

Received 11 September 2022, accepted 21 September 2022, date of publication 28 September 2022, date of current version 7 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3210525

RESEARCH ARTICLE

Dynamic Modeling With Integrated Concept Drift Detection for Predicting Real-Time Energy Consumption of Industrial Machines

ABDULGANI KAHRAMAN¹, MEHMED KANTARDZIC¹, AND MUHAMMED KOTAN²

¹Department of Computer Science and Engineering, University of Louisville, Louisville, KY 40292, USA

²Department of Information Systems Engineering, Sakarya University, 54050 Sakarya, Turkey

Corresponding author: Abdulgani Kahraman (a0kahr01@louisville.edu)

ABSTRACT Industrial machinery is a significant energy consumer, and its CO_2 emissions have increased dramatically in recent years. Therefore, energy efficiency is becoming crucial for businesses, governments, as well as the planet. Estimating the power consumption of industrial machines with greater accuracy assists management and optimizes machine operation parameters. Real-time industrial machine datasets present several challenges, such as changes in the data over time, unknown running conditions, missing data, etc. Most research publications focus on the accuracy of traditional static models of forecasting; however, prediction performance deteriorates over time because data evolves. We implemented deep learning as a prediction model for three distinct real-world industrial datasets. The proposed method, dynamic modeling with memory (DMWM), improved overall prediction performance compared with conventional approaches by identifying concept drifts and optimizing the number of required models in response to industrial datasets' recurring machine energy consumption patterns.

INDEX TERMS Concept drift, deep learning, energy efficiency, energy consumption prediction, industrial machines.

I. INTRODUCTION

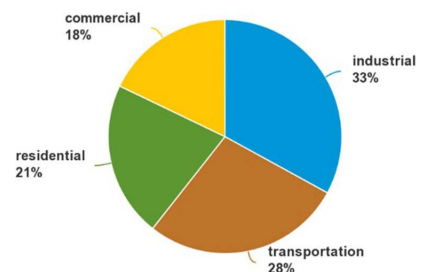
Since the industrial revolution, the burning of fossil fuels such as coal, oil, and gas has caused long-term changes to the climate, including temperature and weather patterns [1]. The European Commission has set the goal of making Europe “climate-neutral” by 2050 as part of its European Green Deal, and achieving this target will require a significant reduction in greenhouse gas emissions [2]. In the U.S., the industrial sector is the most significant consumer of power, accounting for 33% of national energy consumption in 2021, as shown in Figure 1 [3]. The industry also accounts for roughly 30% of emissions, making it one of the substantial sources of greenhouse gases [3]. Decreasing these emissions is crucial, as their influences are irreversible from an environmental perspective.

Due to recent difficulties such as the global coronavirus pandemic, companies, governments, and the world have

The associate editor coordinating the review of this manuscript and approving it for publication was Adamu Murtala Zungeru¹.

Share of total U.S. energy consumption by end-use sectors, 2021

Total = 97.33 quadrillion British thermal units



Source: U.S. Energy Information Administration, *Monthly Energy Review*, Table 2.1, April 2022, preliminary data

Note: Sum of individual percentages may not equal 100 because of independent rounding.

FIGURE 1. U.S. Energy usage percentages by sectors in 2021 [3].

faced a bottleneck in energy supplies. In addition, more than a third of manufacturing enterprises do not set energy efficiency targets, and most do not have a system to track progress [4]. The combination of demand uncertainty, energy

supply difficulties, and high production costs can significantly impact the development of pricing schemes and investment strategies for most industries [5], [6].

There are some technological solutions. Smart grids can help with electricity management by supplying valuable data that can be used to make better decisions [7]. More profound insights into energy consumption patterns are also opening the path to new knowledge and innovation in hourly load forecasting [8]. A power grid needs accurate short-term electrical load forecasts to operate more efficiently [9]. Precise short-term energy prediction is therefore critical to increasing energy efficiency and minimizing blackouts [10].

Moreover, short-term projections are helpful for planning and allocating energy loads since they predict energy usage for any interval from one hour or less to one week [10]. An accurate short-term load forecasting model can improve the reliability of the energy market by increasing efficiency, decreasing production costs, and preventing overproduction and underproduction. This, in turn, is useful for the control of manufacturing and the supply chain [11]. Scheduled operators can be informed of accurate real-time consumption and storage levels to optimize energy use in the industry [12].

It is difficult to explicitly observe or detect any specific patterns or sudden shifts in machine energy consumption. Uncertainties in streamed datasets may include variances in the data known as concept drift [13] or model degradation [14]. While the allocation of the input variable may or may not vary, concept drift is defined as a change in the conditional distribution of the target variable [15]. Active techniques are more concerned with identifying concept drift up front and updating the model afterward [16]. By contrast, passive models use learning algorithms that are updated under the presumption that concept drift is present in continuously changing data [16], [17]. However, passive techniques demand significant processing work and can only adapt to current data with constant updates, which increases the computational load [17].

Energy prediction models based on physics have been presented in the past to comprehend the energy use pattern of a machine tool; nevertheless, uncertainties about both the machine and its operating environment make it difficult to predict energy consumption with certainty [18]. Furthermore, many researchers use synthetic datasets, in which all concept drift points are known in advance. However, working with real-world industrial datasets is far more challenging as they typically include many missing values or unrecorded features, as well as unknown concept drifts and machine running conditions [13], [19].

As a result, conventional machine-learning algorithms struggle to handle changes in real-world streaming data [15]. Traditional electricity energy forecasting models are often trained once and then not re-trained with new data; thus, the prediction performance of the original model deteriorates since industrial data generally changes over time [20]. Most researchers traditionally split all available data into certain parts for training and testing [12], [21], [22]. However, if there

is any degradation in the model prediction performance for future data, it is not easy to accurately detect possible concept drift points.

More research on detecting such drifts is essential because few papers deal with regression settings with a concept drift solution for real-world industrial datasets [14]. This paper enhances the detection of change points for the real-time energy consumption of industrial machines and improves prediction performance by detecting concept drifts thanks to a specially designed dynamic modeling for industrial machines. The model solution developed here can be implemented for similar regression problems with an evolving data issue when a system has a complicated structure that includes several unrecorded features. Moreover, instead of using common data-split approaches (standard fixed-size chunks, etc.), the data is divided into dynamic portions based on machine inactive conditions to detect possible changes more accurately, known as the data-driven method. The proposed model, named dynamic modeling with memory (DMWM), actively reveals possible concept drift points and re-trains the model based on the latest data samples after controlling memorized models. Furthermore, DMWM detects repetitive energy consumption regimes and utilizes old models to optimize the total number of models with less computational cost. As a result, the proposed method outperforms the traditional approach and is designed to optimize the total number of developed models with increased accuracy.

To test the proposed method, three different real-world industrial machines' energy consumption datasets are utilized, and the results are evaluated. The rest of this paper is organized as follows. Section 2 reviews works relevant to this study. The methodology of analysis and dataset details are shared in the third section. Section 4 covers the results and discussion. Lastly, the conclusion and future recommendations are given in Section 5.

II. RELATED WORK

Many researchers have studied ways to predict energy consumption using machine learning techniques. However, there is minimal research into existing concept drift methods used for real-world streamed industrial datasets [15]. Most of the proposed methods prove that concept drift detection abides by the "no-free-lunch" theory, which is a comprehensive approach expressing it can be difficult to discover a method that works for all machine learning problems [14]. To the best of the authors' knowledge, few research papers investigate real-time industrial energy consumption datasets with a possible concept drift solution.

Energy is generally considered to be one of the essential factors for a country's economic development. Nowadays, most countries have to deal with rising energy demand because of their growing populations, and industrial needs [21]. Demand response significantly impacts the power system, easing the balance between production and usage; furthermore, smart grids offer access to vast amounts of data, significantly impacting power system monitoring [23], [24].

Uncertainties in demand, generation, and costs can affect an industry's development of pricing schemes and investment strategies [25], [26]. Most companies have started reducing their energy expenses to ensure a better profit margin and reduce emissions to minimize their environmental impacts. Furthermore, load forecasting enables better energy management, and thus lower costs [5], [13].

Several methods have been developed to improve these forecasts, some more successful than others. Traditional regressive approaches for multi-step time-series data, such as autoregressive integrated moving average (ARIMA), fail to capture non-linear patterns and features [27]. An alternative approach is to combine predicted hourly energy consumption data to create a simulation that can highlight correlations between manufacturing schedules and energy consumption, especially for heating, ventilation, and air conditioning (HVAC) system demands [4].

For example, Zhang *et al.* [28] constructed a model that combines transformer and k-means techniques for every 23 hours of training data divided into clusters. At the same time, the transformer model is trained to predict the next hour's power usage, with the predicted value being added to the trained k-means cluster and the cluster's centroid acting as the final predicted value. Ramos *et al.* [5] further used an artificial neural network (ANN) to re-train the model before every forecasting day, resulting in an updated forecasting model based on 16 months of data split into five-minute increments from an industrial plant.

On a smaller scale, Bhinge *et al.* [18] have shown how to anticipate the energy consumption of a machine tool using a non-parametric regression model. According to their results, this Gaussian process model can explain the complex interactions between input machining parameters and output energy consumption. On a larger scale, Li *et al.* [29] studied energy consumption forecasts in the oil and gas industry with a hybrid model artificial neural network and extreme learning machine.

Rahman *et al.* [21] tried to solve the issue of data filtering using non-predictive factors and feature ratings for a real-time steel industrial dataset, with support vector machines (SVM) providing the most accurate prediction results. Another similar paper shows that the random forest (RF) model outperforms other regression models in predicting steel industrial energy usage [22].

