–(5) represent each ML predictive model that predicts the five target values using four features. After the five ML predictive models complete predictions for all work schedules, the prediction values are corrected for trajectory generation. This process is described in Algorithm 1. According to Algorithm 1, the start time of the first work schedule (t_{i11}) was set to 1 s, and the start temperature (tp_{i11}) was set to 20°. The start time (t_{in1}) and the start temperature (tp_{in1}) of the next work schedule were set to the end time and end temperature of the previous work schedule, respectively. Line 6 indicates the process of setting the work schedule information used for prediction. In lines 7-11, the five pre-trained ML predictive models predicted each target value.

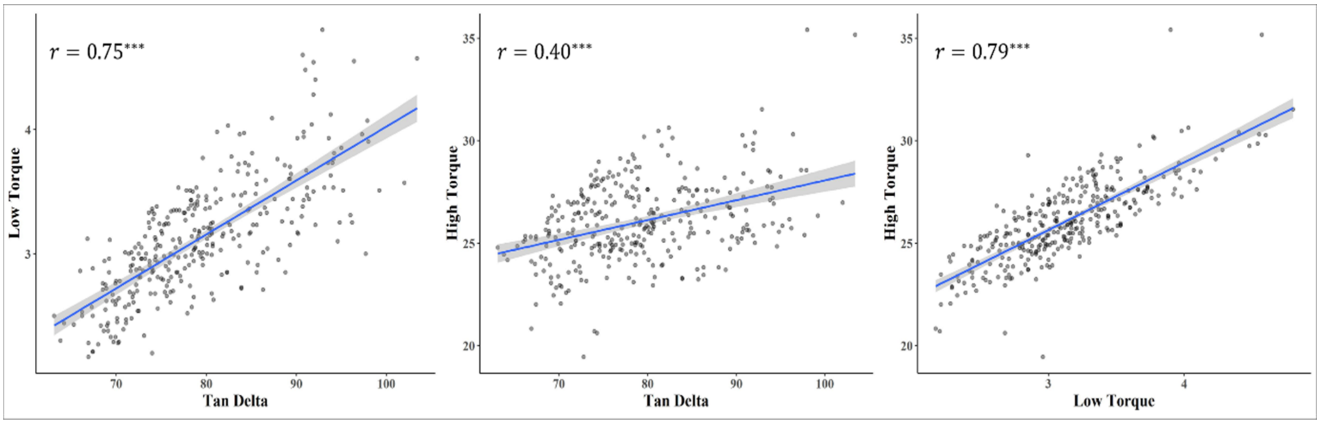


FIGURE 6. **Correlation among quality indicators.**

Then, in lines 12-15, correction is performed for predicted values, which are out of the normal range. For example, if the predicted maximum temperature (y_{1n}) is lower than the other predicted temperatures, it will be reset as the highest value among the lowest and endpoint temperatures. Similarly, if the time to achieve the maximum temperature exceeds the end time of the work schedule, the time will be reset as the mean of the minimum and end temperature. Figure 5 shows the trajectory generated by PWM according to Algorithm 1. In Figure 5, the vertical line is dividing the work schedules (j_{in}) of the recipe (J_{i1}). The red dots indicate the temperature predicted by PWM. The black line connects the red dots. In a certain work schedule, when the predicted values are corrected according to Algorithm 1, the five values are summarized as one or two predicted values.

$$\delta_i \leftarrow f_\delta(lt_i, ht_i, f_\delta(\max temp_i, \min temp_i, t_{inK}, hd_i, dtw_i, fre_i)) \tag{6}$$

$$lt_i \leftarrow f_\delta(\delta_i, ht_i, f_{lt}(\max temp_i, \min temp_i, t_{inK}, hd_i, dtw_i, fre_i)) \tag{7}$$

$$ht_i \leftarrow f_\delta(\delta_{i_i}, lt_i, f_\delta(\max temp_i, \min temp_i, t_{inK}, hd_i, dtw_i, fre_i)) \tag{8}$$

Figure 6 shows the linear relationship between the tan delta, low torque, and high torque. The values have positive significant coefficient correlations of 0.75, 0.40, and 0.79 with each other. We devised a cross-updating prediction using the similarity of the target values.

