IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 11, 2021, accepted September 23, 2021, date of publication September 24, 2021, date of current version October 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3115665

A Multimodal Deep Learning-Based Fault Detection Model for a Plastic Injection Molding Process

GYEONGHO KIM[1], JAE GYEONG CHOI[1], MINJOO KU[1], HYEWON CHO[1], AND SUNGHOON LIM[1,2]

[1]Department of Industrial Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea
[2]Institute for the 4th Industrial Revolution, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea

Corresponding author: Sunghoon Lim (sunghoonlim@unist.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) under Grant 2019R1F1A1059346 and Grant 2021R1F1A1046416, and in part by the AI Collaboration Project Fund of Ulsan National Institute of Science and Technology (UNIST) under Grant 1.210095.

ABSTRACT The authors of this work propose a deep learning-based fault detection model that can be implemented in the field of plastic injection molding. Compared to conventional approaches to fault detection in this domain, recent deep learning approaches prove useful for on-site problems involving complex underlying dynamics with a large number of variables. In addition, the advent of advanced sensors that generate data types in multiple modalities prompts the need for multimodal learning with deep neural networks to detect faults. This process is able to facilitate information from various modalities in an end-to-end learning fashion. The proposed deep learning-based approach opts for an early fusion scheme, in which the low-level feature representations of modalities are combined. A case study involving real-world data, obtained from a car parts company and related to a car window side molding process, validates that the proposed model outperforms late fusion methods and conventional models in solving the problem.

INDEX TERMS Machine learning, deep learning, multimodal learning, early fusion, industrial AI, plastic injection molding.

I. INTRODUCTION

Currently, in the manufacturing industry, the plastic injection molding process is an important branch of the production process for creating plastic components. A variety of plastic products are generated from this process, spanning from everyday products to subparts for automobiles and larger pieces of machinery. The plastic injection molding process consists of multiple stages, including the injection of plastic materials, heating them into a mold, and cooling them to render them into a flawless product. However, due to the inherent complexity of the process, a large number of various parameters need to be manipulated. Most are the setting parameters for the injection molding machine, while some are measured variables. For instance, these variables include temperature, pressure, and temporal information that are dependent on machine types and measurements [1]. The values for variables vary due to environmental changes during the plastic injection molding process as well as manual manipulation of machines. In addition, the plastic injection molding process requires many parameters to be set properly, with the relationships of these parameters not clearly established. Furthermore, since highly nonlinear dynamics underlie the process, it is much more challenging to find and set the proper parameters for the injection molding machine [2].

Given the complicated conditions of the injection molding process, detecting faulty products in advance of completing actual production is challenging. Nevertheless, fault detection has a benefit in terms of production efficiency. It not only helps avoid abnormal event progression [3] but also reduces time and effort spent by human experts during an on-site inspection. In order to perform high-quality fault detection, a detection model should capture meaningful patterns, which may exist in a latent space, to help classify faulty products from the others. Thus, an accurate fault detection model could be the basis for fully automated fault detection in real-time, which would enhance the overall efficiency of the plastic injection molding process.

Since their revival in the early 2010s, deep neural networks have proven to work well for many real-world and highly nonlinear problems, such as object detection, speech recognition, language modeling, and recommender systems [4]–[6]. A deep neural network's capability to model data nonlinearities or phenomena can also be well suited for existing

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

industrial problems. Another advantage of a deep neural network is its flexible applicability, in the sense that neural networks could easily adapt to specific industrial applications with some modifications to architectural details. The hyperparameters for deep neural networks (e.g., number of hidden units or layers, types of connections between layers, activations) can be adjusted in order to make a model suitable for a specific type of problem encountered. Especially in the manufacturing industry, where the number of variables may vary by machine, factory, and company, a deep neural network's adaptability is tremendously valuable.

At real industrial sites, *multimodality* is commonly present in data. The multimodal data come from multiple sources in multiple types. For instance, a single process might simultaneously yield both time-series data and tabular data. One example would be the following: during the production process for a vehicle part, the stationary values of the setting parameters for machinery, as well as the measured values for changing variables throughout the process, such as temperature or pressure, would be collected. The heterogeneity of data from the process enables a wider and deeper understanding of the underlying dynamics of the injection molding process. For instance, the addition of time-series data to tabular data increases the amount of hidden information in the data. Even though the tabular data contain much information about process initialization, time-series data append information regarding process realization and thus enable a deeper understanding of the process. Therefore, incorporating these multimodal data would help the fault detection model become more robust to noise that may exist in other data modalities. For instance, even if a minor machine malfunction occurs, affecting some parts of the input data, the remaining intact data could be utilized to make the model prediction consistent regardless of minor malfunctions.

Therefore, the authors of this work propose a supervised deep neural network-based fault detection model in the field of plastic injection molding. In addition, different types of multimodal fusions are employed and compared. Early fusion methods that combine data in feature-level representation have proven to outperform late fusion methods that conduct decision-level fusion. In particular, using a multimodal data fusion method that combines information from different data types proves to be practical, since complementary information is used for fault detection in a single production process. Furthermore, the proposed model allows for end-to-end training not only for the convenience of training but also for performance improvement.

The remainder of the paper is organized as follows. Section II discusses the literature related to this work. The selected methods for the proposed approach are illustrated in Section III. In Section IV, experiments and evaluations of the proposed approach using real-world data collected from the plastic injection molding process are conducted, and the results are analyzed in Section V. Finally, in Section VI, a conclusion is drawn, and further research topics are discussed.

II. LITERATURE REVIEW

This section illustrates previous works related to 1) fault detection and production analysis in manufacturing, including the plastic injection molding process domain; 2) time-series data analysis and its applications in the plastic injection molding domain; and 3) multimodal learning approaches using deep neural networks.

A. FAULT DETECTION IN A MANUFACTURING DOMAIN

The majority of studies for fault detection and diagnosis in the manufacturing domain have utilized various classical algorithms mostly based on statistical approaches. For instance, models that intuitively match the multistage manufacturing processes, such as the state-space model [7] and Bayesian network [8], have been used for fault detection. Especially in the injection molding domain, similar statistical approaches and numerical simulations have been utilized until the present [9]–[14]. Reyes *et al.* use an approach called case-based reasoning (CBR), which is a problem-solving method of machine learning that relies on learning and reasoning based on previous process events [15]. CBR has been frequently used for process design, control, and fault detection [16], [17]. Zhang *et al.* propose fault detection in plastic injection molding processes based on statistical quality monitoring [18]. Using multi-way principal component analysis (MPCA), statistical variables are reduced to lower dimensions to monitor the overall trajectories, which provide important information about the causes of abnormal processes. However, most conventional methods in the domain lack efficiency as well as the ability to exploit highly complex data and underlying dynamics. Moreover, as the complexity of features collected during manufacturing processes increases, and as databases store larger amounts of data in the current smart manufacturing era, the need for applying more sophisticated methods for fault detection, such as machine learning, keeps growing [19].

Due to these aforementioned conditions, machine learning-based approaches have recently started to supersede conventional approaches, such as statistical approaches and numerical simulations, owing to their superior performance in various tasks. In order to detect faulty products in manufacturing processes, manual feature extraction methods, such as signal processing and transformations, have been widely used [20]. A number of studies utilizing machine learning methods take a two-stage approach. In the first stage, raw features are projected into a latent space. In the second stage, the machine learning classifier is trained to find a decision boundary for fault detection. The idea behind this approach is to use a proper projection method to better detect faulty instances in a low-dimensional manifold (feature space) [21]. In a similar vein, Fisher's linear discriminant analysis, coupled with a cosine transform to separate faulty instances from the rest, is suggested [22]. Applying a principal component analysis (PCA)-based projection with machine learning classifiers, such as the Gaussian mixture model, has also been proposed [23]. It is worth noting that the k-nearest neighbors (k-NN) algorithm has been popularly used as a classifier after the feature projection stage [24]–[26]. Li and Zhang apply diffusion maps for feature extraction and k-NN for classification [21]. Similarly, Zhou *et al.* apply a random projection and employ k-NN for fault detection in semiconductor manufacturing processes [27]. Recently, Fan *et al.* exploit a k-means algorithm to identify important features among multiple sensors and employ naïve Bayes and k-NN in an ensemble scheme for fault detection [28]. Even though such nonparametric models (e.g., k-NN) are commonly used, significant limitations exist. First, these models tend to take longer to train when the data are large in size. In addition, they are inherently prone to overfitting to the training data and being sensitive to outliers. Especially in the plastic injection molding domain, there have been several machine learning-based fault detection approaches. Support vector machine (SVM) and multiple linear regression (MLR) are employed to predict the quality of the manufacturing processes [29], [30]. Ventura and Berjaga make a comparison of statistical discriminant analysis techniques, SVMs, and partial least squares [31]. Recently, Lee *et al.* have utilized k-NN with dynamic time warping (DTW) for fault detection in the injection molding processes [32]. For a better understanding of the predictive models' decision mechanisms, decision tree (DT)-based ensemble models, such as bagging, random forest (RF), and gradient boosting machine (GBM), have been used for fault diagnosis [33].