The best results for each machine learning model may vary since differences in the data split ratio, tuning of hyperparameters, feature selection, etc., can affect the results. To estimate the energy consumption of grinding and milling machines with numerous feature extraction techniques, a novel data-driven energy prediction strategy with deep learning was employed in [30] to remove extraneous features. Another equivalent study was conducted to forecast the energy consumption of electric arc furnaces, with the results showing that deep neural networks (DNN) beat decision trees (DT), linear regression (LR), and SVM in terms of forecast accuracy [31].

Other forecast studies focus on practical applications. To estimate the cutting energy of machining, higher feed rates, and greater spindle speeds require less energy according to used ANN inputs [32]. Real-time operational data on variable feed tonnage, bearing pressure, and spindle speed from semi-autogenous grinding (SAG) mills was employed by Avalos *et al.* [33]. They used several deep learning and machine learning approaches to forecast the SAG mills' energy needs. Their findings demonstrated that one of the most outstanding prediction performances for SAG mill energy consumption was achieved by neural networks.

Bermeo-Ayerbe *et al.* [34] used industrial testing data with three artificially generated concept drifts. Compared with the non-adaptive model, their proposed adaptive strategy outperformed the conventional approaches in terms of energy prediction performance. Mariano-Hernández *et al.* [20] utilized active and passive concept drift detection approaches in the DT and deep learning models. The results indicate that constant re-training of the decision tree models, together with change detection methods, can improve their ability to adapt to changes in the total electrical consumption of a building. Meanwhile, Jayaratne *et al.* [35] have proposed an unsupervised machine learning algorithm for continuous concept drift detection in a synthetic dataset; the proposed model distinguishes between abrupt and reoccurring drifts.

Due to the time-varying data distribution in the concept-drift context, it is impossible to pre-set the variance into the typical surrogate gradient (SG) approach [36]. In order to address the exponential degradation of long-term memory in LSTM, Zheng *et al.* [36] developed a novel adaptive and hybrid spiking (AHS) module that worked in conjunction with two attention mechanisms adjusting the attention score using the negative log-likelihood function to reduce the effects of concept drift. Experiment findings demonstrate that the suggested technique outperformed the latest models in the literature.

You *et al.* [37] created a unique learning technique to simulate concept drift during inference, which can aid the model's future generalization. Additionally, they suggest enhancing the framework by utilizing related series in concept drift modeling to reduce the effects of disturbances and randomness in time series data. Extensive tests were done on three real-world datasets confirming the suggested approach's efficiency. For instance, compared to cutting-edge approaches, the proposed approach delivered a relative improvement of 33% in stock price prediction.

In this research, instead of finding the best prediction model, we mainly focus on enhancing overall prediction performance by detecting concept drift points in order to more accurately predict real-world short-term industrial energy consumption over time.

III. METHODOLOGY

Accurately predicting the energy consumption of industrial machines is a complex process even for experts in the field [38]. The precision of a forecast degrades because any

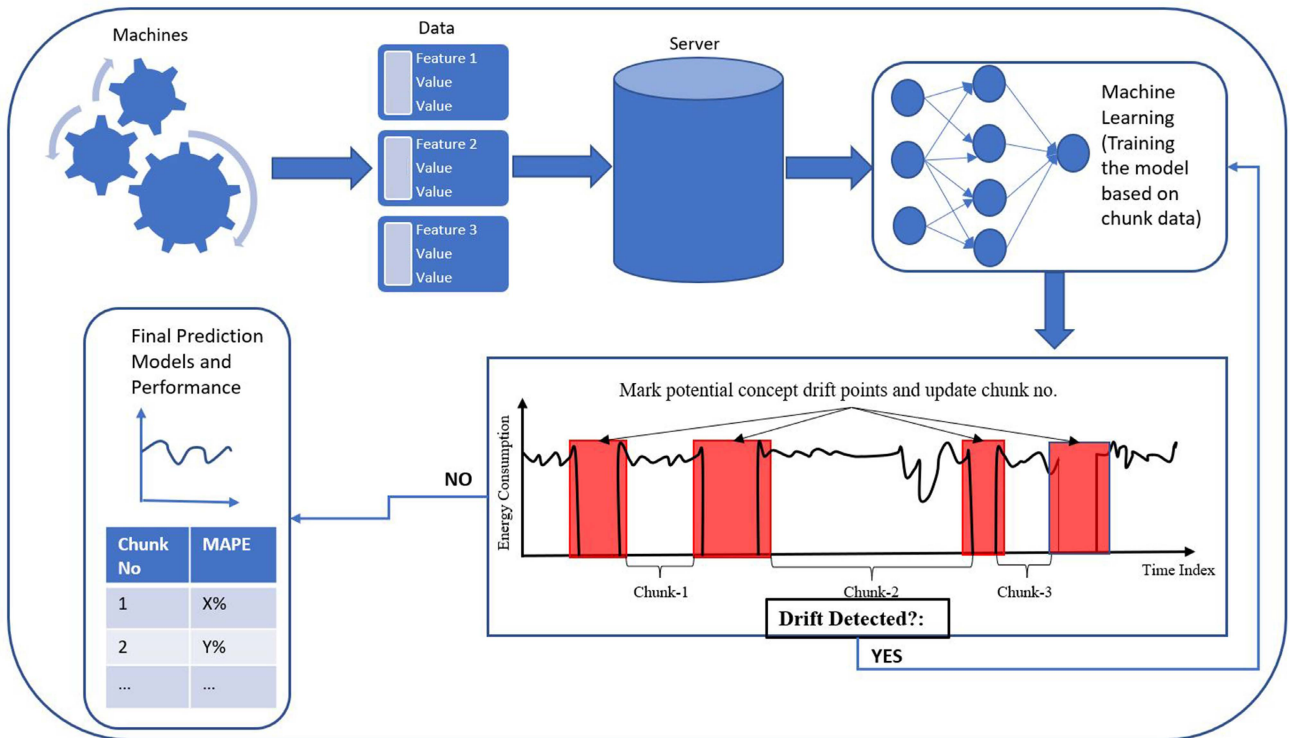


FIGURE 2. Designed framework for energy consumption analysis of industrial datasets.

system experiences unforeseen changes in machine running conditions. Therefore, an adaptive method is necessary to update the model [34]. According to the literature, concept drift can be detected using active or passive methods, as described in Section 2 [14], [15], [35], [39]. Active methods are primed to detect changes and are re-trained when a trigger is recognized, while passive methods are re-trained at regular intervals regardless of whether a change has occurred [20]. Equations 1 and 2 identify the changes in the distributions.

As an equation:

$$P_t(y|X) \neq P_{t+i}(y|X) \tag{1}$$

$$\frac{P_t(X|y) \times P_t(y)}{P_t(X)} \neq \frac{P_{t+i}(X|y) \times P_{t+i}(y)}{P_{t+i}(X)} \tag{2}$$

Concept drift is a change in the joint probability, $P_t(y|X)$ indicates the posterior probability distribution of the target labels, and $P_t(X|y)$ indicates the class-conditional probability density distribution. While $P_t(X)$ is the probability distribution of the input data, $P_t(y)$ indicates the prior probability distribution of the target labels evolving, which is represented with (t) and $(t + i)$ [14].

Furthermore, most industrial datasets usually have an excess of missing values and an absence of essential attributes. Operator mistakes such as a lack of maintenance, aggressive running of machines, poor quality of materials, excessive speed, and high pressure can all have a significant impact on machines' power usage. However, most real-world industrial streaming data do not include all these changes in machine operations [38], [40]. In order to make a more

accurate prediction of energy usage, the issue of uncertainty or unknown dynamics in streaming data must first be addressed [8], [41]. As a result, it is necessary to develop a method that can consider and ideally solve all those problems.

Most industrial machines work 24 hours a day, seven days a week, creating massive amounts of data under different running conditions. Modeling must also consider potential concept drift cases over time since industrial machines generally do not have stable working conditions. Additionally, it is important to check whether a different type of regime is repeated or not. We developed a novel approach that can solve all those problematic issues by using dynamic modeling for real-time industrial energy consumption prediction.

The method developed for dynamic energy consumption prediction with concept drift detection is illustrated in Figure 2. Each coming data portion is numbered according to machine inactive points. The minimum duration of the inactivity threshold value can be decided by a user based on similar research papers or experimental results. Those points identified as representing a potential concept drift and upcoming data samples will be given a different chunk number. Each inactive point is considered a potential change point when the machine is idle for more than a decided number of consecutive hours. During these unused periods, the machine may have significant maintenance or material changes that will impact its future power usage. If the prediction error rate for upcoming samples increases dramatically, the current chunk will be considered a concept drift zone. A new model will be trained based on the current data portion to adapt to the machine's new running conditions.

TABLE 1. Datasets' features with units.

SAG Mill		Steel Industrial		Ventilation Machine	
Feature Names	Units	Feature Names	Units	Feature Names	Units
Feed ParticularSize	(cm)	Lagging Current Reactive Power	(kVarh)	Humidity	(%)
Feed ParticularSize 2	(cm)	Leading Current Reactive Power	(kVarh)	Density	(%)
Density	(%)	tCO2	(Ppm)	Internal Load Percentage	(%)
Fresh Feed Amount	(TPH)	Lagging Current Power Factor	(%)	Temperature1	C
Sound	(dB)	Leading Current Power Factor	(%)	Temperature2	C
Pressure	(psi)	Number of Seconds from Midnight	(sec)	Load	BTU
Energy Consumption	(kWh)	Energy Consumption	(kWh)	Speed	(mph)
				Pressure	(psi)
				Energy Consumption	(kWh)

Since no detailed records are available about the machine's changing status, such as variations in running conditions, operator mistakes, material changes, etc., this proposed dynamic approach overcomes such unknown situations using all available data. It resolves problems by re-training the prediction model and adapting to new or developing operating conditions. Furthermore, the suggested data-driven method keeps similar running condition samples in the same chunk instead of using fixed-size sampling. Thanks to this, each chunk has a different sample size, but more relevant data samples are in the same portion.