E. QUALITY PREDICTION USING CROSS-UPDATING PREDICTION

After a trajectory indicating the temperature change using the PWM is created, the quality of the synthetic rubber is predicted. Because the quality of synthetic rubber is extremely sensitive to the temperature change during the kneading operation, six features were defined, as shown in Figure 4. The global maximum ($\max tp_{ink}$) and minimum temperatures ($\min tp_{ink}$), and job time (t_{inK}) are calculated using the generated trajectories. Hausdorff distance (hd_i), dynamic time wrapping (dtw_i) and the Frechet inception distance (fre_i) that indicate the similarity of the trajectories are compared with one reference trajectory and the remaining n-1 trajectories. Predictive ML models used to predict tan-delta (δ_i), low torque (lt_i), and high torque (ht_i) are trained using the defined features. When training was complete, the predictions were cross-updated to improve performance by continuously updating the predictions using each other's prediction values.

IV. RESULT AND DISCUSSION

In this section, we describe the temperature change prediction of the kneading operation and experimental results of the quality prediction of synthetic rubber. To prove the performance of the PWM presented in this study, we compared the performance of the proposed model with that of an RNN, LSTM and Seq2Seq-based ensemble model (RNN + ENS, LSTM + ENS, Seq2Seq + ENS). Next, the quality prediction results according to the defined features and cross-updating prediction results are described. The proposed methodology has been verified with the kneading operation data provided by DRB, which is a rubber production company in South Korea. 400 kneading operation records were used for learning, of which 70% were used for training, and the remaining 30% were set as test data.

A. PREDICTION ERROR (MAE) FOR TEMPERATURE PREDICTION

To generate the temperature trajectory using the PWM, the predicted values of the time, and the maximum, minimum, and end temperatures are required. For each predictive goal, a performance experiment was performed on four ML models (LR, GAM, RF, and SVM) to construct a PWM that predicts the temperature during the kneading operation, and the results are summarized in Table 1 and Figure 7. Table 1 summarizes the mean absolute errors (MAE) of predictions of the four ML models for each prediction target, and Figure 7 shows

TABLE 1. Prediction error (MAE) for temperature prediction.

ML models	Maximum temperature (°C)	Minimum temperature (°C)	End temperature (°C)	Maximum temperature-time (seconds)	Minimum temperature-time (seconds)
LR	4.47	2.79	6.9	15.15	13.93
GAM	2.81	1.69	3.96	5.76	5.63
RF	**1.45**	**1.42**	**2.09**	**1.32**	**2.06**
SVM	1.75	13.75	2.29	2.3	2.49

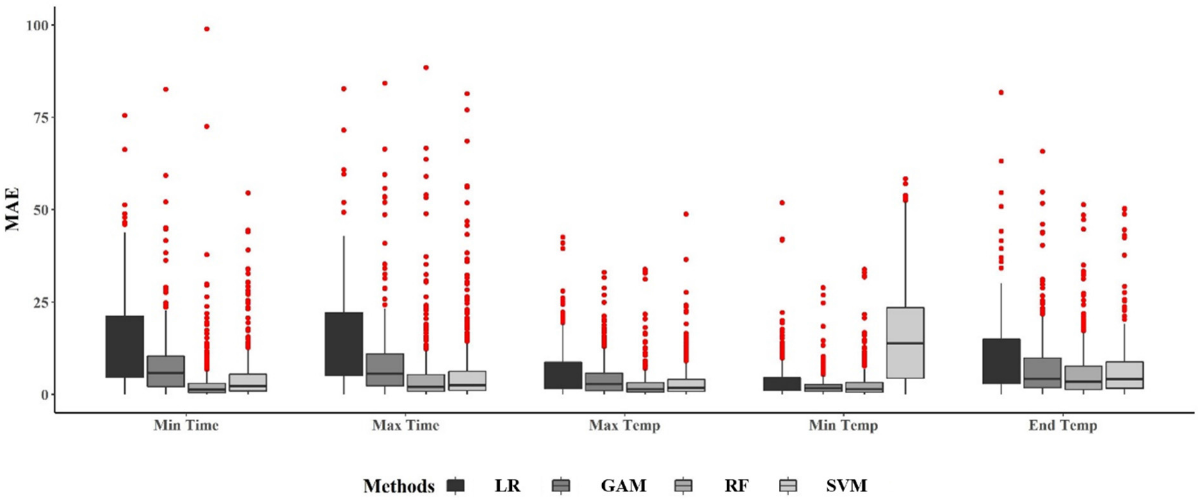


FIGURE 7. Prediction error of ML predictive model.

the distribution of the prediction errors for each model. According to Table 1, the RF shows the best performance for five prediction targets. LR showed the lowest performance among the four ML models, and SVM has a substantial error in minimum temperature prediction. The RF shows the most stable prediction results for five prediction targets. As shown in Figure 7, the dotted points are outliers, which indicate a large error, and there are cases where the predicted value of the ML model is exceptionally incorrect. If there is a large error in each prediction goal, there is a problem in generating a temperature trajectory.

