Multi-faceted Modelling for Strip Breakage in Cold Rolling Using Machine Learning

Zheyuan Chen[a], Ying Liu[a*], Agustin Valera-Medina[a], Fiona Robinson[b]

and Michael Packianather[a]

[a] *Institute of Mechanical and Manufacturing Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK*

[b] *Cogent Power Ltd, Orb Electrical Steels, PO Box 30, Stephenson Street, Newport, NP19 0RB, UK*

**Corresponding author. E-mail: LiuY81@Cardiff.ac.uk.*

ABSTRACT

In the cold rolling process of steel strip products, strip breakage is an undesired production failure which can lead to yield loss, reduced work speed and equipment damage. In order to perform a root cause analysis, conventional physics-based approaches which focus on mechanical and metallurgical principles have been applied in a retrospective manner. With the advancement of data acquisition technologies, a large amount of process monitoring data is collected by various sensors deployed along the cold rolling process; however, conventional approaches cannot take advantage of these data. In this paper, a machine learning-based approach is proposed to characterise and model strip breakage in a predictive manner. First, in order to match the temporal characteristic of strip breakage which occurs instantaneously, historical multivariate time-series data of a cold rolling process were extracted in a run-to-failure manner, and a sliding window strategy was adopted for data annotation. Second, breakage-centric features were identified from three facets — physics-based approaches, empirical knowledge and data-driven features. Finally, these features were used as inputs for strip breakage modelling using recurrent neural networks (RNNs), which are specialised in discovering underlying patterns embedded in time-series data. An experimental study using real-world data collected from a cold-rolled electrical steel strip manufacturer revealed the effectiveness of the proposed approach.

Keywords: Strip Breakage; Cold Rolling; Process Modelling; Quality Improvement; Machine Learning; Recurrent Neural Network.

1. INTRODUCTION

Due to its high efficiency and accuracy, the cold rolling process is a primary metal forming process for the manufacturing of steel strips [1]. An increasing demand for cold-rolled products has aroused widespread concern for maintaining the production continuity of cold rolling. However, cold rolling can encounter certain unexpected production failures which cause unplanned interruptions of the process. Strip breakage is one of the most common and undesired production failures for the cold rolling of strip products [2]. This failure has serious consequences, such as yield loss due to unplanned stops of the rolling mill, extended downtime caused by severe damage of work rolls and altered rolling performance for subsequent rolling when production resumes following a strip breakage [3-5].

Numerous studies of strip breakage have been conducted, and their approaches can be generally classified into two different categories. The first type of approach, which is referred to as the conventional approach, addresses strip breakage by employing mechanical or metallurgical theories. According to related research [3, 6-8], the causes of strip breakage are various and can be generally classified as equipment factors, material defects, improper operation, sensor malfunction or production adjustment. The limitation of the conventional approach is its retrospective manner which focuses on cause analyses after the occurrence of this failure rather than a predictive approach.

The data-driven approach, the second type of approach, has been employed within the last two decades with the advancement of technologies which facilitate data acquisition and storage for complex manufacturing processes [9]. With the deployment of various sensors and accurate measurement devices throughout the modern cold rolling process, process data such as coil entry and exit speed, forward and backwards tension, roll gap position and eccentricity of the cold rolling system are measured in real-time, and a large amount of multivariate time-series data is collected and stored. In this data-rich environment, data-driven approaches to investigating strip breakage have previously been applied in a handful of works [5, 10, 11]. Despite the advantage of being able to extract useful knowledge and make appropriate decisions using the data-driven approach, three questions have rarely been explored. First, these works were conducted with the aim of quality characterisation [12], which is the primary step for quality improvement, rather than quality prediction [13]. Second, the rationale for determining the variables for breakage modelling was not explained and justified. Third, the granularity of the data used in these works cannot match the temporal characteristic of strip breakage, which occurs instantaneously.

In light of these questions, we propose a predictive, data-driven approach to model strip breakage, one which uses multi-faceted features. Recurrent neural networks (RNNs) were applied to take full advantage of multivariate time-series data. In previous data-driven studies of strip breakage, it is often not clear why certain features are chosen, or from which facet should we select features. In this work, three breakage-centric feature sets are identified from three facets: physics-based approaches, empirical knowledge and data-driven features. Furthermore, in the actual production of cold-rolled strip, the steel strip shifts rapidly in the mill, where the rolling condition can change within milliseconds. The time-series process data of cold rolling is collected in a run-to-failure manner to match the temporal characteristic of this instantaneous production failure. A sliding window strategy is applied to segment and annotate whether a strip will break within the next time window (denoted as predicted window).

In addition, considering cold rolling process data may be characterised as multivariate time-series, deep learning architectures may be applied because of their robust capability to manipulate multivariate time-series data compared with more conventional approaches [14]. Among various deep learning architectures, recurrent neural networks (RNNs) retain the recent memories of input patterns, which makes them suitable for time-series processing [15]. Notably, as a variant of RNNs, the long short-term memory (LSTM) network is capable of capturing long term memories due to its fully-trained recurrent models with adaptive gates [16]. Inspired by the studies cited above, to discover the underlying relationship between real-time measured rolling variables and omens of strip breakage, an approach based on RNNs is proposed for the modelling of strip breakage.

To the best of our knowledge, the questions previously raised have not been investigated in any previous strip breakage studies, yet their answers might provide significant benefits in terms of decision-making for the occurrence of strip breakage. In actual cold rolling practice, if such a prediction can be made on a micro-level [17] with an adequately predicted window, a planned stop action can be taken to the mill in advance instead of a passive fast stop which will often result in severe damage to equipment.