Deep learning techniques have been gaining prominence due to their ability to learn feature representations, excellent generalization abilities, and capacity to model the sort of intricate relationships frequently found in massive datasets [8]. According to the [42], multi-layer perceptron (MLP) networks are superior for predicting energy consumption compared with standard regression models. However, according to the no-free-lunch theory, there are no specific models for certain problems [13].

If the dataset has an unused period of at least 12 hours or more marked as a potential new regime point, chunk numbers are given according to these inactive points. While energy consumption values are used as an output feature, the other available features are utilized as inputs for deep learning models so that the model can predict upcoming duration energy consumption based on previous input values. The input features for each dataset are shown in Table 1.

Various combinations were attempted to find the best hyperparameter values for DNN prediction structures using stochastic gradient descent (SGD) as an optimizer. Each training data chunk was split into 70% training, 15% validation, and 15% testing for training purposes. Hyperparameters were used in a variety of different combinations for DNN models during the training, a process also known as a grid search. Two layers, each containing 50 neurons, were selected as they provided more accuracy for the neural network structure than three layers. For activation function, tanh and rectifier functions were used, and the best one was selected. Similarly, different epoch sizes from 2 to 100 and learning rate variations in the range of 0.1 to 1 were used to find the optimum values, with early stop criteria used to avoid overfitting. The

experiments were performed on a computer with an AMD Ryzen 7 pro 4.20 GHz processor and 32 GB RAM.

A. DATASET DETAILS AND PREPROCESSING

All datasets were collected from industrial factories that work seven days a week, 24 hours a day. They operate in an open environment with no heaters or cooling facilities; therefore, the outside temperature does not affect energy use [12], [22]. By looking for a unit root, an Augmented Dickey-Fuller (ADF) test can tell if the series is stationary or not. The series has a unit root and is non-stationary, which is the null hypothesis for this test. The null hypothesis is rejected in the ADF test for all three different datasets, indicating that they are stationary.

The first dataset was gathered from a North American mining company for a SAG mill. There are a total of 103,732 data samples for the three years of data with a 15-minute resolution, and there are a total of 67 different chunks according to machine inactive points. The second dataset belongs to a South Korean Steel Factory. The steel industry's features and energy usage (kWh) were recorded every 15 minutes for 365 days. There are 35,041 samples with 35 different chunks for the selected threshold of minimum inactive duration. The third dataset was collected from an underground mine ventilation machine system from a Turkish Mining Company. There is a total of 28,153 hourly samples among the available data for around three years, with 22 different chunks. Missing values are replaced with the average of the corresponding feature so that we can use each record as much as possible.

Table 1 lists the names of features with each dataset's data type and unit. For a neural network to map inputs to outputs, the range of the feature values used to train the model is critical and varies widely. Model convergence can be ensured via data normalization, limiting the possibility of exploding gradients and slower learning processes [4]. On the time series, z-transformation is one of the most common techniques.

By using:

$$(z_i) = \frac{(x_i - X_m)}{(s)} \quad (3)$$

z_i = z-transformed sample observations
 x_i = original values of the sample

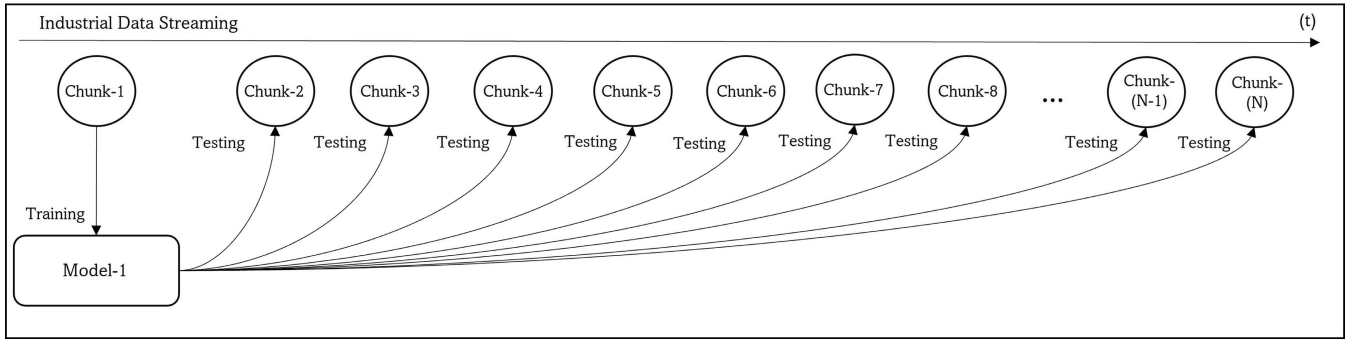


FIGURE 3. Framework of static modeling.

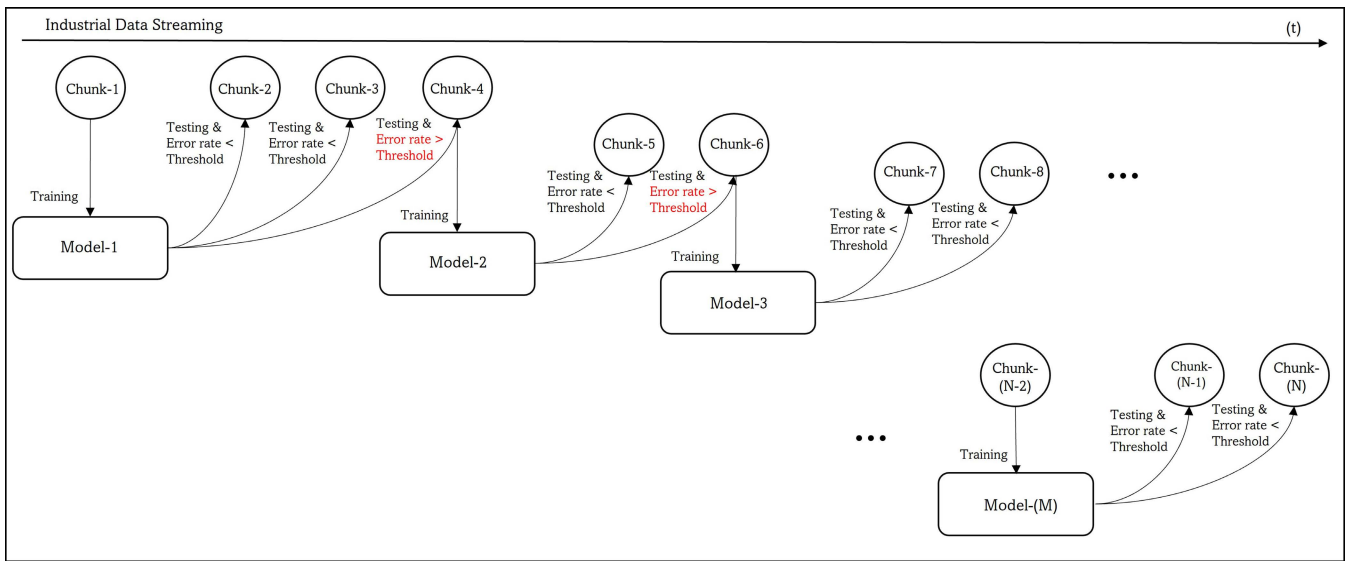


FIGURE 4. Framework of DMWOM.

X_m = sample mean

s = standard deviation of the sample

Only numerical time series are supported by this standardization method. Furthermore, it only alters the distribution's mean and standard deviation, not its shape. Additionally, energy usage is predicted using weather data, and occupancy characteristics as predictors [22]. However, since the industrial facilities are in an open environment with no heaters or cooling equipment, outside weather factors do not affect energy use for industrial machine working conditions [12], [22].

B. PERFORMANCE METRICS

Feedback-based strategies are necessary to assess learners' performance when dealing with concept drift [15]. For continuous variables, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are the most commonly used performance metrics for measuring accuracy and the average size of errors in a prediction set [43]. However, because RMSE gives significant errors a greater load, it is advantageous when such errors should be avoided [4], [9].

The RMSE examines the discrepancies between real and estimated values, which are calculated using Equation 3. MAE determines how close estimations or expectations are to actual outcomes. As indicated in Equation 4, this is calculated by averaging the absolute divergences between the expected and actual values. MAPE is a percentage-based measure of the precision of estimated values compared with real values, computed using Equation 5. The limitations of MAPE are paired with the MAE, which displays how much inaccuracy is expected from the forecast on average, assisting in determining which models are superior. Because the MAE has trouble differentiating between significant and minor mistakes, it was paired with the RMSE to be safe [9], [20].