For example, if the predicted maximum or minimum temperature time is too large, the correction must be performed according to Algorithm 1 because it exceeds the job time (t_{inK}) assumed by the kneading operation recipe. Therefore, the PWM should be configured using an ML model that is suitable for predicting the minimum and maximum temperature times. Table 1 shows that RF has the highest performance for all the predictive goals. This is because RF is a representative ensemble model that can overcome the overfitting problem. LR and GAM are based on linear relationships and exhibit low performance in maximum/minimum temperature time prediction with non-linear relationships. SVM can predict non-linear relationships through kernel transformation, but it is observed to overfit in the maximum temperature prediction. We configured the PWM with RF models that show the highest prediction performance.

B. TEMPERATURE TRAJECTORY GENERATION

After configuring the PWM, it receives the recipe information and generates a temperature trajectory for the synthetic rubber quality. To predict the quality, it is necessary to generate an accurate temperature trajectory. In this section, we describe the experimental results obtained using the proposed methodology. To prove the limitations of the time-series prediction model mentioned in Section II.B, and the necessity of PWM, we compared it with the LSTM methodology for temperature prediction. To overcome the problem of insufficient data, after receiving the initial temperature value suggested in a previous study, the individually trained model predicted the future temperature change and proceeded with the ensemble method to aggregate the results [18]. Table 2 describes the prediction errors of PWM's, RNN + ENS's. LSTM + ENS's and Seq2Seq + ENS's generated trajectory results. The error prediction of the PWM was smaller than that of the RNN + ENS, LSTM + ENS and Seq2Seq + ENS at all intervals. Because the temperature increases as the kneading operation progresses, the prediction error increases as time passes. According to Table 2, The Seq2Seq case shows that the prediction error is substantially larger than the other methods. RNN and LSTM also have absolute errors greater than 10% as the operation time increases. In contrast, PWM has an absolute error of less than 10% for all operation time.

In Figure 8, the gray lines represent the real temperatures during the kneading operations. The black lines are the

TABLE 2. Prediction error for generated temperature trajectory.

Operation time (seconds)	PWM (MAE)	RNN + ENSEMBLE (MAE)	LSTM + ENSEMBLE (MAE)	Seq2Seq + ENSEMBLE (MAE)
0 ~ 60	**2.07**	4.25	3.45	14.2
61 ~ 120	**4.89**	14.8	8.88	51.8
121 ~ 180	**4.50**	19.8	15.5	55.1
181 ~ 240	**6.12**	13.9	14.9	42.2
241 ~ 300	**5.53**	10.7	9.16	33.1
301 ~	**8.26**	18.1	11.6	39.8