The remainder of this paper is structured as follows. In Section 2, a review of relevant studies of the cold rolling process, strip breakage analyses, and sequential pattern mining is provided. Section 3 outlines the flowchart of the proposed methodology for strip breakage modelling using three facets. Section 4 reports an experimental study using real-world cold rolling data to demonstrate the effectiveness of the method, followed by a result analysis and discussion. Section 5 concludes the paper.

2. LITERATURE REVIEW

2.1 Cold Rolling Process and its Mechanical Models

An essential process in the metal processing of sheets and strips, cold rolling is widely applied due to its high accuracy, efficiency and production rate. Cold rolling can be conducted using a single stand in a reverse manner or continuous stands [1]. One of the primary processes in electrical steel strip production, cold rolling enhances strip properties by changing the microstructure and thickness of the steel. These enhanced properties include surface smoothness, tensile strength, yield strength and hardness [10].

Analysis of rolling is dated back to the pioneering work of Orowan [18], who developed a comprehensive theory based on an extension of the slab method by introducing non-homogeneity of plastic deformation of the sheet and elastic deformation of the rolls. Sims [19] developed analytical expressions of pressure distribution, roll force and roll torque by avoiding most of the numerical integration in Orowan's theory.

Various mechanical models regarding the cold rolling process have been developed and presented. These models generally consist of rolling parameters such as tension, roll force, torque and yield strength of the strip as well as several operating parameters. The model developed by Orowan was one of the most comprehensive among these cold rolling process models [18]. However, these conventional rolling force formulas only provided not more than reasonably accurate approximations. By contrast, mathematical modelling of the cold rolling process is desirable and conducive to actual rolling operation practices [20]. Nevertheless, in these theoretical analyses of the rolling process, enormous factors such as friction conditions, roll flattening, deflection of the rolling mill and temperature make these approaches problematic and time-consuming. Since exact values of these variables cannot be measured during the rolling process, other parameters are needed for better accuracy of the mathematical models [21]. As stated above, complex and various operation conditions limit the construction of physical models and make the modelling of the cold rolling process difficult.

2.2 Strip Breakages in the Cold Rolling Process

Similarly to metal forming processes, cold rolling can encounter certain defects with regard to the final product. According to technical reports, common defects in the sheet metal rolling process are edge cracking, central burst, surface defects and buckling of the strip. Among these defects, strip breakage requires special consideration because it not only significantly increases production costs but can also cause serious damage to rolls and mill accessories [22]. As one of the most common production failures in the cold rolling process, much research has been conducted on strip breakage.

Research on strip breakage has typically been conducted in a retrospective manner which focuses on root cause analysis. The causes of strip breakage in the cold rolling process have been proved to be various and have been thoroughly analysed [2, 3, 5, 7, 8, 10, 11, 23-25]. From these works, the causes of strip breakage can be classified into two categories. One category is general causes which are theoretically supported, such as mechanical defects or rolling chatter. The other is specific causes which are on a case by case basis, such as equipment malfunction, improper operation or rolling schedule adjustment.

For the general causes of strip breakage in the cold rolling process, chatters is a primary cause. Rolling and other metal forming processes are subject to the occurrence of a variety of self-excited periodic vibrations which are collectively referred to as 'chatter' [23]. Chatter usually reaches its maximum amplitude within a few seconds or even milliseconds and causes undesirable periodic variations in product gauge and surface finish. Under extreme conditions of chatter, strip breakage or damage to the rolling mill may occur [24].

Among the specific breakage causes, the equipment factor has been frequently analysed. In one case study [7], a servo-valve malfunction resulted in high-pressure fluctuation leading to inter-frame tension deviations and crushing the strip on one side. Other equipment malfunctions such as the piston rod protrusion of HGC (hydraulic gap control) and poor tension meter detection accuracy have also been discussed [2]. Apart from equipment factors, inappropriate operation parameter settings also account for the occurrence of strip snap. Several operating parameters related to the working roll, such as diameter disparity between the top and bottom working rolls, levelness of the bottom working roll and convexity degree of the working rolls, have been discussed as significant strip breakage causes. In addition, working rolls, levelness and perpendicular of the deflector rolls have also been proved to generate strip breakages [3].

2.3 Sequential Pattern Mining and Pattern-Based Classification

Sequential pattern mining is a widely researched topic. It refers to the mining of frequently occurring events in order as patterns. It was first introduced [26] in the Apriori family of algorithms. The algorithms perform pattern mining in sequences of itemsets (events) and find frequent patterns in the input. A large variety of algorithms, similar to the Apriori algorithms, have been introduced, such as Sequential PAttern Discovery using Equivalence classes (SPADE) [27], PrefixSpan [28] and Sequential PAttern Mining (SPAM) [29]. These algorithms all use the support measure to determine frequency. The support of a sequence is simply the proportion of entries in the data that it appears in frequency. The ability to address the complex data structure of sequences is what sets sequential pattern mining apart from standard data mining [30]. Sequential pattern mining can access and obtain information that may be hidden in the structure of a sequence. The collective behaviour and hidden relations between such data can contain decisive information [31].

As was reviewed in Section 2.2, with severe thickness variations of the strip and noticeable tension fluctuations, the chatter will typically result in unexpected strip breakage [23]. Since this chatter incident is a sudden, self-excited periodic vibration of specific process variables, inspired by works cited in Section 2.3, sequential pattern mining is considered to capture patterns of chatters in the cold rolling process.