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2} \tag{4}$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |Y_i - \hat{Y}_i| \tag{5}$$

$$MAPE = \frac{100}{m} \sum_{i=1}^m \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \tag{6}$$

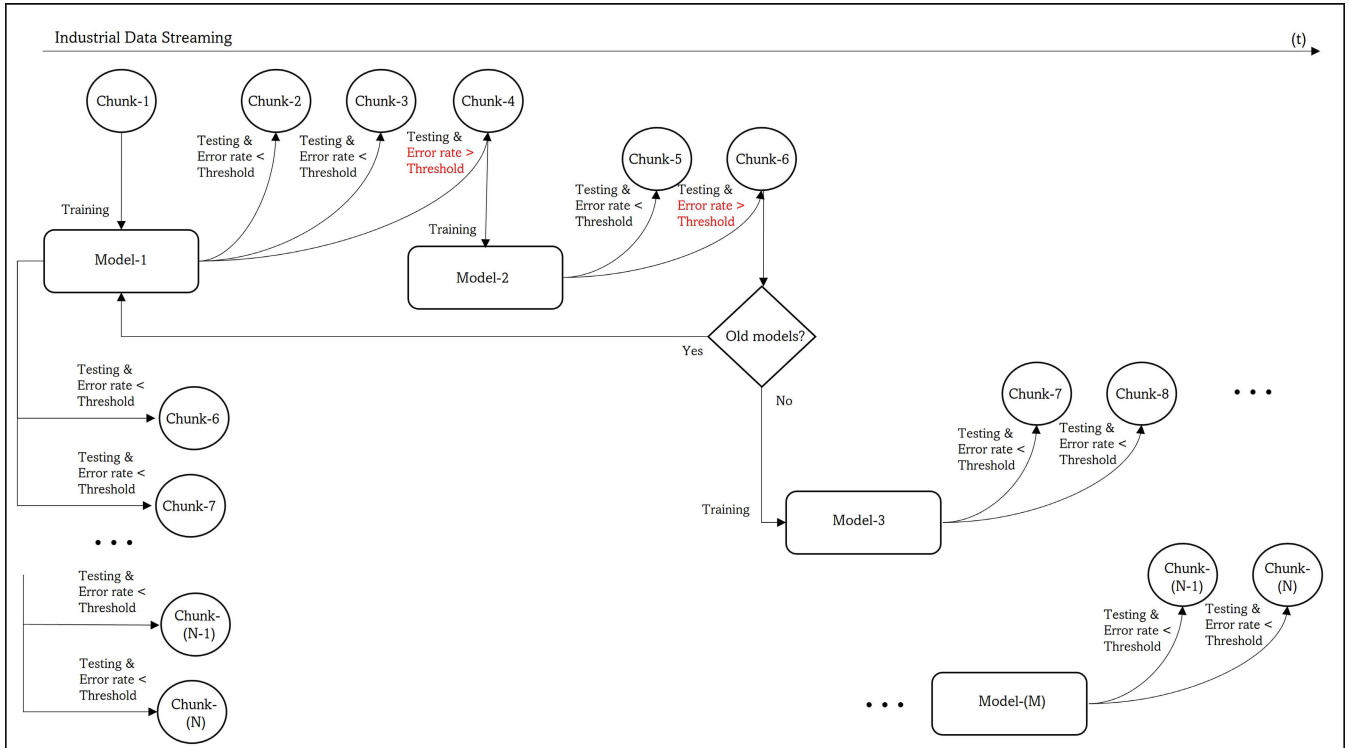


FIGURE 5. Framework of DMWM.

The test set contains m samples, with Y_i representing the sample’s actual value and \hat{Y}_i representing the sample’s anticipated value. The lower the value of these error parameters, the more accurate the model. MAPE was also utilized to determine the threshold values and assess the model’s overall performance since inference of MAPE is more straightforward than other error rates.

C. STATIC (TRADITIONAL) MODELING

Most researchers split the whole data into training and testing parts. However, static modeling does not detect possible concept drifts for streamed data over time. Furthermore, machine running conditions change over time, and a number of unknown working conditions and other hidden features may exist. Traditional static modeling is implemented so that a comparison can be made for results with proposed dynamic modeling.

Figure 3 shows the steps for the static (traditional) method. It trains one model based on the earliest part of the data, and the rest of the other upcoming chunks are used for testing. When upcoming data evolve, the static modeling does not solve the performance degradation problem.

D. DYNAMIC MODELING WITHOUT MEMORY (DMWOM)

In the best-known active drift detection systems, the user determines the window size, which uses the most recent batch of data with the most recent training instances [20]. However, standard drift detection methods are not designed for complex data streams from industrial machines. Using a fixed sample

size for training chunks might miss necessary samples since the user decides the size [20]. Specific periods of inactivity are considered the starting point for a new chunk. Thanks to the proposed dynamic approach, each chunk might have a different sample size and duration according to consecutive inactive time durations.

The steps of the DMWOM method are illustrated in Figure 4. First, a counter is designed to give a chunk number for each coming data part between two consecutive machine inactive points. According to Chunk-1, the prediction model is trained to be used over the upcoming chunks. A threshold is set for the model performance error rate. When the model error exceeds the threshold, a new model is trained based on the current chunk data, such as Chunk-4 and Chunk-6 for Figure 4. In other words, if the threshold value is exceeded n times, n models will be developed. However, since the system does not have a memory for old models, DMWOM only develops one model based on the recent chunk. Compared with a static model, DMWOM provides a better overall prediction performance since it builds a new model when the error rate exceeds the threshold chosen by the user.

E. DYNAMIC MODELING WITH MEMORY (DMWM)

Since DMWOM does not detect repetitive regimes, it builds a new model when error rates pass the threshold. However, optimizing the number of models is necessary for future streaming data. This is where a DMWM model can be applied. First, trained models will be stored in a memory, and before developing a new model for required chunks, all

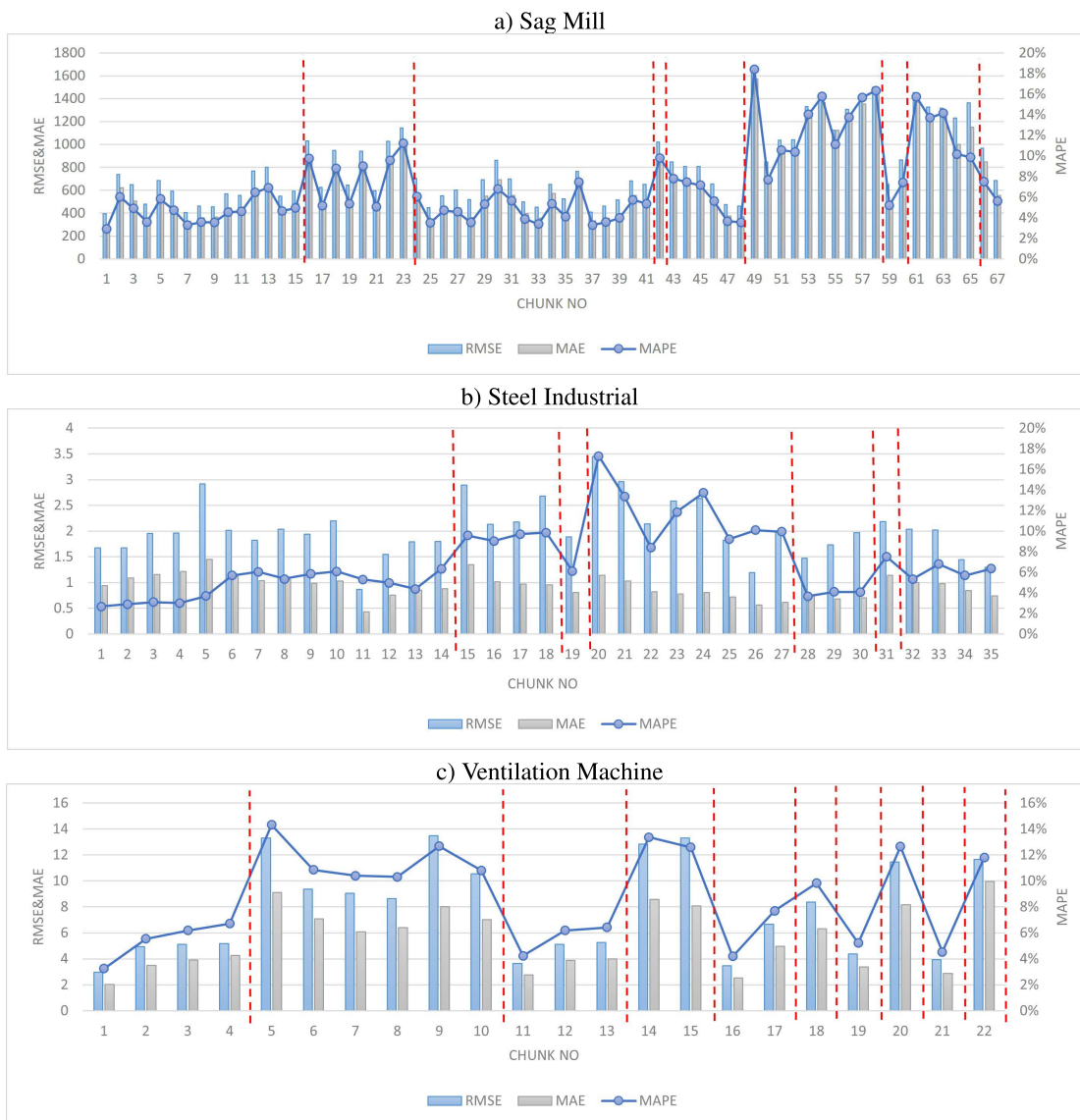


FIGURE 6. Static modeling energy consumption prediction performances for a) SAG Mill, b) Steel industrial, and c) Ventilation machine.

previous models stored will be used. If any of the former trained models provides an error rate lower than the threshold, it will be used and a repetitive regime will be considered for the machine running conditions as shown in Figure 5.

If old models cannot provide an error rate lower than the threshold, a new model will be trained based on recent chunk data. For Chunk-6, when the error rate passes the threshold, before developing a new model, DMWM checks the old models. If Model-1 gives an error rate lower than the threshold, it will be used for future chunks. If older models cannot provide an error rate lower than the threshold, then a new model will be trained based on Chunk-6 data (Model-3 in Figure 5), and this new model will be used for upcoming chunks. The identical approach will be repeated for future chunks so that the old model can provide an error rate lower than the threshold, which means recurring machine running conditions.

IV. RESULTS AND DISCUSSION

Since each dataset has a different recorded period and various chunk sizes, based on a selected minimum of machine inactive durations for the “SAG Mill,” “Steel Industrial,” and “Underground Mine Ventilation Machine” chunks, the average sample sizes are 1236, 824, and 1279, respectively. Energy consumption (EC) prediction performances will be shown in this section for three different datasets.