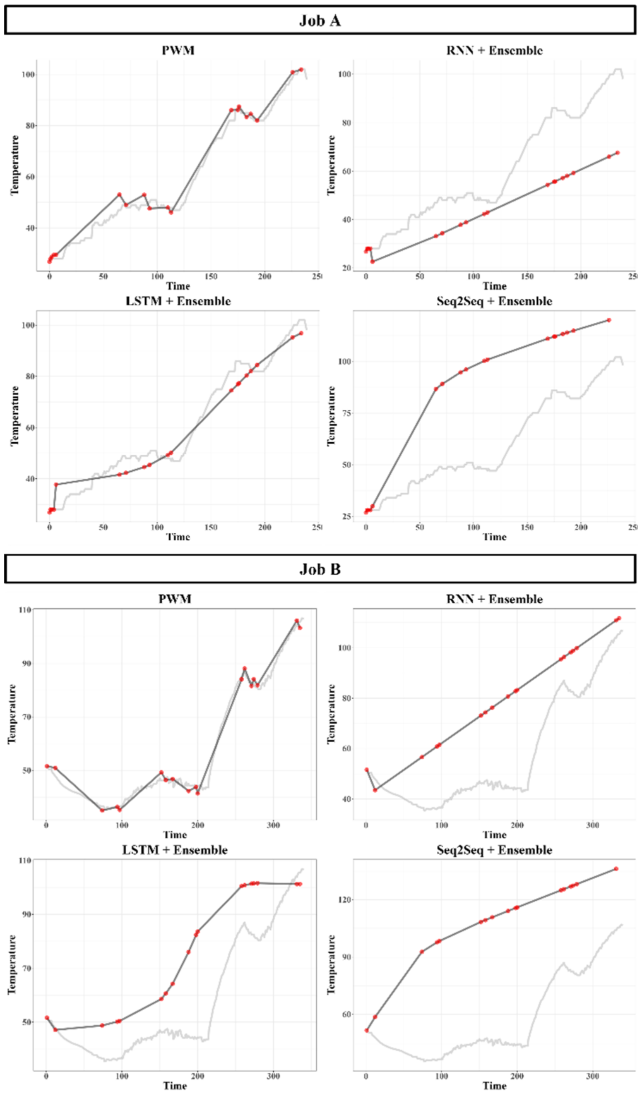


FIGURE 8. Result of generated temperature trajectory.

temperature predicted by each methodology. The red dots indicate the predicted time by PWM.

The trajectory generated by the PWM is closer to the actual trajectory compared with that generated by the RNN, LSTM and Seq2Seq + LSTM model.

The trajectory generated by the PWM follows the trend of temperature rise and fall during the kneading operation, whereas that generated by RNN, LSTM and Seq2Seq + LSTM generally shows a temperature rising trend. In the case of the LSTM+ENS model, the temperature was predicted to be higher than the actual temperature. As mentioned in Section II, the time-series prediction model has poor prediction accuracy and reliability in various situations and has a limitation in that it cannot find a global optimum, especially when the data are not sufficiently secured. The PWM predicts the temperature decrease as well as the temperature increase. Therefore, when predicting the quality of synthetic rubber, PWM-generated information may yield better results.

TABLE 3. Result of cross updating prediction.

Quality	Before updating (MAE)	After updating (MAE)
Tan-delta (δ_i)	7.18	**6.86**
Low torque (lt_i)	**0.28**	0.29
High torque (ht_i)	1.22	**1.10**

C. RESULTS OF QUALITY PREDICTION

In this section, the quality of synthetic rubber is predicted using the trajectory generated by the PWM. For this, features such as the global maximum temperature ($\max tp_{ink}$), minimum temperature ($\min tp_{ink}$), and total job time (t_{inK}) were used, which were presented in Section III.A and Figure 4. Additionally, three trajectory similarity indicators, Hausdorff distance (hd_i), dynamic time wrapping (dtw_i), and Frechet inception distance (fre_i) are used in the prediction. Cross-updating of the prediction was performed to improve the synthetic rubber quality prediction performance, and the results are summarized in Table 3.

Table 3 presents the prediction results of the three quality measures using the cross-updating method. Because the quality of the synthetic rubber is sensitive and changes depending on these three values, more accurate predictions can be made to perform a more efficient kneading operation recipe simulation. By applying the cross-updating method, we achieved improved prediction performance. In the prediction of tan-delta (δ_i) and high torque (ht_i), a reduced error was achieved by applying the cross-updating method. Low torque (lt_i) also shows a performance improvement, which is not as significant as those of δ_i and ht_i.

V. CONCLUSION

Because synthetic rubber is an essential material in various industries, rubber production companies are constantly experimenting with producing products of various qualities. However, approximately 3–5 months and equivalent costs are required to create one optimal recipe. Additionally, the working environment of synthetic rubber manufacturing has side effects that adversely affect workers' health. Therefore, a crucial economic and environmental issue is determining the recipe design method that reduces the number of experiments. In this study, a new approach was proposed to reduce the number of experiments and costs required for the production of optimal synthetic rubber recipes.

A new data-driven method that can predict the results of the kneading operation during the recipe design stage without experimentation was developed. The PWM method was proposed to predict the temperature change of the Banbury mixer and cross-update the prediction to predict the synthetic rubber quality. In the temperature change prediction, a comparison with the widely used time-series prediction model was performed to validate the performance of the proposed method. The experiments revealed that our proposed methodology outperforms the time-series prediction methodology. The PWM predicts the temperature better than the time-series prediction methodology and reflects the time point of the temperature change. In summary, as opposed to the time-series model, our proposed method trains the nonlinear relationship between features effectively. Additionally, the cross-updating prediction method was applied for synthetic rubber quality prediction, and it was proved that it performed better than the general learning method. A more accurate prediction method reduces costs and errors in the relevant field. The methodology presented in this paper is expected to be useful in various manufacturing industries where data collection and reflection are imperative. The data used in this study are limited in that they do not utilize the information on materials used in the manufacture of synthetic rubber.