3. METHODOLOGY

In this paper, we propose a machine learning-based approach that consists of three main stages; the flowchart of the proposed approach is shown in Figure 1. We start with a data collection and pre-processing step, where data is extracted from a cold-rolled steel strip plant. In stage 2, we examine three different facets to generate candidate sets of features. Combinations of these three feature sets are constructed to present different scenarios for strip breakage modelling. In stage 3, a sequence-to-vector RNN architecture is applied for modelling and evaluating the predictive performance of the scenarios.

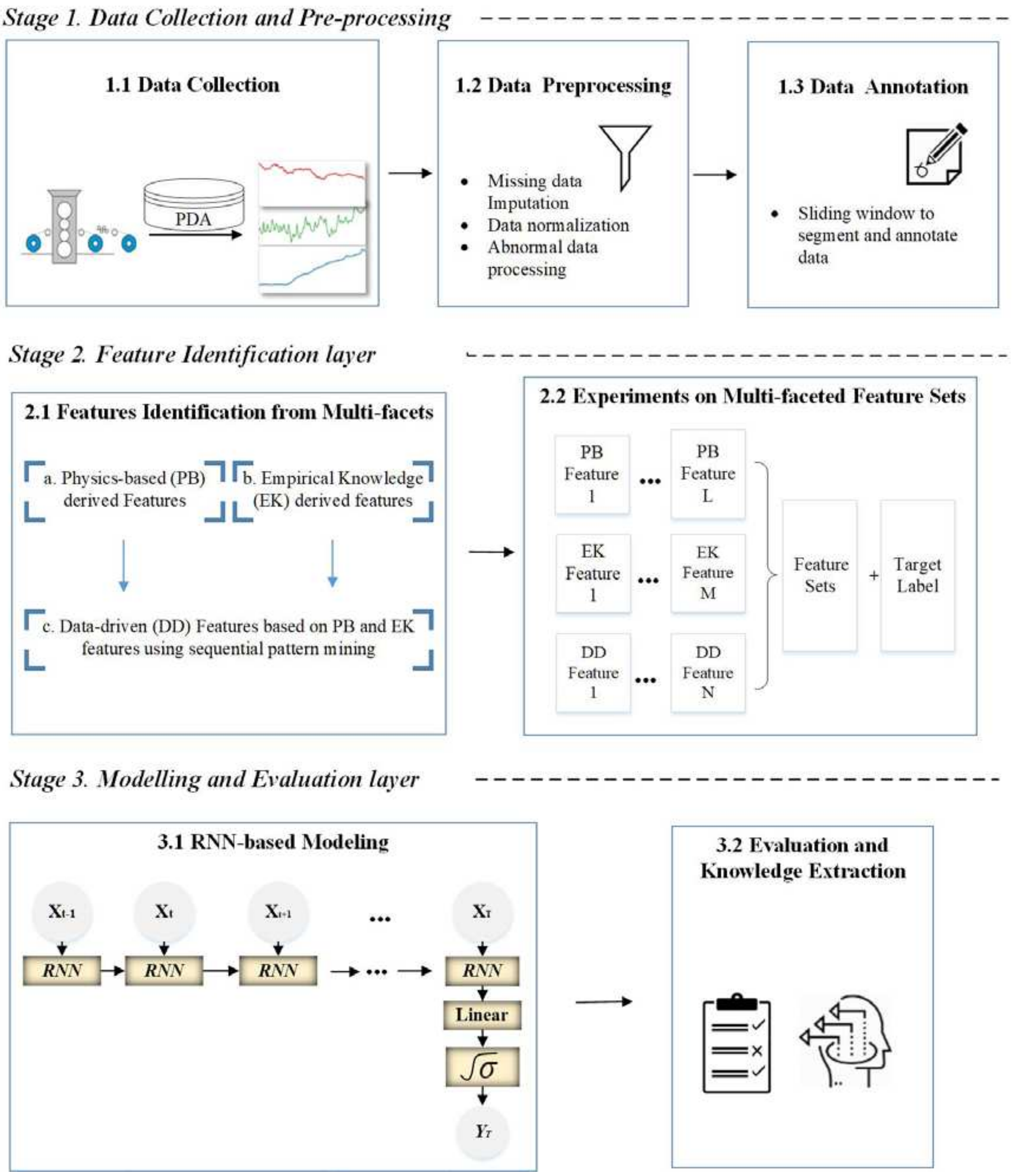

Figure 1. Flowchart of multi-faceted strip breakage modelling

3.1 Data Pre-processing and Time Window Processing

The collection of cold rolling process monitoring data was first conducted. Since strip breakage is an incident that occurs instantaneously, temporal observations that extend far from the breakage point into the past are likely to have lower support for breakage modelling. To collect informative and predictive time-series data, we should incorporate the concept of recency to breakage in the collection of process monitoring data [32]. In this context, data was collected in a run-to-failure manner, from the strip breakage time point backwards in time to obtain the observations most recent to breakage. Due to the high correlation of neighbouring data, a sliding window strategy was adopted to segment the raw time-series data into a collection of pieces; an illustration of this strategy is shown in Figure 2. This strategy is applied to capture the momentary variations before strip breakage. In addition, we can take better advantage of time-series data since a time window conveys more information than a single time point.

In this sliding window strategy, an instance is a two-dimensional matrix containing $\boldsymbol{r}$ sampling points (i.e. the window length) with $\boldsymbol{N}$ attributes. By sliding the window backwards in time from the breakage point following a selected step size, the total window length is segmented into $\boldsymbol{M}$ instances. For both training and testing data, the label of each instance is determined by the interval between the last sample point and the strip breakage point. If the interval is wider than the predefined predicted window length, the corresponding instance is labelled as 0 representing no breakage. Otherwise, if the interval is within the preset predicted window length, the corresponding matrix instance is labelled as 1, which represents the coming breakage. In this manner, the label of the classification task is a binary number representing whether the strip will break within a specific predicted time window.