A. RESULTS FOR STATIC APPROACH

First, we developed a DNN model based on Chunk-1, and upcoming chunks were used for testing. The static approach does not build new models when performance degrades over time. All error rates for each chunk are illustrated in Figure 6. There are two vertical axes since MAPE values have different ranges than RMSE and MAE. SAG Mill EC prediction performance is shown in Figure-6a; error rate increases

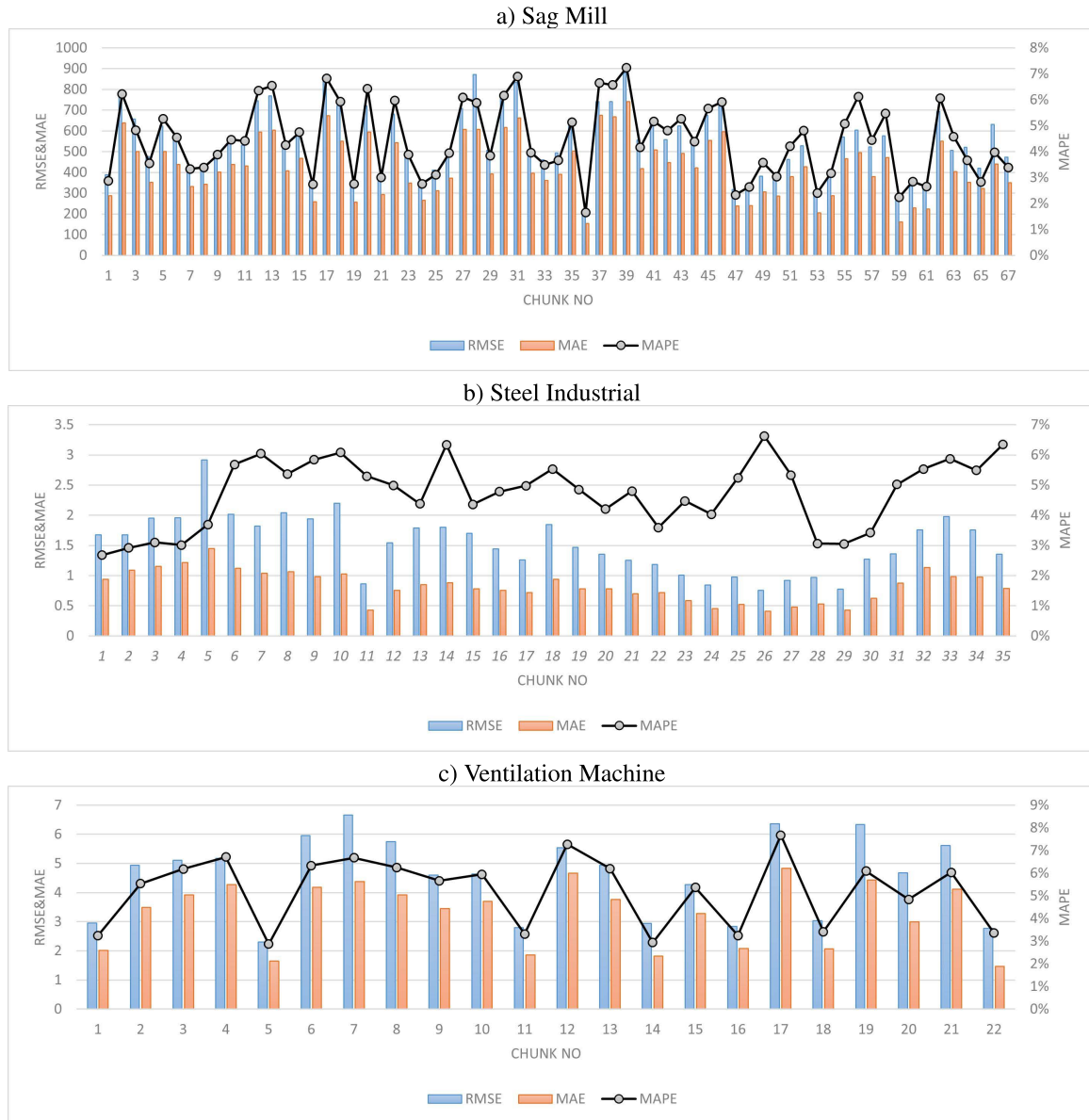


FIGURE 7. DMWOM energy consumption prediction performances for a) SAG Mill, b) Steel industrial, and c) Ventilation machine.

are shown from Chunk-16 to Chunk-23, in Chunk-42, from Chunk-49 to Chunk-58, and between Chunk-61 and Chunk-64, with the dotted red lines demarking concept drift points.

Figure-6b for Steel Industrial EC shows forecast performance accuracy diminishing between Chunk-15 and Chunk-18, and between Chunk-20 and Chunk-27, because of data changes. Lastly, the Ventilation machine EC prediction in Figure-6c shows degradation from Chunk-5 to Chunk-10, and in chunks 14, 15, 18, 20, and 22.

Static modeling does not provide solutions for potential data changes within streaming data. Because of this issue, dynamic modeling is necessary for an industrial dataset with a concept drift problem. We split the dynamic modeling process into two parts: without memory, and with memory. The purpose of the memory approach is that if there are any similar repetitive regimes, they will be detected. As a

result, it allows us to optimize the number of models and find potential repetitive machine running conditions.

B. RESULTS FOR DYNAMIC MODELING WITHOUT MEMORY (DMWOM)

Prediction performance degrades over time for upcoming chunks, so it is necessary to build a new model based on recent data to handle the concept drift problem. DMWOM uses only one recently trained model for predicting upcoming streaming data. According to the selected threshold error rate, a new model will be developed for the current chunk. A new model will be trained according to the last chunk when the prediction error exceeds the decided threshold values. It is selected as the static model prediction performance average value for this paper, but users can select it differently.

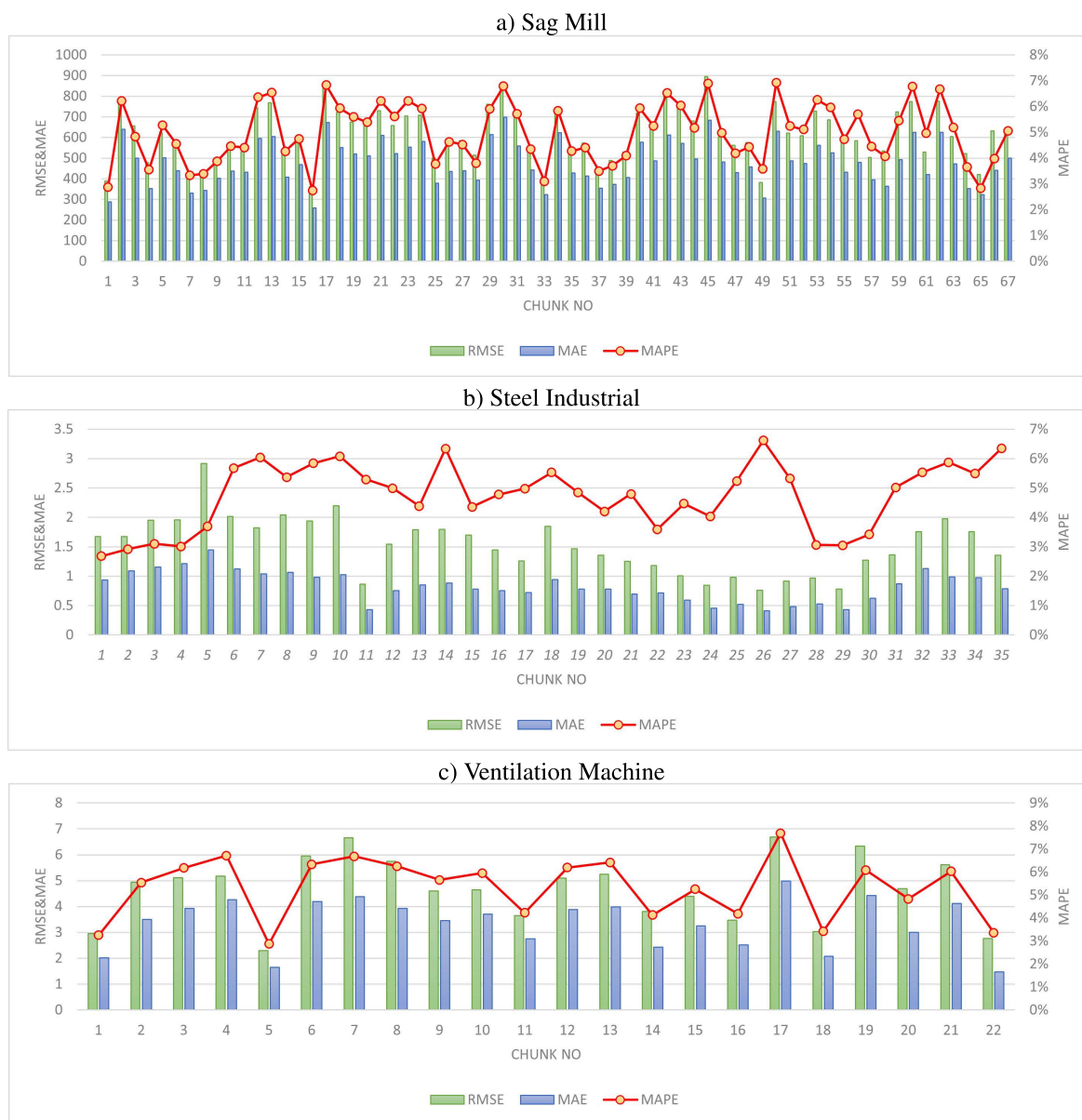


FIGURE 8. DMWM energy consumption prediction performances for a) SAG Mill, b) Steel industrial, and c) Ventilation machine.