It is expected that the methodology presented in this paper will exhibit better performance if the material information is used in the experiment and if a more detailed experimental schedule can be used. The methodology we present in this study is to predict the planned recipe accurately. It is helpful for the recipe generation test because predictions can be made quickly for various recipes.

The proposed methodology has the advantage of accelerating experimental data collection by performing simulations by changing parameters in various ways without consuming time and costs. Because various attempts can be tested without risk, experimental design using PWM is economical. Moreover, environmental benefits arise because workers can reduce the number of experiments. The promising results presented in this paper can be exploited to build a system that designs the optimal recipe for kneading operation. Therefore, future study will focus on quickly designing the optimal recipe using PWM.

REFERENCES

[1] A. Hasan, Rochmadi, H. Sulistyo, and S. Honggokusumo, "Rubber mixing process and its relationship with bound rubber and crosslink density," *IOP Conf. Series: Mater. Sci. Eng.*, vol. 213, Jun. 2017, Art. no. 012048, doi: 10.1088/1757-899X/213/1/012048.

[2] S. Kim, "A review of occupational cancer in the rubber industry," *Rubber Technol.*, vol. 20, no. 3, pp. 153–170,2019.

[3] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. Cambridge, MA, USA: MIT Press, 2018.

[4] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Machine learning for predictive modelling based on small data in biomedical engineering," *IFAC-PapersOnLine*, vol. 48, no. 20, pp. 469–474, 2015.

[5] F. Mateo, J. J. Carrasco, A. Sellami, M. Millán-Giraldo, M. Domínguez, and E. Soria-Olivas, "Machine learning methods to forecast temperature in buildings," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1061–1068, Mar. 2013, doi: 10.1016/j.eswa.2012.08.030.

[6] M. Zeynoddin, H. Bonakdari, I. Ebtehaj, F. Esmaeilbeiki, B. Gharabaghi, and D. Z. Haghi, "A reliable linear stochastic daily soil temperature forecast model," *Soil Tillage Res.*, vol. 189, pp. 73–87, Jun. 2019, doi: 10.1016/j.still.2018.12.023.

[7] S. Wolff, F. O'Donncha, and B. Chen, "Statistical and machine learning ensemble modelling to forecast sea surface temperature," *J. Mar. Syst.*, vol. 208, Aug. 2020, Art. no. 103347, doi: 10.1016/j.jmarsys.2020.103347.

[8] S. L. Handler, H. D. Reeves, and A. McGovern, "Development of a probabilistic subfreezing road temperature nowcast and forecast using machine learning," *Weather Forecasting*, vol. 35, no. 5, pp. 1845–1863, Oct. 2020, doi: 10.1175/WAF-D-19-0159.1.

[9] N. Lupolova, S. J. Lycett, and D. L. Gally, "A guide to machine learning for bacterial host attribution using genome sequence data," *Microbial Genomics*, vol. 5, no. 12, Dec. 2019, doi: 10.1099/mgen.0.000317.

[10] Z. Feng and Y. Zhu, "A survey on trajectory data mining: Techniques and applications," *IEEE Access*, vol. 4, pp. 2056–2067, 2016, doi: 10.1109/ACCESS.2016.2553681.

[11] S. Shang, L. Chen, Z. Wei, C. S. Jensen, K. Zheng, and P. Kalnis, "Trajectory similarity join in spatial networks," *Proc. VLDB Endowment*, vol. 10, no. 11, pp. 1178–1189, Aug. 2017, doi: 10.14778/3137628.3137630.

[12] N. Magdy, M. A. Sakr, T. Mostafa, and K. El-Bahnasy, "Review on trajectory similarity measures," in *Proc. IEEE 7th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2015, pp. 613–619, doi: 10.1109/IntelCIS.2015.7397286.

[13] J. Gross and J. Groß, *Linear Regression*, vol. 175. Springer, 2003.

[14] T. J. Hastie and J. R. Tibshirani, "Monographs on statistics and applied probability," *Generalized Additive Models*, vol. 43, pp. 205–208, 1990.