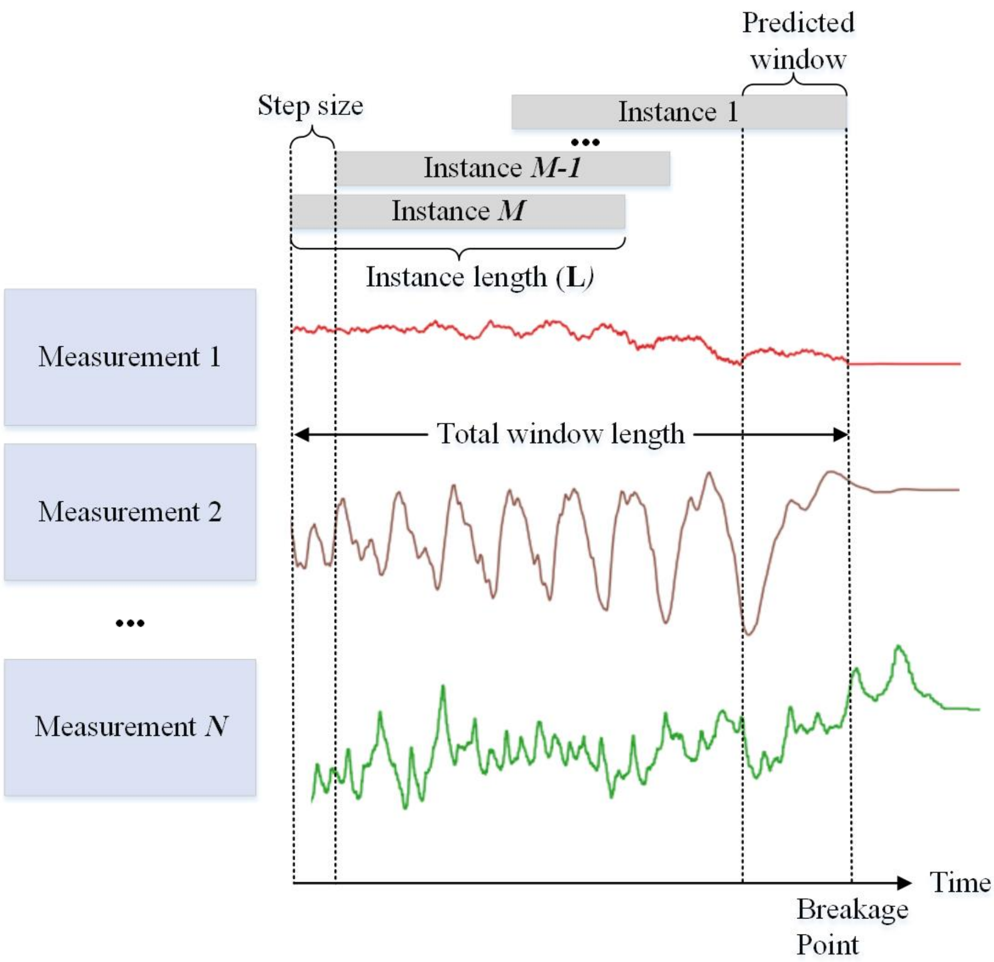

Figure 2. Illustration of the proposed sliding window strategy

3.2 Breakage Relevant Feature Identification from Multi-facets

Because there are typically thousands of measurements being taken throughout the cold rolling process, it is necessary to select or construct a subset of the most relevant features. Considering the complexity of strip breakage causes, we are not sure whether the algorithms can make use of all the features. When choosing features, the domain of the text in the dataset is also of significant influence. Certain features work better either for a specific domain or in a non-domain dependent dataset [33]. We choose to determine candidate feature sets from the following three facets:

- The **physics-based** (PB) feature set contains features directly derived from previous physics-based models of strip breakage failure. This facet is selected to capture the general causes of strip breakage, as was reviewed in Section 2.2.
- The **empirical knowledge** (EK) feature set is a feature set that would capture specific relevant and discriminative data by looking at the informative factors that result in a strip breakage and referring to domain experts. This facet is selected to capture the specific causes of strip breakage, as was reviewed in Section 2.3, since these causes vary from mill to mill.
- The **data-derived** (DD) features are binary features derived from sequential patterns based on PB and EK features.

3.3 Multi-Faceted Modelling for Strip Breakage Using Machine Learning

3.3.1 Modelling procedure

Regarding the prediction task for whether a strip will break within a predicted window, a supervised machine learning approach based on sequence-to-vector RNN architecture is proposed. The task is set as a binary classification problem to classify whether a strip will break within a specific time window. To be specific, an instance entering the proposed RNN architecture is a two-dimensional matrix containing r sampling points (i.e. the window length), with N attributes.

Unlike a standard neuron network, RNN consists of a series of recurrent neurons, and the output from a recurrent neuron is connected to the next recurrent neuron, as shown in Figure 3. One issue associated with the standard RNN is the 'fading memory' problem. Once the number of time steps becomes large, the 'future' time steps will contain virtually no memory of the first inputs, as there is no structure in the standard recurrent layer that individually controls the flow of the memory itself. To solve this problem, the LSTM network, a family of the recurrent cell which incorporates the standard recurrent layer along with additional 'memory' control gates, has been proposed [34]. An LSTM network is formed exactly like a simple RNN except that memory blocks replace the nonlinear units in the hidden layer. Indeed, LSTM blocks may be mixed with simple units if required — although it is typically not necessary. Also, as with other RNNs, the hidden layer can be attached to any differentiable output layer, depending on whether the required task is a regression or classification [35].