More recently, deep learning has been introduced in the quality monitoring process using the backpropagation neural network approach [34], [35]. In addition, a deep learning model is applied to process optimization for multivariate process planning [36], transfer learning to improve product quality [1], and minimizing the gap between real data and simulation data [2]. Deep learning approaches have also been proposed for fault detection in manufacturing processes. Shao *et al.* employ a deep belief network (DBN) for fault diagnosis [37], and Lee *et al.* use a convolutional neural network (CNN) for fault detection [38]. In the study, CNN is trained to extract fault features using multivariate sensor signals, thus locating variables that cause faulty processes as well as time information. Kim *et al.* propose a self-attentive CNN for the fault detection of lengthy sensor signals in semiconductor manufacturing processes [39]. Deep learning approaches have proven effective at detecting faulty products in manufacturing processes without bothersome procedures, including manual feature selection and extraction. Approaches for predicting defects at different levels, inducing multi-class classification, also exist [40], [41]. In a nutshell, deep learning applications for the injection molding process are used to detect sortable defects and to reduce defective products.

B. TIME-SERIES ANALYSIS IN A MANUFACTURING DOMAIN

The advent of advanced sensors, the so-called smart sensors in the injection molding industry, has enabled the generation of large volumes of time-series data that are collected in real time [42]–[44]. In the real world, an enormous amount of data exists in time-series form, in which the data are collected in a sequential manner. The ubiquity of time-series data further emphasizes the importance of time-series analysis. Time-series analysis can be also applied to real-world event prediction [45], [46]. In addition, the massive amount of time-series data collected from advanced types of sensors in the industry prompts the need for applying more sophisticated techniques, such as machine learning and deep learning.

There have been many studies on generic time-series data classification. A collective time-series classification framework called COTE (Collective of Transformation Ensembles), which is based on ensembles of nearest neighbor classifiers, is presented [47]. Afterward, HIVE-COTE (Hierarchical Vote Collective of Transformation-Based Ensembles), which improves COTE with a hierarchical voting scheme, is proposed [48]. These algorithms show powerful performance with high accuracy in addressing time-series classification problems. HIVE-COTE proves particularly useful in the medical field when predicting epileptic seizures in scalp electroencephalogram (EEG) [49]. Despite the improved performances of these algorithms, a deep learning approach is required to better understand data generated in the manufacturing industry because of the data's large number of variables and the nature of dependencies, as previously mentioned [50]. Various studies have proposed a convolutional neural network (CNN)-based method, such as fully convolutional networks and fully undecimated CNNs, for time-series classification [52].

In the manufacturing as well as injection molding domain, statistical approaches are used for process control and fault detection [53]. In particular, various transform techniques are used as a feature extraction method for multivariate time-series data collected from multiple sensors [54]–[57]. After applying an adequate feature transformation, statistical data mining techniques are used for fault detection or classification [58]. Sánchez-Fernández *et al.* propose a multivariate statistical process control for fault production by capturing anomalous products using temporal signals [59]. Zhu *et al.* propose time-series alignment kernels (TSAKs) to handle multivariate time-series sensor data from the manufacturing process [60]. The study subsequently applies SVM for fault detection. However, certain transformation techniques yield limited expressive representations from raw time-series data. Thus, several deep learning methods, which are capable of learning meaningful patterns from sequential time-series data, are applied to overcome the limitations of approaches that utilize existing time-series transform techniques. Huang *et al.* present a two-stage architecture for fault detection in motor vibration time-series data from the manufacturing process [61]. The recurrent neural network (RNN)-based variational autoencoder (VAE) is used for dimensionality reduction and feature extraction. Then, PCA and linear discriminant analysis (LDA) are applied to further improve fault detection performance. Several studies proposing deep learning methods for simultaneously extracting and detecting faulty products also exist. Chadha *et al.* use CNN and its variants to conduct fault detection with time-series data in multiple manufacturing processes [62]. Lee *et al.* utilize several kinds of deep neural networks in a one-class classification setting for imbalanced time-series data in the die-casting process [32]. CNN is often used for fault detection in multivariate time-series data from multiple sensors due to its capability to learn key features with stacked convolutions and pooling layers [63]. Compared to existing studies in other manufacturing domains, there are few pieces of research on time-series analysis in the plastic injection molding domain. The implementation of recurrent neural networks (RNNs) using long short-term memory (LSTM) units for the predictive model is presented, as they are taken repeatedly at every shot during plastic injection molding [64]. In addition, due to the characteristics of defect prediction, average recall is preferred to classification accuracy because of the skewed dataset.

C. MULTIMODAL LEARNING IN A MANUFACTURING DOMAIN

The more information derived from data collected during any process, the more accurate the understanding capability of a deep learning model becomes. The concept of multimodal learning uses this same basis. Multimodal learning incorporates related information from different investigative data sources [42]. A fusion of multimodal data allows for robust detection due to diverse representations of a single process and complementary information that exists among several modalities [65], [66]. The input data, which is composed of multiple types, such as sensor data with different temporal information and an image represented by different features, can be used to construct a multimodal learning model [67]–[69]. Especially in the manufacturing domain, multimodal data from various types of sensors and fusion methods have guaranteed complementary effects. In addition, fusing classifiers (i.e., predictions) constructed from multiple sources of sensor data has been found to greatly increase efficiency and accuracy [70]. A stochastic process, such as the hidden Markov model (HMM), has been used to address multimodal characteristics during process monitoring [71]. Some machine learning-based approaches address multimodal learning with feature extraction, which in some cases can be especially useful where domain knowledge integration is required [72].

Since the early 2010s, deep neural networks, such as deep Boltzmann machine and autoencoder, have been extensively applied in order for multimodal data to effectively learn a fusion of features [42], [73]. Particularly, in domains where the data normally exist in multiple modalities, such as an activity, context recognition [74], pose estimation [75], emotion recognition [76], [77], and medical diagnosis [43], [78], deep neural networks have proven superior to other methods. Various kinds of deep neural networks are employed because they can handle multiple modalities effortlessly. They can learn nonlinear correlations among modalities for information fusions [79] as well as complementary information among modalities [80], thus fully leveraging the potential of multimodal data. As a predictive model in the manufacturing domain, a multimodal neural network is employed to learn combined representations of semi-structured data, which are comprised of structured categorical and text-based features [81]. To diagnose faults in a gearbox, multimodal deep support vector classification (MDSVC) is proposed, where Gaussian-Bernoulli deep Boltzmann machine (GDBM) learns features in each modality, and SVM is applied to fuse three modalities [82].

This work proposes a holistic approach, based on deep neural networks, to detect faults in the field of plastic injection molding. It differentiates itself from previous studies in that it does not use artificial data from simulations to validate the proposed approach. Instead, this approach uses a large amount of real-world data from the plastic injection molding process. Rather than apply manual feature extraction methods (e.g., data projection, transformation techniques) and use an additional classifier, the proposed deep neural network-based approach is trained to extract meaningful representations and simultaneously classify faulty products. Furthermore, a deep neural network-based approach with a multimodal learning scheme is effective in solving complex real-world problems, such as problems in the plastic injection molding domain. Thorough experiments with various methods and types of multimodal fusions validate the improved performance of the proposed approach.

III. METHOD

In this work, a deep neural network-based approach that uses multimodal data to detect faulty injection molding instances is proposed. Considering the different types of data collected before and during the plastic injection molding process, several state-of-the-art deep learning methods are selected and combined to enhance the performance of the fault detection model. The proposed approach in Fig. 1 is trained in a supervised manner, where each instance of the data contains a corresponding target denoting whether it is faulty or not. In the following sections, several neural network models used in the proposed framework are described in detail.

A. MULTI-LAYER PERCEPTRON

A multi-layer perceptron (MLP) is a general feed-forward neural network that has been used in various domains. Composed of multiple fully connected neural layers (some of which might contain dropout neurons) and nonlinear activations in between the layers, MLP behaves well in modeling complex functions. The neural network is trained with backpropagation, which uses a gradient descent algorithm to update the parameters characterizing the layers. These days, other types of more advanced neural networks, such as CNNs or RNNs, provide more powerful performance in certain fields. Nonetheless, MLPs are still useful at adapting to a number of domains with high flexibility in their architecture.