TABLE 2. Performance metrics for models and datasets.

	SAG Mill			Steel Industrial			Ventilation Machine		
	Time(min)	MAPE	T. Num. of Mod.	Time(min)	MAPE	T. Num. of Mod.	Time(min)	MAPE	T. Num. of Mod.
Static Model	0.77	7.42%	1	0.45	7.06%	1	0.35	8.63%	1
DMWOM	5.68	4.45%	16	0.98	4.74%	4	1.37	5.24%	7
DMWM	2.38	4.98%	5	0.98	4.74%	4	0.88	5.33%	4

Error rates for DMWOM are shown in Figure 7 for SAG Mill, Steel Industrial, and Ventilation Machine, and DMWOM had 16, four, and seven different models, respectively. Overall, DMWOM prediction performances were 4.45%, 4.74%, and 5.24%, respectively, while static modeling general MAPE prediction performances were 7.42%, 7.06%, and 8.63%, in that order. It can be seen that general EC prediction performance precision increased thanks to the dynamic approach. However, the number of models has also risen since DMWOM develops a new model when the error rate exceeds

the threshold. To decrease the total number of models without decreasing prediction performance, DMWM is designed as a solution that also detects possible machine repetitive energy consumption regimes.

C. RESULTS FOR DYNAMIC MODELING WITH MEMORY (DMWM)

In contrast to DMWOM applications, DMWM uses all previously trained models when MAPE exceeds the threshold rather than creating a new model for the most recent chunk.

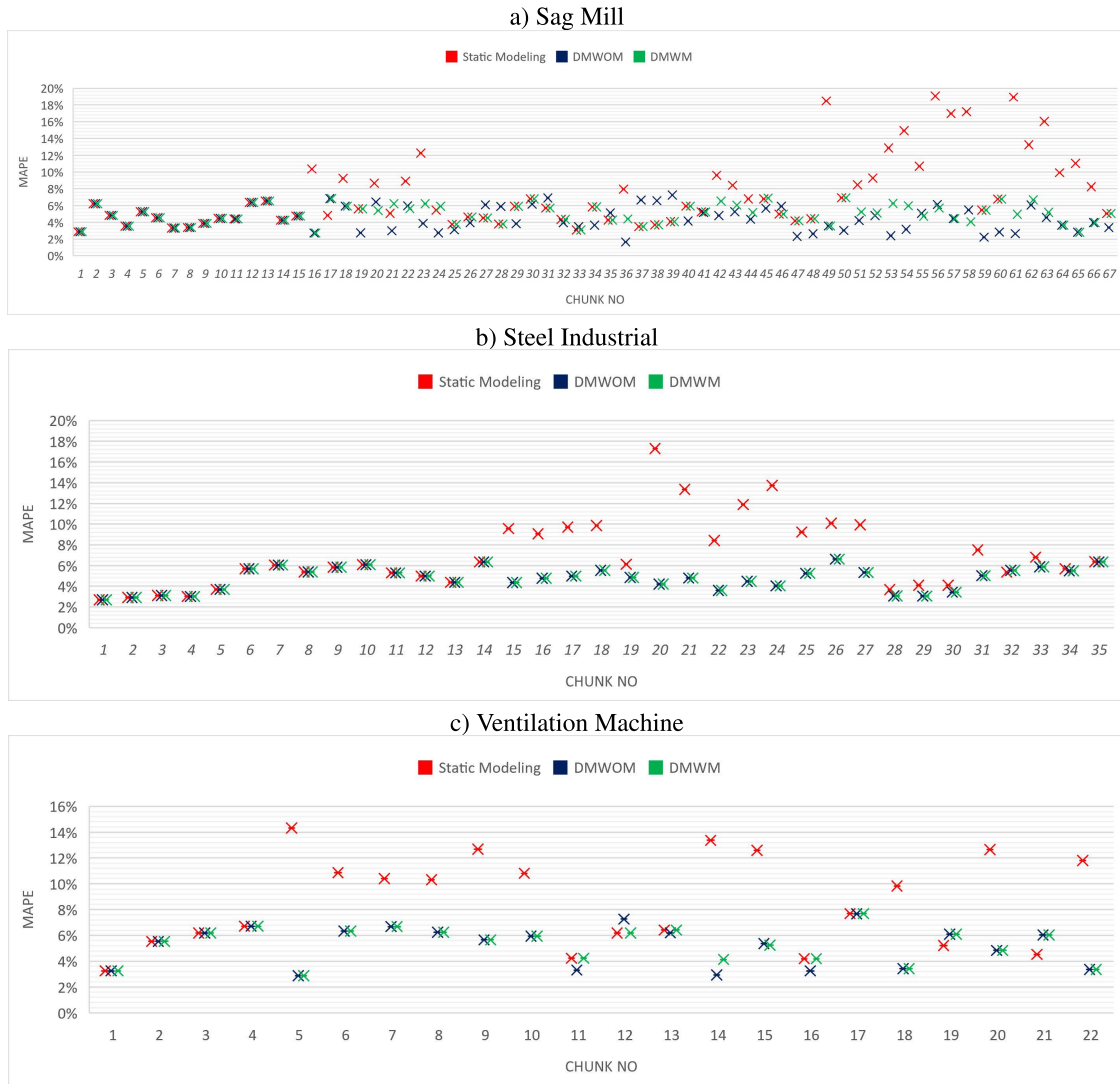


FIGURE 9. Comparison of static modeling, DMWOM, and DMWM energy consumption prediction performance in MAPE for a) SAG Mill, b) Steel industrial, and c) Ventilation machine.

A new model will be developed for the latest chunk of samples if any of the earlier models are unable to deliver an error rate below the threshold. The number of models has been optimized thanks to DMWM, which also determines whether repeated regimes exist or not.

Figure 8 displays DMWM error rates for the SAG Mill, Steel Industrial, and Ventilation Machine datasets. While DMWOM used 16, four, and seven distinct models for the available datasets, DMWM had five, four, and four models, respectively. While DMWM decreased the total number of models thanks to the repetitive regimes for SAG Mill and Ventilation Machine, it did not find any repetitive regimes for the Steel Industrial dataset. DMWM provided overall error rates slightly higher than DMWOM. For SAG Mill, Steel Industrial, and Ventilation Machine datasets, DMWM overall MAPE prediction performances were 4.98%, 4.74%, and 5.33%, and DMWOM overall MAPE values were 4.45%, 4.74 %, and 5.24%, respectively. The absence of repetitive models means no repetitive regimes in available Steel Indus-

trial data. However, the dataset has a limited time duration of only one year, and further data might have repetitive regimes that can be detected thanks to the DMWM.

D. COMPARISON OF STATIC MODELING, DMWOM, AND DMWM

DMWOM and DMWM had a better performance due to creating a new model for chunks that passes the threshold. In a long runtime, DMWM can detect possible repetitive regimes for industrial machines’ energy consumption and reduce the total number of models. Figure 9 shows each chunk MAPE rate for static modeling, DMWOM, and DMWM.

One of the distinctive chunks of prediction performance is shown as a plot for each dataset in Figure 10, with x-axes representing real EC values and y-axes illustrating prediction values. Compared with the static approach, EC prediction performance accuracies increased for specific chunks thanks to the proposed model after detecting concept drift. SAG Mill EC, Steel Industrial EC, and Ventilation Machine EC