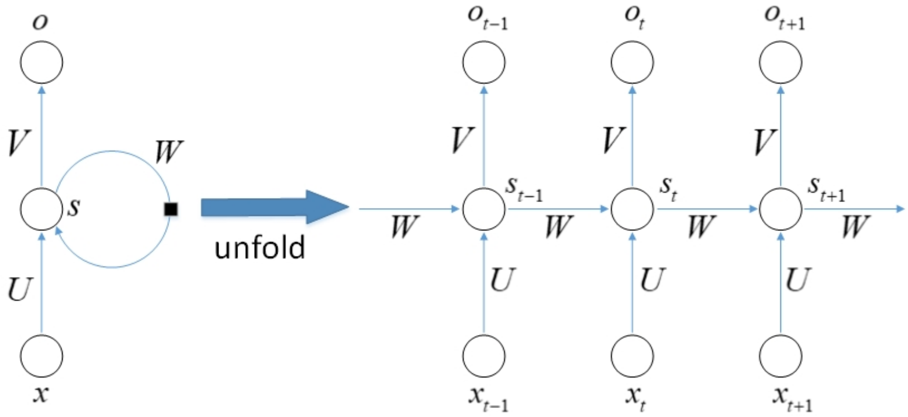

Figure 3. RNN structure illustration

The proposed RNN is applied to multivariate time-series classification as follows: input data for a time slice represented as a matrix instance ($r \times N$) is fed into the recurrent layers, and only the output of the last neuron is fed into the linear layer (the rest are ignored). This output is subsequently fed into a linear layer which is embedded with an activation function to make binary predictions.

3.3.2 Evaluation metrics

The area under the receiver operating characteristic curve (AUC) is selected to evaluate the performance of the proposed modelling methodology since AUC is suitable for binary classification. Conventionally, accuracy rate is a standard metric for binary classification. Nevertheless, regarding the unbalanced dataset in this study, accuracy rate is not an appropriate index since it does not distinguish between numbers of correctly classified instances of different classes [36]. To evaluate the classification performance of positive and negative classes independently, we can obtain the following two metrics from the confusion matrix [37]:

True positive rate (TPR) is the percentage of positive instances correctly classified:

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

True negative rate (TNR) is the percentage of negative instances correctly classified:

$$TNR = \frac{TN}{FP+TN} \tag{2}$$

However, neither of these measures is adequate for evaluating the classification performance since classification intends to achieve good results for both positive and negative classes. In this case, the receiver operating characteristic (ROC) graph offer a means to combine these measures to produce an evaluation measurement [38]. The AUC provides a single measure of a classifier's performance for the evaluation regarding which model is on average better. Regarding the unbalanced characterisation of the dataset, AUC is used in this study as the metric for the classification of performance evaluation [39].

4 EXPERIMENTAL STUDY

The experimental study was conducted using the historical data provided by a cold-rolled silicon electrical steel manufacturer. Due to excellent electrical and magnetic properties, cold-rolled silicon steel strip is a primary functional material widely used for the manufacturing of transformer cores and motors in the power electronics industry [40]. For the production of silicon steel strips, cold rolling is an essential process which compresses and squeezes the incoming strip into the thinner outcoming strip to enhance properties such as yield strength, surface smoothness and permeability. Compared with general low carbon steel strips, silicon steel strips are lower in toughness due to a high silicon content [41]. In this context, strip breakages therefore more frequently occur in the cold rolling process of high silicon steel strips.

In this steel strip manufacturer, brittle high silicon electrical steel strips are cold-rolled five passes back and forth, decreasing the original thickness by 90% on a reversing mill where undesired strip breakages occur from time to time. Breakages of strip coils result in yield loss, reduced speed of work and failure to achieve the final target thickness. Consequently, strip breakage production failure increasingly aroused attention from this steel company. Furthermore, in this company, causes of strip breakage are currently identified retrospectively rather than in a predictive manner. The company can benefit from an effective strip breakage prediction model by taking countermeasures beforehand to increase productivity and prevent further damage.

An initial experimental study was conducted to predict strip breakage and gain insights into feature sets constructed from different facets. In addition to experiments on different feature sets, to discover the appropriate predicted window length for the evaluation of prediction performance and actual production practice, experiments exploring different predicted window lengths were also conducted.

4.1 Data Acquisition and Preparation

The raw data regarding the cold rolling process in this study were extracted from a production data acquisition (PDA) system installed on-site. Cold rolling process variables such as speed, tension, eccentricity and roll gap position can be measured by this automated system, which is employed with accurate measurement devices. Data are monitored and recorded continuously in real-time at a frequency of 100 Hz to record the continuous working condition of the mill. Due to the high correlation between neighbouring data points, a lower sampling rate results in a distortion, compared with raw data. Therefore, to get the most information possible from this PDA-recorded data, we used full resolution data under the sampling rate of 100 Hz. Additionally, using full-resolution data enables a detailed calculation of the breakage point, resulting in more accurate labels for classification.

Since the causes of strip breakage are currently identified retrospectively in this company, each broken strip coil is marked with a specific cause manually by shop floor engineers. These causes are generally classified into material causes, non-material causes and unknown causes. It should be noted that, as was reviewed in Section 2.1, a primary cause of strip breakage is the issue of incoming raw material which is annealed and pickled hot-rolled strips. The case is similar in this company as well. However, the information regarding these hot-rolling strips was collected in a batch level (i.e. the measurements were taken on each coil) in practical cold rolling operations, and no detailed material data were collected at a second level. Since the objective of the experimental study is to predict strip breakage at a micro-level, the material issue was not in its scope. In this context, data were obtained from coils with a unified material grade.