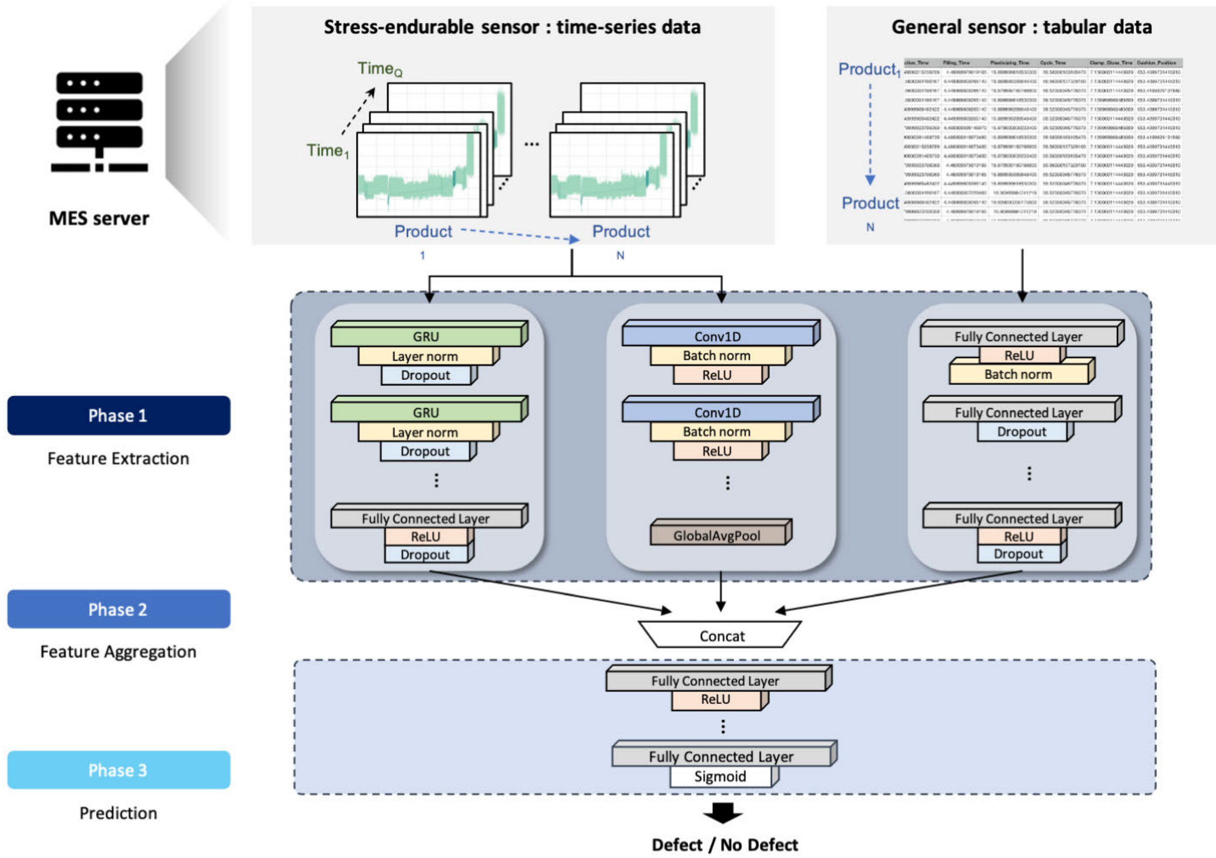

FIGURE 1. An overview of the proposed model.

(1) indicates the general computation of a single layer of MLP.

$$h_{i+1} = g(W_i h_i). \tag{1}$$

h is an i^{th} hidden representation, W is a trainable weight matrix, and g is an activation function.

B. CONVOLUTIONAL NEURAL NETWORK

CNN is one of the most widely used types of neural networks in the computer vision domain due to its capability of extracting hierarchical patterns in image and video recognition [83]. It can find not only local but also compositional relationships within input data by stacking multiple convolutional layers with nonlinear activations coupled with pooling layers. However, CNN has also has been used for pattern recognition in sequential data, such as voice and natural language [51]. Instead of using 2-D kernels or so-called filters that traverse the input data to extract meaningful patterns, the model suggested in this work utilizes a kernel with a dimension of one, a so-called 1-D convolution. The convolution computation of an input and convolutional filter k of length l is defined as:

$$f * k(i) = \sum_{j=1}^{l} k(j) f(i - j + \frac{1}{2}). \tag{2}$$

C. GATED RECURRENT UNIT

A gated recurrent unit (GRU) [84] is a compelling RNN popularly used for sequence modeling, time-series classification, language modeling, and for any kind of data expressed in sequential order. RNN is able to find patterns in sequential data, since it recurrently updates the state of its hidden layer using the input as well as the previous hidden state. (3) indicates the recurrent computation of RNN for a single time step.

$$h_t = tanh(h_{t-1}, x_t). \tag{3}$$

Analogous to LSTM networks [85], GRU addresses long-term dependencies in sequences. In fact, RNNs are prone to losing long-term dependencies because the gradients back-propagated during training either explode or vanish as the length of the input sequences increases. Both GRU and LSTM retain a unique internal mechanism (i.e., a gate) that controls the amount of information either to be forgotten or updated at each time step during the sequential computation in training. This inherent functionality thus prevents a gradient vanishing or exploding problem from happening and so is appropriate for lengthy sequences. In general, it is known that GRU 1) has a simpler architecture, 2) provides similar or slightly better performance, and 3) takes a shorter time for training than LSTM [86]. The computation of GRU at time step t is defined

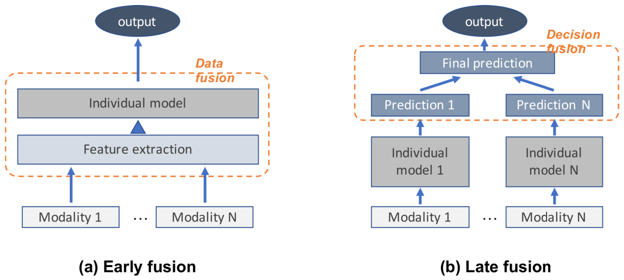

FIGURE 2. The early fusion and late fusion.

as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}). \tag{4}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}). \tag{5}$$

$$\hat{h} = tanh(W_h x_t + U_h(r_t \odot h_{t-1})). \tag{6}$$

$$h_t = z_t \odot \hat{h} + (1 - z_t) \odot h_{t-1}. \tag{7}$$

$W_z, W_r, W_h, U_z, U_r, U_h$ are the weight matrices containing corresponding bias terms (bias terms are removed for readability), $\odot$ is Hadamard product.

D. PROPOSED METHOD

The proposed approach of this work combines all previously detailed neural networks (i.e., MLP, CNN, GRU) through an early fusion manner in a multimodal learning setting. Early fusion is a fusion type where low-level feature representations of modalities are combined to yield a single prediction, whereas late fusion integrates high-level representations (i.e., decisions) after each modality has been turned into a decision via an individual prediction model [65]. The early fusion and late fusion schemes are illustrated in Fig. 2. The framework of the proposed approach is illustrated in Fig. 1 in detail. Three submodels are positioned in parallel. Each submodel individually takes one corresponding input type at a time. During a feature extraction phase, the submodels map the raw input features onto a latent space. In a feature aggregation phase, the transformed data representations are then integrated into upper layers for a single fused prediction.

Rather than training each model individually, the entire model is trained in an end-to-end fashion. The end-to-end training scheme is used, because the model is expected to automatically learn not only hidden representations of the data that help to discriminate faulty instances but also harmonious combinations of submodels that contribute to better performances. Since end-to-end training only takes an input and an output for training without any manipulations in terms of human intervention, the error signal is expected to be effectively backpropagated throughout the whole model as well as each of the submodels, increasing overall performances. In the application section, both types of early and late fusion methods are employed in order to validate the superior performance of an early fusion method. For a detailed comparison, the submodels are replaced with the others to investigate the performance of submodels in pairs.