[15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[16] M. Awad and R. Khanna, "Support vector regression," *Efficient Learn. Mach.*, pp. 67–80, 2015.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[18] M. Rahaman, W. Mu, J. Odqvist, and P. Hedström, "Machine learning to predict the martensite start temperature in steels," *Metall. Mater. Trans. A*, vol. 50, no. 5, pp. 2081–2091, May 2019.

[19] Y. Tarutani, K. Hashimoto, G. Hasegawa, Y. Nakamura, T. Tamura, K. Matsudax, and M. Matsuoka, "Reducing power consumption in data center by predicting temperature distribution and air conditioner efficiency with machine learning," in *Proc. IEEE Int. Conf. Cloud Eng. (ICE)*, Apr. 2016, pp. 226–227, doi: 10.1109/IC2E.2016.39.

[20] S. Francik and S. Kurpaska, "The use of artificial neural networks for forecasting of air temperature inside a heated foil tunnel," *Sensors*, vol. 20, p. 652, Jan. 2020.

[21] J. Athavale, M. Yoda, and Y. Joshi, "Comparison of data driven modeling approaches for temperature prediction in data centers," *Int. J. Heat Mass Transf.*, vol. 135, pp. 1039–1052, Jun. 2019.

[22] A. M. Bahman and S. A. Ebrahim, "Prediction of the minimum film boiling temperature using artificial neural network," *Int. J. Heat Mass Transf.*, vol. 155, Jul. 2020, Art. no. 119834, doi: 10.1016/j.ijheatmasstransfer.2020.119834.

[23] K. M. Kim, P. Hurley, and J. P. Duarte, "High-resolution prediction of quenching behavior using machine learning based on optical fiber temperature measurement," *Int. J. Heat Mass Transf.*, vol. 184, Mar. 2022, Art. no. 122338, doi: 10.1016/j.ijheatmasstransfer.2021.122338.

[24] W. Chen, Q. Yang, S. Gao, T. Zhang, and H. Han, "Temperature monitoring and prediction under different transmission modes," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, Art. no. 107140, doi: 10.1016/j.compeleceng.2021.107140.

[25] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587, doi: 10.1016/j.ymssp.2019.106587.

[26] T. Han and Y.-F. Li, "Out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles," *Rel. Eng. Syst. Saf.*, vol. 226, Oct. 2022, Art. no. 108648, doi: 10.1016/j.ress.2022.108648.

[27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

[28] D. Liu, W. Zhang, Y. Tang, and Y. Jian, "Prediction of hydration heat of mass concrete based on the SVR model," *IEEE Access*, vol. 9, pp. 62935–62945, 2021, doi: 10.1109/ACCESS.2021.3075212.

[29] X. Zhang, Y. Wang, X. He, H. Ji, Y. Li, X. Duan, and F. Guo, "Prediction of vehicle driver's facial air temperature with SVR, ANN, and GRU," *IEEE Access*, vol. 10, pp. 20212–20222, 2022, doi: 10.1109/ACCESS.2022.3149523.

[30] A. A. Milad, I. Adwan, S. A. Majeed, Z. A. Memon, M. Bilema, H. A. Omar, M. G. M. Abdolrasol, A. Usman, and N. I. M. Yusoff, "Development of a hybrid machine learning model for asphalt pavement temperature prediction," *IEEE Access*, vol. 9, pp. 158041–158056, 2021, doi: 10.1109/ACCESS.2021.3129979.

KIKUN PARK received the B.S. degree in applied statistics from Gachon University, South Korea, in 2017. He is currently pursuing the Ph.D. degree with Pusan National University, Busan, South Korea. His research interests include machine learning and scheduling problem.

HANBYEOUL PARK received the B.S. degree in applied industrial engineering from Kyungsung University, South Korea, in 2021. He is currently pursuing the M.S. degree with Pusan National University, Busan, South Korea. His research interests include machine learning and deep learning about natural language process and graph theory.

HYERIM BAE (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in industrial engineering from Seoul National University, South Korea. From 2002 to 2003, he was at Samsung Credit Card Corporation, Seoul, South Korea. Since 2005, he has been a Professor with the Department of Industrial Engineering, Pusan National University, South Korea. His research interests include information system design, cloud computing, business process management systems, process mining, big data analytics for operational intelligence, analyzing huge volumes of event logs from port logistics, and ship building industries using process mining techniques.