Pre-processing of data was conducted concerning the momentary manner of strip breakage with the aim of taking better advantage of time-series data. To be specific, the collection and pre-processing of the process data were conducted from the strip breakage point backwards in time. Following this, data cleaning was conducted to deal with abnormal and missing data. There were abnormal negative values in variables such as entry and exit speed, indicating the rolling direction (since the mill was reversing), and the absolute value was therefore taken. Furthermore, values were missing within the dataset. In consideration of the correlation of neighbouring data points, forward imputation was applied, which imputed any missing value to be the same as its previous measurement. In this context, data were collected from the strip coils broken due to non-material causes, and these coils possessed the same incoming material grade. It should also be noted that for each selected coil only broke once. Thus, the dataset contained 33 broken strip coils marked with non-material causes under the same incoming material grade, covering three months of production. The dataset was divided into training and testing sets before training. The training set consisted of 27 coils, and the test set contained 6 coils. In consideration of obtaining a manageable dataset size while at the same time considering recency, the parameter set of the sliding strategy was determined as shown in Table 1.

Table 1. Parameters in sliding window strategy (unit: second).

Instance length	Step size	Predicted window length	Time backwards from breakage point
58	0.01	0.5	60

To be specific, under this parameter setting, each coil in the training set could generate 201 instances; 50 of the 201 were marked as break and the remainder were labelled non-break.

4.2 Experimental Setup

In order to gain insights into predictive performance among feature sets, the first experimental study was a performance comparison of models based on different combinations of feature sets identified from multiple facets. Subsequently, based on the results from the first experimental study, further exploratory experiments were conducted to discover an appropriate predicted time window length.

In the first experimental study, models were built based on feature sets identified from EK, physical-based models and DD approaches. First, the feature set derived from EK was created since these features are informative and include specific strip breakage causes in the cold rolling system of this company. Details of this feature set are listed in Table 2.

Table 2. Details of features relevant to strip breakage based on empirical knowledge.

No.	Name	Description
1	Raw entry speed (m/min)	Strip speed measured on the entry side
2	Raw exit Speed (m/min)	Strip speed measured on the exit side
3	Total load feedback	The force pushing the load apart, equal to the pressure on the strip
4	Front capsule force	Force applied on the front capsule

5	Back capsule force	Force applied on the back capsule
6	LR tension	A force applied to the pull strip from the side of the left reel into the rolls
7	RR tension	A force applied to the pull strip from the side of the right reel into the rolls
8	Exit gauge deviation (mm)	Strip thickness deviation measured on the exit side
9	Raw gap position (mm)	A separation distance of work rolls with no elastic deformation
10	Eccentricity trim (mm)	A periodic trim to handle the non-circularity of the rolls which result in periodic variation in the roll gap
11	Measured slip (%)	The displacement ratio between the strip coil and the working roll

Second, in addition to the features identified from EK which cover specific strip breakage causes as reviewed in Section 2.2, features identified from PB models were considered as the second facet to be included in general causes of strip breakage. Diameter disparity between the top and bottom working rolls as well as left and right deflector roll were first considered [3]. Since chatter is a vital aspect of strip breakage, the causes of chatter were also considered. As chatter was proved to be the manifestation of torsional vibrations of the roll-spindle shaft system [42], the frequency of vertical and torsional vibrations of work roll and spindle are typically considered in chatter modelling. As there were only data measuring the working roll, the frequency was derived by taking the spectrum of the work roll position signal. In this context, six PB features were constructed for further experiments. Details of the features constructed from physical-based models for strip breakage analysis are listed in Table 3.

Table 3. Details of features constructed from physical-based models.

No.	Name	Description
1	Diameter disparity of work roll (%)	Top working roll diameter divided by bottom working roll
2	Diameter disparity of deflector roll (%)	Left deflector roll diameter divided by right deflector roll
3	Vertical vibration frequency of top work rolls (Hz)	The spectrum of the top work roll position signal
4	Vertical vibration frequency of bottom work roll (Hz)	The spectrum of the bottom work roll position signal
5	Work roll location error	Mean error of location of rolls
6	Angular velocity error	Mean error of angular velocity of rolls

Third, as a main non-material cause of strip breakage, chatter can cause drastic variation in primary variables [43]. It would be beneficial to discover the sequential variation pattern regarding the selected primary features and apply these patterns as features for the prediction task.

In this context, sequential pattern mining was conducted on the 17 selected EK and PB features listed in Table 2 and 3. First, to handle the complex time-series, temporal abstraction [44] was conducted to transform the numeric time-series variables into a qualitative high-level form. To be specific, each primary variable was converted into an interval-based representation. When the mill encounters a chatter, the values of certain primary features fluctuate remarkably. Under this condition, each primary feature was segmented using 10th, 25th, 75th and 90th percentiles of the numeric values. Five states were defined as Very Low (VL), Low (L), Normal (N), High (H) and Very High (VH). For instance, a value between the 10th and 25th percentiles was segmented as low, and a value above the 90th was very high. Following the temporal abstraction of the global one-minute data for each coil, PrefixSpan [45] sequential pattern mining algorithms were applied on the selected 27 broken coils in the training dataset. An abstracted state corresponds to an event (itemset) within a sequence. To be specific, the abstracted state for each primary feature consisted of a sequence within a broken coil, and the sequential pattern mining was then conducted for all 27 coils. The sequential pattern mining was conducted using the open-source Java-based package SPMF [46]. For the PrefixSpan algorithm, the Min support was set to 0.4 and Max pattern length to 10. The pattern with the highest support was chosen as the most frequent sequential pattern of strip breakage for this primary feature.