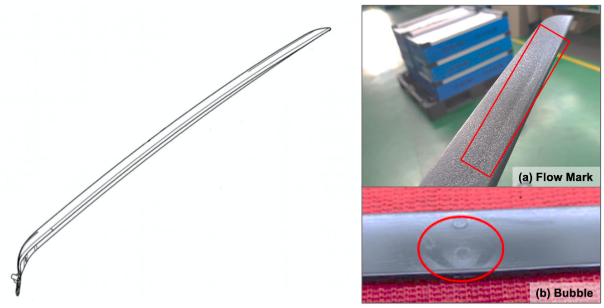

FIGURE 3. Illustrations of car windshield side moldings. Left: the detail of the car windshield side molding, Right: fault types of damaged products. (a) Flow marks are roughness of the surface caused by the high viscosity of the resin, and (b) bubbles are marks caused by unexpected gas emissions or lack of pressure.

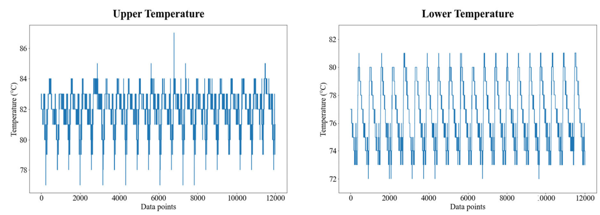

FIGURE 4. The time waveform of the time-series data.

IV. APPLICATIONS

This section introduces a case study involving real-world data collected through a production process for a car windshield side molding. A windshield side molding is the black plastic trim that wraps on the sides of a car windshield. This molding covers the gap between the edge of the windshield and the frame of the car, which helps drain water down the sides of the windshield. Fig. 3 shows the diagram of the manufactured windshield side molding (left) and the prominent defective cases (right).

The dataset used in this case study is collected by the manufacturing execution system (MES) of a car parts company in Ulsan, Republic of Korea. MES is a high-quality and profit-oriented production system that collects various information on the production site and controls the aggregation, analysis, monitoring, and production process from general and advanced sensors. Two sensors generate distinctive modal data that will be referred to below as tabular data and time-series data, respectively. General sensors collect data per instance. Meanwhile, advanced sensors, which are stress-endurable sensors that can endure higher pressures and temperatures to directly measure the resin's state, collect data at multiple timestamps per instance, with a sampling frequency of 5Hz. Fig. 4 shows the time waveform of the time-series data collected from the advanced sensors. The data used in this case study are collected over eight days from the injection molding machine, as shown in Fig. 5. The machinery is a hydraulic type that generates clamping force by directly applying fluid pressure to the mold with a hydraulic cylinder and has easy adjustment and setting of its clamping force. Each instance consists of 23 variable measurements from general sensors with a time-series of

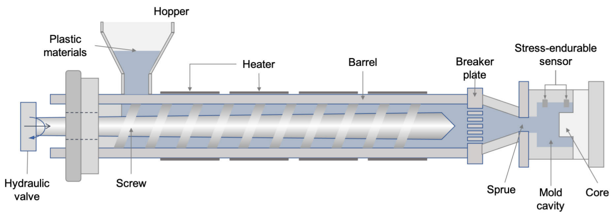

FIGURE 5. The internal structure of the molding machine.

two variables for a total of more than 2,000 instances of both data types. In this regard, the label for every instance of the data is annotated by company staff members with expertise, where defective (i.e., faulty; mostly having flow marks or bubbles on the surface) products are labeled as '1' and normal products as '0'. According to [87], the model with prior domain knowledge outperforms a domain-agnostic model, so it is postulated that instilling domain knowledge into the model is beneficial. Drawing from reflections and experiences shared by factory workers, the prediction model is devised using features that affect the presence of defects amongst all features. The selected features in tabular and time-series data are described in Tables 1 and 2, respectively.

The proposed multimodal deep learning-based fault detection model takes two types of multivariate data at a time: tabular and time-series data. Depending on the type, data are reshaped into a suitable form for the input shape of the model. Tabular data are changed into shape (N, D_1), while time-series data are changed into shape (N, D_2, T), where N denotes the number of instances in the data, D_1 and D_2 indicate the numbers of variables for tabular and time-series data, respectively, and T denotes the number of timesteps. All data are standardized for normalization. The time-series data used in this case study have varying lengths from 250 to 350 timesteps per instance. RNN is capable of dealing with the variable-length time-series, however, CNN and others are not. Thus, a zero-padding is used so that each time-series has an equivalent length. For the later experiments, the data are divided into training and test sets after random shuffling, with a ratio of 8:2.

In order to prevent overfitting in the neural networks, regularization techniques are introduced, such as dropout, batch normalization, and layer normalization. Dropout randomly disconnects neural connections with a predefined rate in the forward-pass, especially the training step [88]. Using a dropout layer has the corresponding effect of converting one network to numerous sub-networks, thus alleviating generalization errors. Batch normalization [89] is a technique that can stabilize the whole training process while accelerating the learning rate. Instability in training occurs as the variance of input values varies for each layer or activation, called an internal covariate shift. A distinctive feature of batch regularization is that the process of adjusting the mean and variance is not separated as another process but is included in a neural network to control the mean and variance during training process. In other words, it normalizes each layer,

TABLE 1. Description of variables of tabular data.

Category	Variable	Unit	Description
Time	Injection time	sec	Time for injecting plasticizing material
	Filling time	sec	Time for the molten plastic material to be injected into the mold cavity from the barrel through a sprue
	Plasticizing time	sec	Time for plasticizing material to fill into a barrel from the hopper by a screw to initialize the molding cycle
	Cycle time	sec	Time to complete the injection molding cycle
	Clamp closure time	sec	Time consumed to calibrate clamp closure pressure between each molding cycle
Position	Cushion position	mm	The position of the screw when the hold ends
	Switch over position	mm	The ram position where the injection stage switches to the packing stage
	Plasticizing position	mm	The position of the screw after plasticizing
	Clamp open position	mm	The position of clamps opened to harvest molded product from the cavity
Speed, Rate	Maximum injection speed	mm/s	The maximum forward speed of the screw during its injection operation
	Maximum screw RPM	mm/s	The maximum rate at which the plasticizing screw rotates
	Average screw RPM	mm/s	The average rate at which the plasticizing screw rotates
Pressure	Maximum injection pressure	MPa	Maximum pressure applied on the injection screw when a material is being injected into the mold
	Maximum switch over pressure	MPa	The maximum pressure to hold the material after the plasticizing process
	Maximum back pressure	MPa	The maximum resistance of the screw to recover as the metering section pumps molten plastic through the non-return valve to the front of the screw
	Average back pressure	MPa	The average of the resistance of the screw to recover as the metering section pumps molten plastic through the non-return valve to the front of the screw
Temperature	Barrel temperature	°C (six in total)	The appropriate range of the barrel temperature can be set while referring to the recommended temperature range that varies with the grade of the resin and which is provided by each resin manufacturer

TABLE 2. Description of features of time-series data.

Category	Variable	Unit	Description
Temperature	Upper temperature	°C	Upper-stream temperature from a stress-endurable sensor
	Lower temperature	°C	Down-stream temperature from a stress-endurable sensor

adjusting the distribution to avoid deformation. It can prevent gradient vanishing and exploding issues. The layer works well with CNN-based models, since it is dependent on the size of mini-batches, but it does not work with other models, such

TABLE 3. Detailed hyperparameters of the used machine learning models.

Model	Detailed hyperparameter values
SVM	C = 1.0, kernel = rbf, degree = 3
NB	None
RF	#estimators = 20, criterion = gini, max depth = 10
LightGBM	#leaves = 40, learning rate = 0.2, #estimators = 50, max depth = 20

as RNN. As an alternative, layer normalization [90] is used to overcome the drawbacks of batch normalization. It is not contingent upon batch sizes during training, as it normalizes using statistics collected from all units within a layer of the current sample, which makes it effective at stabilizing the hidden state in RNNs.

In order to counteract the problem of a class imbalance, which is often the case with data collected in the manufacturing domain, as well as the one in this work, an effective type of a loss function called focal loss [91] is used for training deep learning models. Based on the cross-entropy function, focal loss employs modulating factors and weighting factors to reduce the relative weights on well-classified instances while focusing on misclassified ones. During experiments, focal loss tends to have better performance.

The experiment is conducted in two stages. In the first stage, deep learning and conventional mathematical methods are applied to each modality. In the second stage, combinations of methods are employed with a late fusion scheme, where decisions from the submodels are averaged to compute the final decision. The proposed early fusion model with deep neural networks is tested, also with ablation. In order to validate the proposed deep learning-based approaches, several other supervised machine learning algorithms, such as support vector machine (SVM), naïve Bayes (NB), random forest (RF), and light gradient boosting machine (LightGBM) [92], along with a conventional time-series classification methods, such as dynamic time warping (DTW) and HIVE-COTE (HC) [48], are implemented for comparison. The hyperparameters for the aforementioned machine learning algorithms are tuned via random search [93], and the details of the model specifications are shown in Tables 3 and 4.