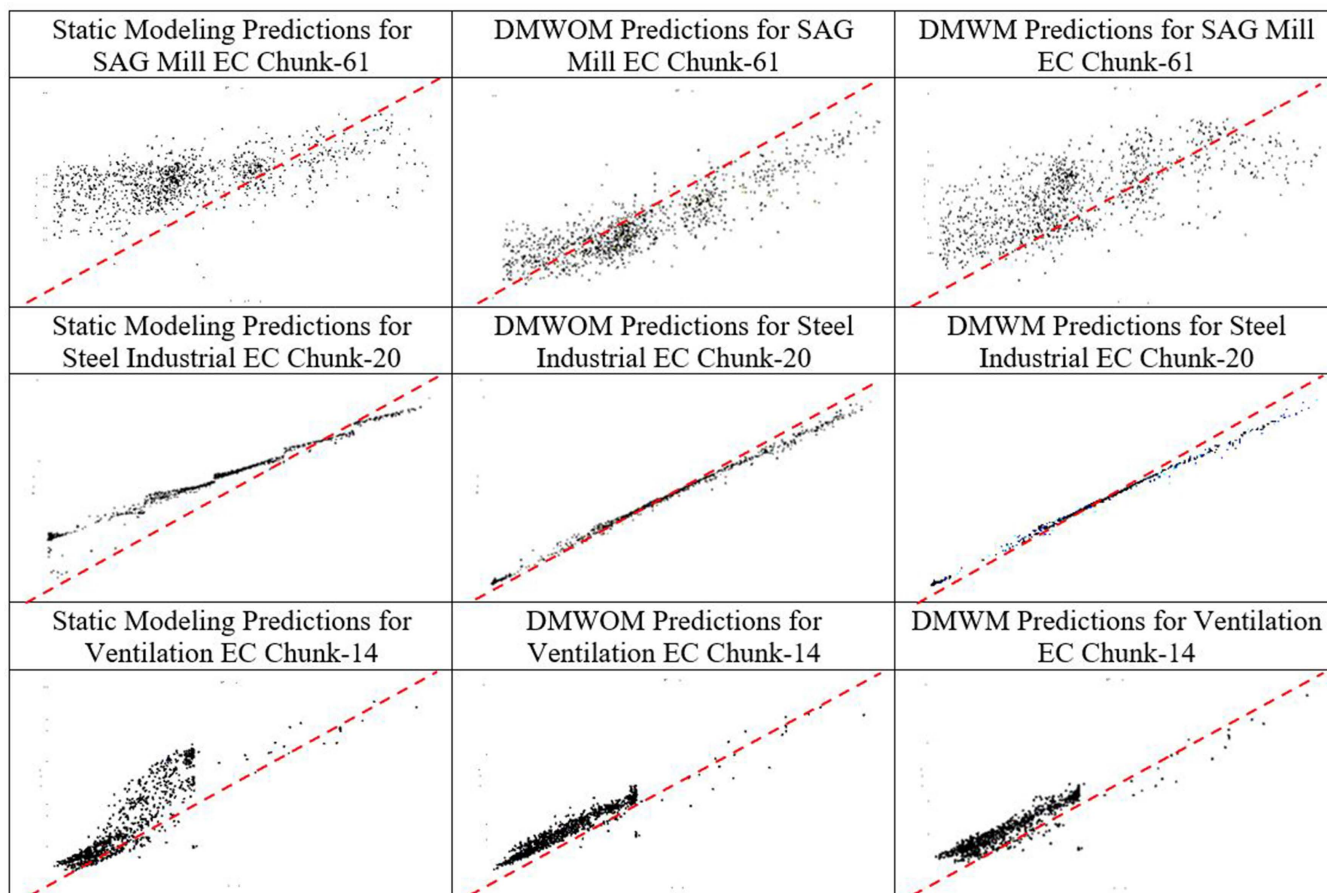


FIGURE 10. Comparison of static modeling, DMWOM, and DMWM energy consumption prediction performance plots for SAG Mill, steel industrial, and ventilation machine.

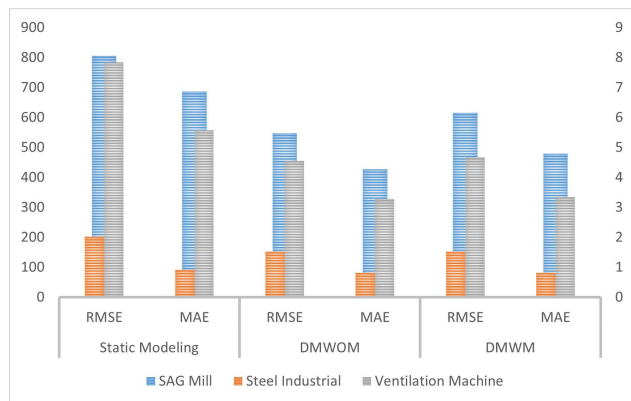


FIGURE 11. Comparison of static modeling, DMWOM, and DMWM energy consumption prediction performance in RMSE and MAE for SAG Mill, steel industrial, and ventilation machine.

prediction plots are illustrated for Chunk-61, Chunk-20, and Chunk-14, respectively.

It can be seen from Figure 10 that static modeling prediction values deviated from the actual values, which indicates concept drifts for streaming data over time for specific chunks. As a result, the proposed dynamic modeling can obtain a better prediction performance for real-world applica-

tions compared with traditional static modeling. Furthermore, DMWM decreased the number of models thanks to detecting repetitive machine running regimes.

Additionally, overall average RMSE and MAE values are shown in Figure 11. Compared with static modeling, DMWOM and DMWM had lower error values. There are two horizontal axes in Figure 11 since each dataset has different EC value ranges.

Comparisons of the methods applied for the datasets in terms of training time, error rates, and the total number of developed models are shown in Table 2. The values are also presented graphically in Figure 12. While dynamic modeling has a lower error rate than static modeling for all datasets, it needs a longer training time and a larger number of models. Dynamic modeling with higher accuracy estimation provides an advantage over the traditional approach in data analysis of industrial machines. While DMWOM needed more models and training time, DMWM was able to achieve approximately the same prediction performance with less model and training time thanks to its ability to use old models.

V. CONCLUSION AND FUTURE WORK

Deep learning architectures have been widely deployed for the forecasting of sensor-based electrical loads. Most

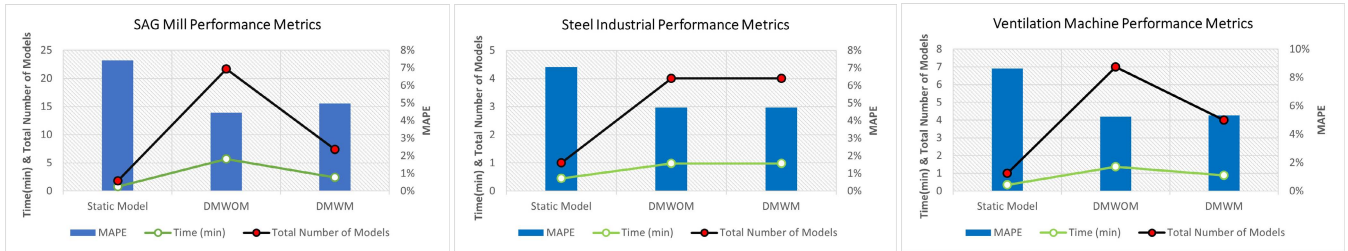


FIGURE 12. Comparison of static modeling, DMWOM, and DMWM energy consumption prediction performance in MAPE for SAG Mill, steel industrial, and ventilation machine.

methods use a model that has been trained only once and then used to predict future loads. As a result, these strategies do not benefit from the latest data, and the performance of the models generally deteriorates over time.

Industrial datasets may contain several unknown features which require a solution with a mandatory modification based on available data and attributes. More integrative solutions for complex systems are required to achieve better prediction performances.

This research has proposed a data-driven dynamic technique with an adjusted concept drift detection method to predict the energy consumption of three different real-world industrial datasets. Compared with a static methodology, the dynamic method maintains a better prediction performance thanks to adaptive modeling. While DMWOM's overall EC prediction performances were 4.45 percent, 4.74 percent, and 5.24 percent, respectively, MAPE's overall prediction performances for static modeling were 7.42 percent, 7.06 percent, and 8.63 percent for the SAG Mill, Steel Industrial, and Ventilation Machine datasets used here. Moreover, DMWM reduced computing complexity by requiring less training while improving prediction accuracy. The proposed method can be tested on various streamed datasets in future works.

In addition, this study has used deep learning as a prediction method. Various machine learning models (SVM, RF, etc.) can be integrated into the proposed method to compare their predictive performance and runtime in future research.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- [1] G. Y. Zhao, Z. Y. Liu, Y. He, H. J. Cao, and Y. B. Guo, "Energy consumption in machining: Classification, prediction, and reduction strategy," *Energy*, vol. 133, pp. 142–157, Aug. 2017.
- [2] J. Walther and M. Weigold, "A systematic review on predicting and forecasting the electrical energy consumption in the manufacturing industry," *Energies*, vol. 14, no. 4, p. 968, 2021.
- [3] Homepage—U.S. Energy Information Administration (EIA). Accessed: Jul. 7, 2022. [Online]. Available: <https://www.eia.gov/energyexplained/use-of-energy/>
- [4] V. J. Mawson and B. R. Hughes, "Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector," *Energy Buildings*, vol. 217, Jun. 2020, Art. no. 109966.
- [5] D. Ramos, P. Faria, Z. Vale, J. Mourinho, and R. Correia, "Industrial facility electricity consumption forecast using artificial neural networks and incremental learning," *Energies*, vol. 13, no. 18, p. 4774, 2020.
- [6] M. Iannario, M. Manisera, D. Piccolo, and P. Zuccolotto, "Sensory analysis in the food industry as a tool for marketing decisions," *Adv. Data Anal. Classification*, vol. 6, no. 4, pp. 303–321, Dec. 2012.
- [7] P. Faria and Z. Vale, "A demand response approach to scheduling constrained load shifting," *Energies*, vol. 12, no. 9, p. 1752, May 2019.
- [8] M. N. Fekri, H. Patel, K. Grolinger, and V. Sharma, "Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116177.
- [9] J. Moon, Y. Kim, M. Son, and E. Hwang, "Hybrid short-term load forecasting scheme using random forest and multilayer perceptron," *Energies*, vol. 11, no. 12, p. 3283, Nov. 2018.
- [10] D. Smith, K. Jaskie, J. Cadigan, J. Marvin, and A. Spanias, "Machine learning for fast short-term energy load forecasting," in *Proc. IEEE Conf. Ind. Cyberphysical Syst. (ICPS)*, vol. 1, Jun. 2020, pp. 433–436.
- [11] N. M. M. Bendaoud and N. Farah, "Using deep learning for short-term load forecasting," *Neural Comput. Appl.*, vol. 32, no. 18, pp. 15029–15041, 2020.
- [12] V. Sathishkumar, M.-B. Lee, J.-H. Lim, C.-S. Shin, C.-W. Park, and Y. Y. Cho, "Hourly steel industry energy consumption prediction using machine learning algorithms," in *Proc. Korea Inf. Process. Soc. Conf.*, 2019, pp. 585–588.
- [13] A. Kahraman, M. Kantardzic, M. M. Kahraman, and M. Kotan, "A data-driven multi-regime approach for predicting energy consumption," *Energies*, vol. 14, no. 20, p. 6763, Oct. 2021.
- [14] F. Bayram, B. S. Ahmed, and A. Kassler, "From concept drift to model degradation: An overview on performance-aware drift detectors," *Knowl.-Based Syst.*, vol. 245, Jun. 2022, Art. no. 108632.
- [15] H. Mehmood, P. Kostakos, M. Cortes, T. Anagnostopoulos, S. Pirttikangas, and E. Gilman, "Concept drift adaptation techniques in distributed environment for real-world data streams," *Smart Cities*, vol. 4, no. 1, pp. 349–371, Mar. 2021.
- [16] R. Elwell and R. Polikar, "Incremental learning of concept drift in non-stationary environments," *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1517–1531, Oct. 2011.
- [17] G. Ditzler and R. Polikar, "Incremental learning of concept drift from streaming imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2283–2301, Oct. 2012.
- [18] R. Bhinge, J. Park, K. H. Law, D. A. Dornfeld, M. Helu, and S. Rachuri, "Toward a generalized energy prediction model for machine tools," *J. Manuf. Sci. Eng.*, vol. 139, no. 4, pp. 1–22, Apr. 2017.
- [19] C. Czech, M. Lesjak, C. Bach, and F. Duddeck, "Data-driven models for crashworthiness optimisation: Intrusive and non-intrusive model order reduction techniques," *Struct. Multidisciplinary Optim.*, vol. 65, no. 7, pp. 1–15, Jul. 2022.
- [20] D. Mariano-Hernández, L. Hernández-Callejo, M. Solís, A. Zorita-Lamadrid, O. Duque-Pérez, L. Gonzalez-Morales, F. S. García, A. Jaramillo-Duque, A. Ospino-Castro, V. Alonso-Gómez, and H. J. Bello, "Analysis of the integration of drift detection methods in learning algorithms for electrical consumption forecasting in smart buildings," *Sustainability*, vol. 14, no. 10, p. 5857, May 2022.
- [21] A. S. Rahman, M. Lee, J. Lim, Y. Cho, and C. Shin, "A prediction model for steel factory manufacturing product based on energy consumption using data mining technique," *J. Sci. Eng. Manage.*, vol. 1, no. 2, pp. 9–16, 2020.