Finally, for an instance which represented a 58-second time window, the same abstraction strategy was adopted. The abstracted instance was matched with 17 sequential patterns generated from 17 selected primary features. Then, 17 binary features were constructed in each instance. The value of each binary feature depended on the matching between an abstracted primary feature and its corresponding pattern. Within an instance, if an abstracted primary feature contained its following sequential pattern, the corresponding binary feature was marked as 1, and vice versa. In this manner, regarding the primary feature set, the 5800×17 instance was transformed into 5800×34 after feature construction.

In terms of the deep learning architecture, the main deep learning model parameters consisted of the type of layer, number of layers, number of nodes in each layer and dropout rate. After several trials, a pyramid shape network structure was designed in accordance with both computation cost and classification performance to balance the trade-off between computation cost and model depth. Rectified linear unit (ReLU) was selected as the activation function. Detailed information for the proposed network is shown in Table 4.

Table 4. Detailed information for each layer of the proposed network model.

Layers	Layer name	Main parameters	Other parameters
Layer 1	Embedding	N/A	N/A
Layer 2	LSTM/GRU/RNN	60 units	Dropout = 0.3
Layer 3	LSTM/GRU/RNN	30 units	Dropout = 0.3
Layer 4	Flatten	N/A	N/A
Layer 5	Fully-connected	30 units	Activation = ReLU
Layer 6	Fully-connected	2 units	NA

To be specific, since this was a binary classification problem, CrossEntropy was used as the loss function. The efficient Adam algorithm [47] was used for optimisation. The model was fit for 100 epochs because it quickly overfits the problem. A batch size of 60 was used to space out weight updates. Python was the utilised platform, and the deep learning models were built using Pytorch [48]. Details of the parameters for the training process are specified in Table 5.

Table 5. Parameters of the training process of strip breakage deep learning prediction model.

Learning rate	Batch size	Epochs	Activation function	Optimiser	Loss function
0.001	60	100	ReLU	Adam	CrossEntropy

A benchmark test was conducted to compare five prevailing machine learning algorithms: random forest (RF) [49, 50], support vector classification (SVC) [51], artificial neuron network (ANN), RNN and gated recurrent unit (GRU). For the conventional RF, SVC and NN algorithms, which are unable to handle the high dimensionality of time-series data directly, hand-crafted features were required, and feature extractions were consequently applied. Six types of features, including time domain and frequency domain, were designed and crafted, as shown in Table 6. The benchmark tests of the conventional algorithms were carried out using the open-source package Scikit-learn [52] with the default hyperparameters. For RF, the number of trees in the forest was set as 100, and the number of features to consider when looking for the best split was set as the square root of the number of input features. For SVC, radial basis function was used as the kernel type, the degree of the polynomial kernel function was set as 3 and the kernel coefficient for radial basis function was set as the reciprocal of the number of features. For ANN, two hidden layers were designed with 100 neurons in each layer; the other parameters were set as with LSTM, as shown in Table 5. In terms of the RNN and GRU network, the input data can be tensor, as with an LSTM network. Therefore, architecture and training parameters for the RNN and GRU network were set to be identical to the LSTM network.

Table 6. List of extracted features.

Domain	Features
Statistical	Root mean square
	Variance
	Maximum
	Peak-to-peak
Frequency	Spectral skewness
	Spectral kurtosis

A graphics processing unit (GPU) was used for the experiment to increase speed and decrease training time. More specifically, the processing system used for the analysis was as follows: CPU Core i7-9700 K 3.8 GHz with 32 GB RAM and GPU NVIDIA GeForce 2080ti.

4.3 Results and Discussion

4.3.1 Experiments using multi-faceted feature sets

Based on the experimental setup, which explored different combination scenarios of the feature sets identified from different facets, the quantitative results of the predictive models are presented in Table 7 under the metrics of AUC evaluated using the test dataset.

Table 7. AUCs of models with the best performance in predicting strip breakage using different algorithms and feature sets.

	Feature sets	Conventional			RNN-based		
		RF	SVM	ANN	RNN	GRU	LSTM
Single feature set	EK	0.519	0.563	0.679	0.772	0.788	**0.820**
	PB	0.507	0.554	0.665	0.749	0.735	**0.807**
Multiple feature sets	EK + DD	0.518	0.561	0.697	0.751	0.792	**0.821**
	PB + DD	0.506	0.552	0.682	0.749	0.787	**0.809**
	EK+PB	0.515	0.559	0.698	0.736	0.757	**0.811**
	EK + PB + DD	0.511	0.561	0.693	0.779	0.801	**0.835**

In Table 7, the performance of various models in which both RNN-based and conventional approaches were applied with different mixes of feature sets is displayed. Generally, due to the default setting of hyperparameter selection and different manner of data representation, the improvement of RNN-based deep learning models compared with traditional methods is enormous. However, as a result of model complexity, hyperparameter selection is required to achieve the desired performance.

Among the models built on primary features, the best performance was obtained when the LSTM network was applied. The RNN-based algorithms outperformed conventional algorithms when only single feature sets were being applied. This may result from hand-crafted features used to represent the raw time-series data limiting modelling performance [53]. Among the models built on multiple feature sets, the LSTM network still outperformed other algorithms. Indeed, as a result of the inclusion of more features, the overall performance of various conventional algorithms improved.