A. IMPLEMENTATION

The experiments are conducted on four GPUs: GTX 1080 Ti, implemented with open-source libraries, TensorFlow [94] and Keras [95]. Adam optimizer [96] is used with a batch size of 128. The learning rate is set to 1e-3 with a weight decay of 1e-4. After epochs without improvement, it is multiplied by $1/\sqrt[3]{2}$. To prevent models from overfitting, early stopping is used in which the training terminates as the validation loss saturates. The source code to reproduce the experiments in this work will be made publicly available.

V. RESULTS AND DISCUSSION

To validate the experiment of this case study, an F_1 score along with accuracy, precision, and recall is used. Precision is the rate of true positives among predictions as positive (i.e., faulty), while recall is the rate of true positives among the actual faulty instances. F_1 score is a harmonic mean of precision and recall, detailed as follows:

TABLE 4. Detailed architectures of the used deep learning models.

Model	Layer	Kernel size	Kernel number	Activation
MLP	1-Fully-connected	16	/	ReLU
	2-Fully-connected	32	/	ReLU
	3-Fully-connected	16	/	ReLU
	4-Fully-connected	8	/	ReLU
GRU	1-GRU	256	/	tanh
	dropout (0.3)	/	/	/
	2-GRU	128	/	tanh
	dropout (0.3)	/	/	/
	1-Fully-connected	32	/	ReLU
	dropout (0.2)	/	/	/
	2-Fully-connected	8	/	ReLU
CNN	1-Convolution	8	16	ReLU
	batchnorm	/	/	/
	2-Convolution	5	24	ReLU
	batchnorm	/	/	/
	3-Convolution	3	32	ReLU
	batchnorm	/	/	/
	1-Global-average-pooling	/	/	/
	1-Fully-connected	16	/	ReLU

$$precision = \frac{true\ positives}{true\ positives + false\ positives}. \tag{8}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}. \tag{9}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \tag{10}$$

Considering the nature of the plastic injection molding process, misclassification errors should be taken into careful consideration. Since the actual data collected during the process are mostly faultless, the data used for training, and perhaps for the test, are imbalanced or strongly skewed. F_1 score is a common metric used for imbalanced classification problems [97] and weights precision and recall equally. Therefore, the metric is particularly emphasized in the performance comparison as an appropriate measure.

The following is a comparison of conventional mathematical methods, including machine learning methods, and deep learning-based methods that use each modality. Tables 5 and 6 show the results of fault detection measured with precision, recall, accuracy, and F_1 score. The data used in corresponding experiments for Table 5 are tabular data consisting of the setting parameter values of the plastic molding injection machinery. According to Table 5, the proposed deep learning-based approach in tabular data, which is MLP, shows the highest detection performance (i.e., F_1 score = 0.6250) among other methods. In spite of similar recall values, MLP provides the

TABLE 5. Performance comparison of methods using only tabular data (setting variables).

Model	Precision	Recall	Accuracy	F_1 score
SVM	0.3846	0.5000	0.9676	0.4347
NB	0.1470	0.5000	0.9154	0.2272
RF	0.5714	0.4000	0.9776	0.4705
LightGBM	0.8000	0.4000	0.9825	0.5333
MLP	0.8333	0.5000	**0.9851**	**0.6250**

TABLE 6. Performance comparison of methods using time-series data (measurement during the process).

Model	Precision	Recall	Accuracy	F_1 score
SVM	0.3333	0.3000	0.9676	0.3157
RF	0.3636	0.4000	0.9676	0.3809
DTW	0.2727	0.3000	0.9626	0.2857
HC [48]	0.5000	0.1000	0.9651	0.1666
GRU	0.4615	0.6000	**0.9726**	**0.5217**
CNN	0.4000	0.4000	0.9702	**0.4000**

highest precision values, thus resulting in a better fault detection performance in terms of F_1 score. LightGBM also shows relatively high performance, implying the potential suitability of the gradient boosting machine models for tabular data.

Table 6 illustrates a comparison of several methods using time-series data. The data used in corresponding experiments for Table 6 are time-series data composed of sequential value measurements during the process. The results demonstrate that deep learning-based approaches have indisputably higher prediction scores than others (i.e., F_1 score $= 0.5217$ for GRU and F_1 score $= 0.4000$ for CNN). Compared to HIVE-COTE [48], an existing state-of-the-art method, GRU and CNN show better performances. In particular, given the sequencing of the data, GRU, which inherently has a recurrent nature, proves superior to CNN. Considering the performance difference between deep learning-based approaches and other approaches, the former is capable of capturing a richer hidden representation of data, hence demonstrating their suitability for fault detection using time-series sensor data.

In both experiments, deep learning-based approaches also show a slight improvement in detection accuracy, albeit with highly skewed non-faulty data. As for the results, it is postulated that using deep learning-based approaches (i.e., MLP, GRU, CNN) on both data types is effective for the fault detection model.

In subsequent experiments, a comparison of fusion-based methods for multimodal learning is conducted. In particular, late fusion approaches, where the predictions of models, each of which is trained with a single modality, are combined (mostly averaged). In addition, deep neural networks are trained in an end-to-end early fusion manner. Lastly, the proposed approach is employed.

Table 7 compares the performances of multimodal late fusion models with machine/deep learning methods on tabular and time-series data and the proposed early fusion model

TABLE 7. Performance comparison of fusion-based methods and the proposed approach.

Type	Model	Precision	Recall	Accuracy	F_1 score
[A]	SVM-SVM	0.5000	0.1000	0.9751	0.1666
	SVM-RF	0.4166	0.5000	0.9701	0.4545
	SVM-DTW	0.3076	0.4000	0.9626	0.3478
	SVM-HC	0.6000	0.3000	0.9676	0.4000
	NB-SVM	0.1470	0.5000	0.9154	0.2272
	NB-RF	0.3636	0.4000	0.9676	0.3809
	NB-DTW	0.2727	0.3000	0.9626	0.2857
	NB-HC	0.1944	0.7000	0.9203	0.3043
	RF-SVM	**1.000**	0.4000	0.9850	0.5714
	RF-RF	0.4375	0.7000	0.9701	0.5384
	RF-DTW	0.4117	0.7000	0.9676	0.5185
	RF-HC	0.8000	0.4000	0.9850	0.5333
	LightGBM-SVM	0.3571	0.5000	0.9651	0.4166
	LightGBM-RF	0.4615	0.6000	0.9726	0.5217
	LightGBM-DTW	0.3571	0.5000	0.9651	0.4166
	LightGBM-HC	0.7142	0.5000	0.9826	0.5882
[B]	SVM-GRU	0.4615	0.6000	0.9726	0.5217
	SVM-CNN	0.3636	0.4000	0.9676	0.3809
	NB-GRU	0.4615	0.6000	0.9129	0.5217
	NB-CNN	0.1842	0.7000	0.9154	0.2916
	RF-GRU	0.4666	0.7000	0.9726	0.5600
	RF-CNN	0.6666	0.6000	0.9825	0.6316
	LightGBM-GRU	0.5000	0.7000	0.9751	0.5833
	LightGBM-CNN	0.4615	0.6000	0.9726	0.5217
	MLP-SVM	0.8333	0.5000	0.9851	0.6250
	MLP-RF	0.4285	0.6000	0.9701	0.5000
	MLP-DTW	0.4000	0.6000	0.9676	0.4800
	MLP-HC	0.7500	0.6000	0.9850	0.6667
[C]	MLP-GRU	0.4666	0.7000	0.9726	0.5600
	MLP-CNN	0.5000	0.7000	0.9751	0.5833
[D]	MLP+GRU	**1.000**	0.5000	0.9876	0.6667
	MLP+CNN	0.6667	0.6000	0.9826	0.6316
[E]	MLP+GRU+CNN(**proposed**)	**1.000**	0.6000	**0.9900**	**0.7500**

with deep learning methods. In the type column, − denotes a late fusion and + denotes an early fusion. The late fusion method, which combines the conventional machine learning model with the deep learning model, gives higher values of F_1 scores than the fusion method composed only of machine learning models (i.e., Types [A] and [B]). A few late fusion methods based on conventional machine learning methods, such as RF-SVM and LightGBM-HC, provide slightly higher detection performance, which implies that conventional machine learning approaches are still effective. In addition, HC [48] seems to improve detection performances when coupled with other classifiers in a late fusion manner. Nevertheless, the majority of late fusion models containing machine learning and deep learning methods show much higher F_1 scores. According to Table 7, the average F_1