- [22] V. E. Sathishkumar, J. Lim, M. Lee, K. Cho, J. Park, C. Shin, and Y. Cho, "Industry energy consumption prediction using data mining techniques," *Int. J. Energy, Inf. Commun.*, vol. 11, no. 1, pp. 7–14, Feb. 2020, doi: 10.21742/IJEIC.2020.11.1.02.
- [23] A. A. Hadi, C. A. S. Silva, E. Hossain, and R. Chaloo, "Algorithm for demand response to maximize the penetration of renewable energy," *IEEE Access*, vol. 8, pp. 55279–55288, 2020.
- [24] S. M. Bhuiyan, J. F. Khan, and G. V. Murphy, "Big data analysis of the electric power PMU data from smart grid," in *Proc. SoutheastCon*, Mar. 2017, pp. 1–5.
- [25] X. Li, L. Gao, G. Wang, F. Gao, and Q. Wu, "Investing and pricing with supply uncertainty in electricity market: A general view combining wholesale and retail market," *China Commun.*, vol. 12, no. 3, pp. 20–34, Mar. 2015.
- [26] A. de Hauteclocque, "Legal uncertainty and competition policy: The case of long-term vertical contracting by dominant firms in the EU electricity markets," in *Proc. 5th Int. Conf. Eur. Electr. Market*, May 2008, pp. 1–8.
- [27] K. L. Lum, H. K. Mun, S. K. Phang, and W. Q. Tan, "Industrial electrical energy consumption forecasting by using temporal convolutional neural networks," in *Proc. MATEC Web Conf.*, vol. 335, 2021, p. 02003.
- [28] J. Zhang, H. Zhang, S. Ding, and X. Zhang, "Power consumption predicting and anomaly detection based on transformer and K-means," *Frontiers Energy Res.*, vol. 9, p. 681, Oct. 2021.
- [29] J. Li, Y. Guo, X. Zhang, and Z. Fu, "Using hybrid machine learning methods to predict and improve the energy consumption efficiency in oil and gas fields," *Mobile Inf. Syst.*, vol. 2021, pp. 1–7, Dec. 2021.
- [30] Y. He, P. Wu, Y. Li, Y. Wang, F. Tao, and Y. Wang, "A generic energy prediction model of machine tools using deep learning algorithms," *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115402.
- [31] C. Chen, Y. Liu, M. Kumar, and J. Qin, "Energy consumption modelling using deep learning technique—A case study of EAF," *Proc. CIRP*, vol. 72, pp. 1063–1068, Jan. 2018.
- [32] G. Kant and K. S. Sangwan, "Predictive modelling for energy consumption in machining using artificial neural network," *Proc. CIRP*, vol. 37, pp. 205–210, Jan. 2015.
- [33] S. Avalos, W. Kracht, and J. M. Ortiz, "Machine learning and deep learning methods in mining operations: A data-driven SAG mill energy consumption prediction application," *Mining, Metall. Explor.*, vol. 37, no. 4, pp. 1197–1212, Aug. 2020.
- [34] M. A. Bermeo-Ayerbe, C. Ocampo-Martinez, and J. Diaz-Rozo, "Data-driven energy prediction modeling for both energy efficiency and maintenance in smart manufacturing systems," *Energy*, vol. 238, Jan. 2022, Art. no. 121691.
- [35] D. Jayaratne, D. De Silva, D. Alahakoon, and X. Yu, "Continuous detection of concept drift in industrial cyber-physical systems using closed loop incremental machine learning," *Discover Artif. Intell.*, vol. 1, no. 1, pp. 1–13, 2021.
- [36] W. Zheng, P. Zhao, G. Chen, H. Zhou, and Y. Tian, "A hybrid spiking neurons embedded LSTM network for multivariate time series learning under concept-drift environment," *IEEE Trans. Knowl. Data Eng.*, early access, May 27, 2022, doi: 10.1109/TKDE.2022.3178176.
- [37] X. You, M. Zhang, D. Ding, F. Feng, and Y. Huang, "Learning to learn the future: Modeling concept drifts in time series prediction," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2021, pp. 2434–2443.
- [38] A. Kahraman, M. Kantardzic, M. M. Kahraman, and M. Kotan, "Sequential pattern mining method for predictive maintenance of large mining trucks," in *Proc. Medit. Forum, Data Sci. Conf.*, J. H. Telalovic and M. Kantardzic, Eds. Cham, Switzerland: Springer, 2021, pp. 126–136.
- [39] L. Baier, M. Hofmann, N. Kühl, M. Mohr, and G. Satzger, "Handling concept drifts in regression problems—The error intersection approach," 2020, *arXiv:2004.00438*.
- [40] J. Reimann, K. Wenzel, M. Friedemann, and M. Putz, "Methodology and model for predicting energy consumption in manufacturing at multiple scales," *Proc. Manuf.*, vol. 21, pp. 694–701, Jan. 2018.
- [41] J. Lu, A. Liu, Y. Song, and G. Zhang, "Data-driven decision support under concept drift in streamed big data," *Complex Intell. Syst.*, vol. 6, no. 1, pp. 157–163, Apr. 2020.
- [42] S. K. Mohapatra, S. Mishra, H. K. Tripathy, A. K. Bhoi, and P. Barsocchi, "A pragmatic investigation of energy consumption and utilization models in the urban sector using predictive intelligence approaches," *Energies*, vol. 14, no. 13, p. 3900, Jun. 2021.
- [43] M. Kalimoldayev, A. Drozdenko, I. Kopylyk, T. Marinich, A. Abdildayeva, and T. Zhukabayeva, "Analysis of modern approaches for the prediction of electric energy consumption," *Open Eng.*, vol. 10, no. 1, pp. 350–361, Apr. 2020.



ABDULGANI KAHRAMAN received the master's degree in computer engineering and science from the University of Louisville, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include data mining, machine learning, green energy, and image processing.



MEHMED KANTARDZIC received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in computer science from the University of Sarajevo, Bosnia and Herzegovina, in 1972, 1976, and 1980, respectively. He served as an Assistant and an Associate Professor with the University of Sarajevo and later as an Associate. Since 2004, he has been a Full Professor with the University of Louisville. He is currently the CSE Associate Chair, the Director of the Data Mining

Laboratory, and the Director of CSE graduate studies with the CSE Department. He is the author of six books, including the textbook *Data Mining: Concepts, Models, Methods, and Algorithms* (John Wiley, third edition, 2019), which is accepted for data mining courses at more than hundred universities in USA and abroad, and translated into Chinese, Japanese, Spanish, and Farsi. He is the author of over 40 peer-reviewed journal publications, 20 book chapters, and more than 200 reviewed papers in the proceedings of international conferences. His recent research projects are supported by NSF, USA Treasury Department, USA Army, and NASA. His research interests include data mining knowledge discovery, machine learning, soft computing, click fraud detection and prevention, concept drift in streaming data, and medical data mining. He was elected as a Fulbright Specialist in information technology, in 2012. For his contributions in data mining field, he was also elected as a member of the Academy of Sciences and Arts of Bosnia and Herzegovina, in March 2019. He has served on an editorial board for several international journals. He is also serving an Associate Editor for *WIREs Data Mining and Knowledge Discovery* journal. He was a Reviewer and a Panel Member of NSF, KSEF-SBIR, NASA, the Austrian Science Foundation (ASF), the Canadian Research Council (NSERC), the Croatian UKF Fund, RAND NPRP, the Chilean Science Foundation, the National Science Center in Poland, and the Norman Hackerman Advanced Research Program (NHARP). He was invited keynote speaker at conferences in USA, Mexico, Bosnia, Egypt, Japan, India, and Algeria, and his service activities include a General Chair or a Program Chair assignments for several international conferences, including IEEE ICMLA 2018, HICSS17, GOCICT 2016, HICSS16, INSS-BigData15, IEEE ICAT13, ICMLA09, IEEE ISSPIT08, and WSIPT07.



MUHAMMED KOTAN received the M.S. and Ph.D. degrees in computer and information engineering from Sakarya University, Turkey, in 2014 and 2020, respectively. He is currently working with the Information Systems Engineering Department, Sakarya University. His current research interests include image processing, computer vision, and artificial intelligence.