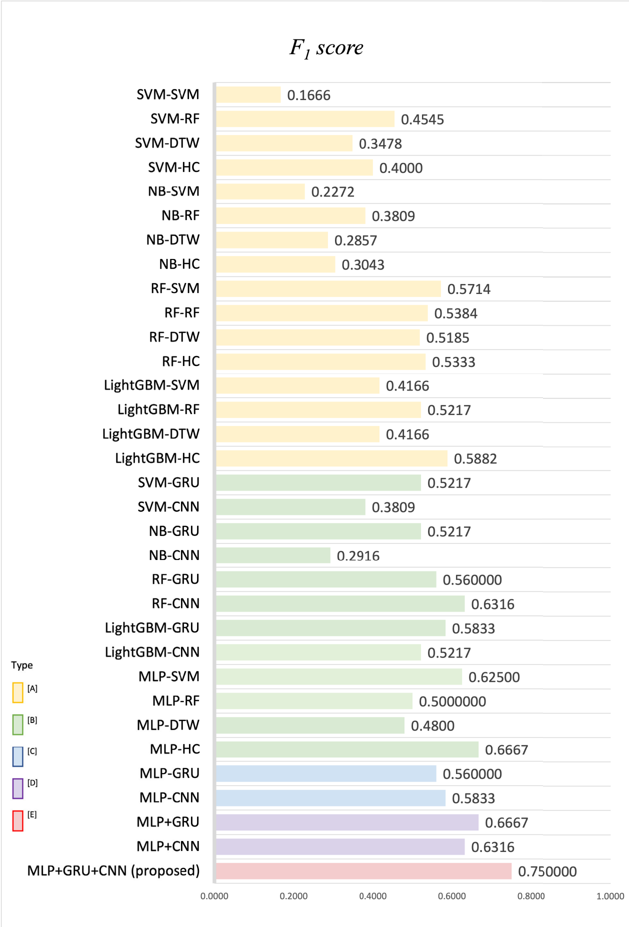

FIGURE 6. ***F_1* score comparison of fusion-based methods and the proposed approach.**

score of Type [C] is 0.5717, while that of Types [B] and [A] are 0.5107 and 0.4038, respectively. Therefore, the late fusion model with deep learning methods mostly outperforms those not based on deep learning or are on par with few existing comparative methods (e.g., HIVE-COTE).

The results indicate that different fusion types (i.e., an early fusion and a late fusion) using the same machine/deep learning methods have different performances (i.e., Types [C] and [D]). An early fusion has a conspicuous improvement on deep learning models in comparison with late fusion multimodal methods. Considering the aforementioned characteristics of the fusion types, the early fusion models might have enabled an effective combination of complementary information from multiple modalities while exploiting rich data representation through deep learning.

Table 7 also indicates that the proposed early fusion approach with three deep learning models (i.e., MLP, GRU, CNN) provides stable and remarkable fault detection performances (i.e., accuracy = 0.9900 and F_1 score = 0.7500 for Type [E]). From the results, consistent improved accuracy is visible when comparing deep learning-based models with the others and when comparing the early fusion approaches with the others. Therefore, the proposed early fusion approach proves most effective for fault detection in the field of plastic injection molding (Fig. 6).

VI. CONCLUSION AND FUTURE WORKS

The authors propose a deep learning-based approach that uses multimodal data in an early fusion manner to automatically detect faulty products during the plastic injection molding process. The proposed early fusion approach outperforms other conventional machine learning methods as well as those that adopt a late fusion. Moreover, the real-world multimodal data used in the case study are comparatively large, hence providing reliability to the results.

The proposed research has three steps. First, the multimodal data, which consist of setting parameter values and sequential measurements as tabular and time-series data, are collected from a plastic injection molding process. Deep neural network-based methods, along with conventional mathematical methods suitable for each data modality, are employed to conduct fault detection. In a multimodal setting, early and late fusion methods based on various mathematical methods, as well as the proposed early fusion approach using deep learning methods, are implemented.

Experimental results illustrate the superiority of deep learning-based models, as well as in a multimodal setting. Furthermore, the early fusion approaches have shown better performances than late fusion in fault detection. In particular, the proposed early fusion approach has shown the most remarkable performance (i.e., accuracy = 0.9900 and F_1 score = 0.7500).

The proposed approach finds its novelty in the application of an early fusion scheme as well as state-of-the-art deep neural networks in the field of plastic injection molding. The proposed approach's outstanding performance indicates its useful applicability in real-world circumstances. A few limitations of this research include that the proposed approach might require data with quite a large number of variables to achieve a decent performance in fault detection, since it exploits rich data representation using deep neural networks. In addition, depending on the size of the model, enough labeled data and computational resources would be needed.

Future works will directly exploit on-site and first-hand knowledge in the design of model architecture. In addition, introducing an explainable and interpretable model in the domain is deemed valuable. Concerning the fact that most deep neural networks are black-box models, which do not yield the reasons behind the predictions, deep learning models that are able to provide interpretable fault detection results would enhance the understanding of faulty processes as well as user satisfaction. Transfer learning [98] is also worth considering in the manufacturing industry. It enables knowledge transfer without much effort to generate new annotations as well as training models from scratch, thus making it applicable and effective in the manufacturing domain. For instance, a model trained with data collected from one machinery could be applied to another machinery. Especially for the purpose of improving the ability to generalize on a different domain (e.g., conditions of machinery), advanced deep neural network architectures that are capable of learning robust feature representations in unseen domains, such as [99], should be further studied. While the proposed early fusion approach could technically distinguish different kinds of faults (e.g., flow marks, bubbles), they are not distinguished in this case study, because MES of the car parts company in Ulsan used in this case study does not consider different kinds of faults. Future work will distinguish different kinds of faults once the data that consider different kinds of faults are available.

ACKNOWLEDGMENT

(Gyeongho Kim and Jae Gyeong Choi are co-first authors.)

REFERENCES

[1] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel, "A review of machine learning for the optimization of production processes," *Int. J. Adv. Manuf. Technol.*, vol. 104, nos. 5–8, pp. 1889–1902, Oct. 2019, doi: 10.1007/s00170-019-03988-5.

[2] H. Tercan, A. Guajardo, J. Heinisch, T. Thiele, C. Hopmann, and T. Meisen, "Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding," *Proc. CIRP*, vol. 72, pp. 185–190, Jan. 2018, doi: 10.1016/j.procir.2018.03.087.

[3] D. Miljkovic, "Fault detection methods: A literature survey," in *Proc. 34th Int. Conv. MIPRO*, 2011, pp. 750–755.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[5] J. G. Choi, C. W. Kong, G. Kim, and S. Lim, "Car crash detection using ensemble deep learning and multimodal data from dashboard cameras," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115400, doi: 10.1016/j.eswa.2021.115400.

[6] B. A. Tama and S. Lim, "Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation," *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100357.

[7] Y. Ding, D. Ceglarek, and J. Shi, "Fault diagnosis of multistage manufacturing processes by using state space approach," *J. Manuf. Sci. Eng.*, vol. 124, no. 2, pp. 313–322, 2002, doi: 10.1115/1.1445155.

[8] D. T. Nguyen, Q. B. Duong, E. Zamai, and M. K. Shahzad, "Fault diagnosis for the complex manufacturing system," *Proc. Inst. Mech. Eng., O, J. Risk Rel.*, vol. 230, no. 2, pp. 178–194, Apr. 2016, doi: 10.1177/1748006X15623089.

[9] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine learning in manufacturing: Advantages, challenges, and applications," *Prod. Manuf. Res.*, vol. 4, no. 1, pp. 23–45, Jan. 2016, doi: 10.1080/21693277.2016.1192517.

[10] C. Hopmann, S. Jeschke, T. Meisen, T. Thiele, H. Tercan, M. Liebenberg, J. Heinisch, and M. Theunissen, "Combined learning processes for injection moulding based on simulation and experimental data," *Proc. AIP Conf.*, vol. 2139, no. 1, Aug. 2019, Art. no. 030003, doi: 10.1063/1.5121656.

[11] J. Yu and Q. Zhang, "Limited rolling time domain-based hybrid tracking control for injection molding process," *IEEE Access*, vol. 7, pp. 67446–67455, 2019, doi: 10.1109/ACCESS.2019.2918020.

[12] M. Soltani, R. Kulkarni, T. Scheinost, T. Groezinger, and A. Zimmermann, "A novel approach for reliability investigation of LEDs on molded interconnect devices based on FE-analysis coupled to injection molding simulation," *IEEE Access*, vol. 7, pp. 56163–56173, 2019, doi: 10.1109/ACCESS.2019.2913786.

[13] S. Mok and C. Kwong, "Application of artificial neural network and fuzzy logic in a case-based system for initial process parameter setting of injection molding," *J. Intell. Manuf.*, vol. 13, pp. 165–176, Jun. 2002.

[14] S. Nasiri and M. R. Khosravani, "Faults and failures prediction in injection molding process," *Int. J. Adv. Manuf. Technol.*, vol. 103, nos. 5–8, pp. 2469–2484, Aug. 2019.

[15] E. R. Reyes, S. Negny, G. C. Robles, and J. M. Le Lann, "Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: Application to process engineering design," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 1–16, May 2015.

[16] M. R. Khosravani, S. Nasiri, and K. Weinberg, "Application of case-based reasoning in a fault detection system on production of drippers," *Appl. Soft Comput.*, vol. 75, pp. 227–232, Feb. 2019.

[17] M. R. Khosravani and S. Nasiri, "Injection molding manufacturing process: Review of case-based reasoning applications," *J. Intell. Manuf.*, vol. 31, no. 4, pp. 847–864, Apr. 2020.

[18] Y. Zhang, T. Mao, Z. Huang, H. Gao, and D. Li, "A statistical quality monitoring method for plastic injection molding using machine built-in sensors," *Int. J. Adv. Manuf. Technol.*, vol. 85, nos. 9–12, pp. 2483–2494, Aug. 2016.

[19] J. Wang, L. Ye, R. X. Gao, C. Li, and L. Zhang, "Digital twin for rotating machinery fault diagnosis in smart manufacturing," *Int. J. Prod. Res.*, vol. 57, no. 12, pp. 3920–3934, 2019, doi: 10.1080/00207543.2018.1552032.

[20] Z. Germán-Salló and G. Strnad, "Signal processing methods in fault detection in manufacturing systems," *Proc. Manuf.*, vol. 22, pp. 613–620, Jan. 2018, doi: 10.1016/j.promfg.2018.03.089.

[21] Y. Li and X. Zhang, "Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection," *Chemometrics Intell. Lab. Syst.*, vol. 136, pp. 47–57, Aug. 2014, doi: 10.1016/j.chemolab.2014.05.003.

[22] F. M. Megahed and J. A. Camelio, "Real-time fault detection in manufacturing environments using face recognition techniques," *J. Intell. Manuf.*, vol. 23, no. 3, pp. 393–408, 2012, doi: 10.1007/s10845-010-0378-3.

[23] J. Yu, "Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 432–444, Aug. 2011, doi: 10.1109/TSM.2011.2154850.

[24] Q. P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007, doi: 10.1109/TSM.2007.907607.

[25] G. Verdier and F. Ariane, "Adaptive Mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 59–68, Feb. 2011, doi: 10.1109/TSM.2010.2065531.

[26] C. Zhang, X. Gao, Y. Li, and L. Feng, "Fault detection strategy based on weighted distance of k nearest neighbors for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 1, pp. 75–81, Feb. 2019, doi: 10.1109/TSM.2018.2857818.

[27] Z. Zhou, C. Wen, and C. Yang, "Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 70–79, Feb. 2015, doi: 10.1109/TSM.2014.2374339.

[28] S.-K.-S. Fan, C.-Y. Hsu, D.-M. Tsai, F. He, and C.-C. Cheng, "Data-driven approach for fault detection and diagnostic in semiconductor manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 1925–1936, Oct. 2020, doi: 10.1109/TASE.2020.2983061.

[29] B. Ribeiro, "Support vector machines for quality monitoring in a plastic injection molding process," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 401–410, Aug. 2005, doi: 10.1109/TSMCC.2004.843228.

[30] M. Gülçür and B. Whiteside, "A study of micromanufacturing process fingerprints in micro-injection moulding for machine learning and industry 4.0 applications," *Int. J. Adv. Manuf. Technol.*, pp. 1943–1954, May 2021, doi: 10.1007/s00170-021-07252-7.

[31] R. Ventura and X. Berjaga, "Comparison of multivariate analysis techniques in plastic injection moulding process," in *Proc. IEEE 20th Conf. Emerg. Technol. Factory Automat. (ETFA)*, Sep. 2015, pp. 1–6, doi: 10.1109/ETFA.2015.7301557.

[32] J. Lee, Y. C. Lee, and J. T. Kim, "Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database," *J. Manuf. Syst.*, vol. 57, pp. 357–366, Oct. 2020, doi: 10.1016/j.jmsy.2020.10.013.

[33] J. Obregon, J. Hong, and J.-Y. Jung, "Rule-based explanations based on ensemble machine learning for detecting sink mark defects in the injection moulding process," *J. Manuf. Syst.*, vol. 60, pp. 392–405, Jul. 2021, doi: 10.1016/j.jmsy.2021.07.001.

[34] W.-C. Chen, P.-H. Tai, M.-W. Wang, W.-J. Deng, and C.-T. Chen, "A neural network-based approach for dynamic quality prediction in a plastic injection molding process," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 843–849, Oct. 2008, doi: 10.1016/j.eswa.2007.07.037.

[35] P. Nagorny, M. Pillet, E. Pairel, R. Le Goff, J. Loureaux, M. Wali, and P. Kiener, "Quality prediction in injection molding," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2017, pp. 141–146.

[36] B. H. M. Sadeghi, "BP-neural network predictor model for plastic injection molding process," *J. Mater. Process. Technol.*, vol. 103, no. 3, pp. 411–416, Jul. 2000, doi: 10.1016/S0924-0136(00)00498-2.

[37] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. X. Gao, "A deep learning approach for fault diagnosis of induction motors in manufacturing," *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1347–1356, 2017, doi: 10.1007/s10033-017-0189-y.

[38] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017, doi: 10.1109/TSM.2017.2676245.

[39] E. Kim, S. Cho, B. Lee, and M. Cho, "Fault detection and diagnosis using self-attentive convolutional neural networks for variable-length sensor data in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 302–309, Aug. 2019, doi: 10.1109/TSM.2019.2917521.

[40] B. Ribeiro, "Support vector machines for quality monitoring in a plastic injection molding process," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 401–410, Aug. 2005, doi: 10.1109/TSMCC.2004.843228.

[41] 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, doi: 10.3390/polym12081812.

[42] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 689–696.

[43] T. Xu, H. Zhang, X. Huang, S. Zhang, and D. N. Metaxas, "Multimodal deep learning for cervical dysplasia diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9901, Oct. 2016, pp. 115–123, doi: 10.1007/978-3-319-46723-8_14.

[44] K. Kammerer, B. Hoppenstedt, R. Pryss, S. Stökler, J. Allgaier, and M. Reichert, "Anomaly detections for manufacturing systems based on sensor data—Insights into two challenging real-world production settings," *Sensors*, vol. 19, no. 24, p. 5370, Dec. 2019, doi: 10.3390/s19245370.

[45] S. Lim and C. S. Tucker, "Mining Twitter data for causal links between tweets and real-world outcomes," *Expert Syst. Appl., X*, vol. 3, Sep. 2019, Art. no. 100007, doi: 10.1016/j.eswax.2019.100007.

[46] S. Lim, S. J. Kim, Y. Park, and N. Kwon, "A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115532.

[47] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances," *Data Mining Knowl. Discovery*, vol. 31, no. 3, pp. 606–660, May 2017, doi: 10.1007/s10618-016-0483-9.

[48] J. Lines, S. Taylor, and A. Bagnall, "HIVE-COTE: The hierarchical vote collective of transformation-based ensembles for time series classification," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1041–1046, doi: 10.1109/ICDM.2016.0133.

[49] P. Peng, H. Wei, L. Xie, and Y. Song, "Epileptic seizure prediction in scalp EEG using an improved HIVE-COTE model," in *Proc. 39th Chin. Control Conf. (CCC)*, Jul. 2020, pp. 6450–6457, doi: 10.23919/CCC50068.2020.9188930.

[50] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019, doi: 10.1007/s10618-019-00619-1.

[51] S. A. Ebrahim, J. Poshtan, S. M. Jamali, and N. A. Ebrahim, "Quantitative and qualitative analysis of time-series classification using deep learning," *IEEE Access*, vol. 8, pp. 90202–90215, 2020, doi: 10.1109/ACCESS.2020.2993538.

[52] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017, doi: 10.21629/JSEE.2017.01.18.

[53] A. Ison and C. Spanos, "Robust fault detection and fault classification of semiconductor manufacturing equipment," in *Proc. Int. Symp. Semiconductor Manuf.*, vol. 2, 1996, pp. 1–6.

[54] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005, doi: 10.1007/s10115-004-0154-9.

[55] F. Morchen, *Time Series Feature Extraction for Data Mining Using DWT and DFT*. Marburg, Germany: Philipps-Univ. Marburg, 2003.

[56] E. K. Lada, J.-C. Lu, and J. R. Wilson, "A wavelet-based procedure for process fault detection," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 1, pp. 79–90, Feb. 2002, doi: 10.1109/66.983447.

[57] J. M. Ko and C. O. Kim, "A multivariate parameter trace analysis for online fault detection in a semiconductor etch tool," *Int. J. Prod. Res.*, vol. 50, no. 23, pp. 6639–6654, Dec. 2012, doi: 10.1080/00207543.2011.611538.

[58] T. Lee and C. O. Kim, "Statistical comparison of fault detection models for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 80–91, Feb. 2015, doi: 10.1109/TSM.2014.2378796.

[59] A. Sánchez-Fernández, F. J. Baldán, G. I. Sainz-Palmero, J. M. Benítez, and M. J. Fuente, "Fault detection based on time series modeling and multivariate statistical process control," *Chemometric Intell. Lab. Syst.*, vol. 182, pp. 57–69, Nov. 2018, doi: 10.1016/j.chemolab.2018.08.003.

[60] F. Zhu, X. Jia, M. Miller, X. Li, F. Li, Y. Wang, and J. Lee, "Methodology for important sensor screening for fault detection and classification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 1, pp. 65–73, Feb. 2021, doi: 10.1109/TSM.2020.3037085.

[61] Y. Huang, C.-H. Chen, and C.-J. Huang, "Motor fault detection and feature extraction using RNN-based variational autoencoder," *IEEE Access*, vol. 7, pp. 139086–139096, 2019, doi: 10.1109/ACCESS.2019.2940769.

[62] G. Singh Chadha, M. Krishnamoorthy, and A. Schwung, "Time series based fault detection in industrial processes using convolutional neural networks," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2019, pp. 173–178, doi: 10.1109/IECON.2019.8926924.

[63] C.-Y. Hsu and W.-C. Liu, "Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing," *J. Intell. Manuf.*, vol. 32, no. 3, pp. 823–836, Mar. 2021, doi: 10.1007/s10845-020-01591-0.

[64] S. Kim, S. Kim, and K. R. Ryu, "Deep learning experiments with skewed data for defect prediction in plastic injection molding," in *Proc. IEEE/ACS 15th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2018, pp. 1–2, doi: 10.1109/AICCSA.2018.8612875.

[65] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[66] W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: A survey," *IEEE Access*, vol. 7, pp. 63373–63394, 2019, doi: 10.1109/ACCESS.2019.2916887.

[67] G. Tse, J. Zhou, S. Woo, C. Ko, R. Lai, T. Liu, and Y. Liu, "Multi-modality machine learning approach for risk stratification in heart failure with left ventricular ejection fraction 45%," *ESC Heart Failure*, vol. 7, no. 6, pp. 3716–3725, 2020.

[68] K. Usman and K. Rajpoot, "Brain tumor classification from multi-modality MRI using wavelets and machine learning," *Pattern Anal. Appl.*, vol. 20, no. 3, pp. 871–881, 2017.

[69] G. van Tulder and M. de Bruijne, "Learning cross-modality representations from multi-modal images," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 638–648, Feb. 2019.

[70] O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela, "Human activity recognition based on a sensor weighting hierarchical classifier," *Soft Comput.*, vol. 17, no. 2, pp. 333–343, Feb. 2013, doi: 10.1007/s00500-012-0896-3.

[71] J. Yu, "Hidden Markov models combining local and global information for nonlinear and multimodal process monitoring," *J. Process Control*, vol. 20, pp. 344–359, Mar. 2010, doi: 10.1016/j.jprocont.2009.12.002.

[72] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple modalities: Face, body gesture, speech," in *Affect and Emotion in Human-Computer Interaction*. Berlin, Germany: Springer, 2008, pp. 92–103.

[73] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 2222–2230.

[74] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–27, Jan. 2018, doi: 10.1145/3161174.

[75] W. Ouyang, X. Chu, and X. Wang, "Multi-source deep learning for human pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2329–2336.

[76] W. Liu, W. L. Zheng, and B. L. Lu, "Emotion recognition using multimodal deep learning," in *Proc. Int. Conf. Neural Inf. Process.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9948, Oct. 2016, pp. 521–529, doi: 10.1007/978-3-319-46672-9-58.

[77] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, "Cross-subject multimodal emotion recognition based on hybrid fusion," *IEEE Access*, vol. 8, pp. 168865–168878, Sep. 2020, doi: 10.1109/access.2020.3023871.

[78] O. Toygar, F. O. Babalola, and Y. Bitirim, "FYO: A novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, vol. 8, pp. 82461–82470, 2020, doi: 10.1109/ACCESS.2020.2991475.

[79] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017, doi: 10.1109/MSP.2017.2738401.

[80] K. Liu, Y. Li, N. Xu, and P. Natarajan, "Learn to combine modalities in multimodal deep learning," 2018, *arXiv:1805.11730*. [Online]. Available: http://arxiv.org/abs/1805.11730

[81] D. Lichtenwalter, P. Burggräf, J. Wagner, and T. Weißer, "Deep multimodal learning for manufacturing problem solving," *Proc. CIRP*, vol. 99, pp. 615–620, Jan. 2021, doi: 10.1016/j.procir.2021.03.083.

[82] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, Nov. 2015, doi: 10.1016/j.neucom.2015.06.008.

[83] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2014, pp. 818–833.

[84] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[85] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[86] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: http://arxiv.org/abs/1412.3555

[87] N. Muralidhar, M. R. Islam, M. Marwah, A. Karpatne, and N. Ramakrishnan, "Incorporating prior domain knowledge into deep neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 36–45, doi: 10.1109/BigData.2018.8621955.

[88] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[89] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[90] J. Lei Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*. [Online]. Available: http://arxiv.org/abs/1607.06450

[91] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[92] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 3147–3155, doi: 10.1109/iccse.2019.8845529.

[93] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012, doi: 10.5555/2188385.2188395.

[94] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[95] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras

[96] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[97] H. He and Y. Ma, *Imbalanced Learning: Foundations, Algorithms, and Applications*. Hoboken, NJ, USA: Wiley, 2013.

[98] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[99] T. Han, Y.-F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2021.3088489.

GYEONGHO KIM received the B.S. degree in industrial engineering from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2021, where he is currently pursuing the combined M.S./Ph.D. degree in industrial engineering. His research interests include semi-supervised learning, anomaly detection, application of machine learning in industries (industrial AI), and knowledge distillation.

JAE GYEONG CHOI received the B.S. degree in design and human engineering from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2019, where she is currently pursuing the combined M.S./Ph.D. degree in industrial engineering. Her research interests include machine learning/deep learning, industrial artificial intelligence, and video/audio processing.

MINJOO KU received the B.S. degree in technology management from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2016, where she is currently pursuing the master's degree in industrial engineering. From 2017 to 2019, she was a Flight Attendant at Kuwait Airways. Her research interests include machine learning/deep learning and industrial artificial intelligence.

HYEWON CHO received the B.S. degree in biomedical science/chemical engineering and the M.S. in biomedical engineering from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2016 and 2018, respectively, where she is currently pursuing the Ph.D. degree in industrial engineering. From 2018 to 2019, she was a Research Analyst at the Consulting Team, IQVIA, providing consulting service based on healthcare data analytics on primary and secondary research basis, from pharmaceutical landscape to recent biotech application. Her research interests include industrial artificial intelligence and machine learning/deep learning application into biotech and/or healthcare services.

SUNGHOON LIM received the B.S. and M.S. degrees in industrial engineering from KAIST, Republic of Korea, in 2005 and 2009, respectively, and the Ph.D. degree in industrial engineering from The Pennsylvania State University, University Park, PA, USA, in 2018. Since 2018, he has been an Assistant Professor with the Department of Industrial Engineering, Ulsan National Institute of Science and Technology (UNIST), Republic of Korea. Since 2021, he has also been the Head of the Institute for the 4th Industrial Revolution, UNIST. His research interests include machine learning/deep learning, industrial artificial intelligence, and smart manufacturing.