Among the models built on the proposed LSTM network, the best performance was achieved when features from all three sets were analysed together. An analysis of the performance of the best model on the test set for the different feature sets indicates that the network is able to make use of extra features. The performance advantage of the inclusion of the KB feature sets for the primary feature set was substantial. The advantage of the inclusion of SPM feature sets for the primary feature set was relatively subtle.

4.3.2 Experiments using different time window lengths

The above experimental studies were to train a model to predict whether a strip would break within the next 0.5 seconds. The 0.5-second window was suggested by the steel plant since they consider it sufficient time to respond. An adequate predicted window can provide enough time to take countermeasures before a strip breakage occurs. For instance, operators can take contingency mitigation countermeasures such as an actively planned stop rather than a passive fast stop, which will often result in severe damage to the rolling mill. However, as a rule of thumb, a wider predicted window often leads to undesirable prediction accuracy. Based on the best model in Section 4.2.1, to gain insights on the trade-off between predicted time window length and prediction performance, the following experiments were designed to explore window sizes from 0.1 to 0.9 seconds.

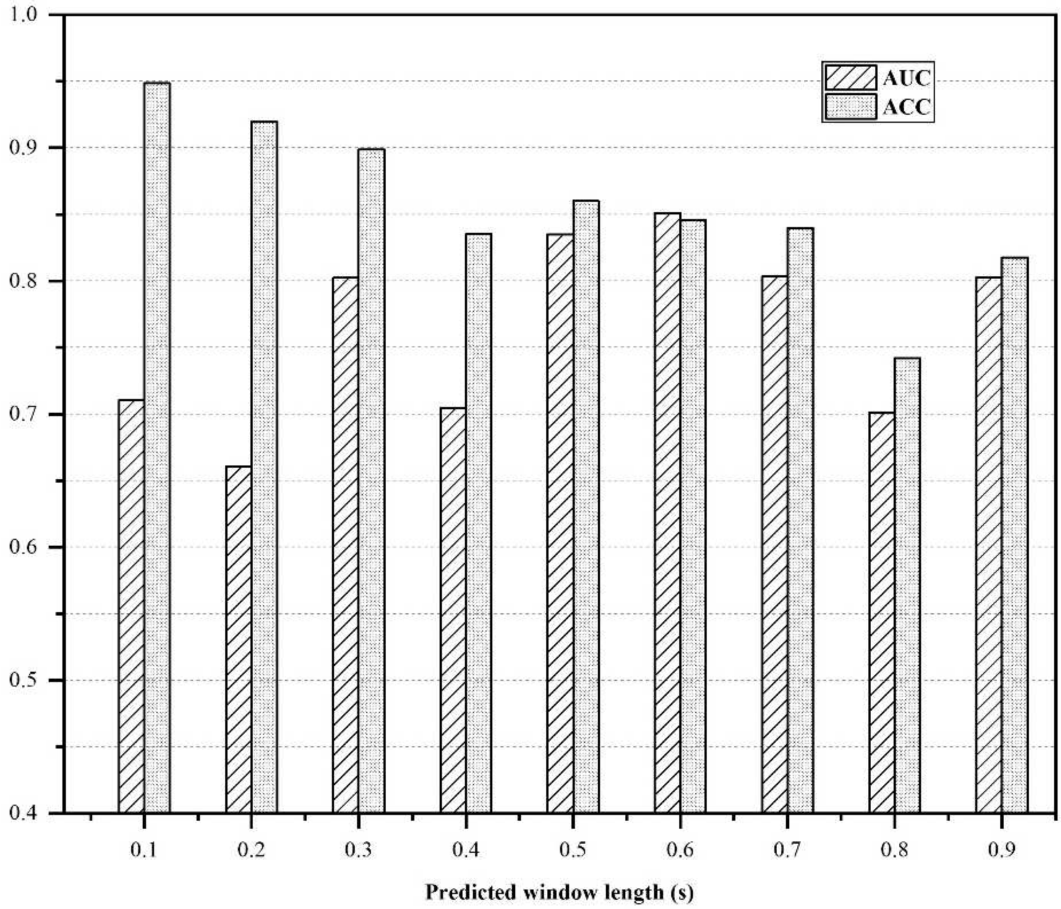

Figure 4. Algorithm performance of models based on different predicted time windows in terms of ACC and AUC.

Based on different window lengths, performance was compared when the algorithm converged. The relationship between the algorithm performance and different predicted window lengths in terms of AUC and ACC are shown in Figure 4. It can be seen that ACC generally decreases with the increment of predicted window length, which conforms with our assumption of the relationship between prediction accuracy and predicted time window. However, according to our proposed sliding window strategy, the selection of a predicted window length will affect the data balance. Even if better performance in terms of ACC is achieved when the window length is narrow, the best AUC is achieved with a predicted window of 0.6 seconds.

5. CONCLUSIONS AND FUTURE WORK

Strip breakage is a severe production failure which occurs instantaneously in the cold rolling process. Prediction of this failure can bring significant benefits to the cold rolling industry in terms of contingency mitigation and quality improvement. In the present study, to minimise the occurrence and impact of strip breakage, we achieved a micro-level prediction of strip breakage based on historical process data. The first contribution of this paper is its exploration of deep learning models applied to a cold rolling process at an event level as compared to a batch level regarding strip breakage failure. The cold rolling mill operator can benefit from utilising this prediction approach in developing their contingency mitigation strategies. According to the predicted information, a planned stop action can be taken to avoid damage from an unplanned fast stop. Understanding the likelihood of strip breakage in the near future can also be vital for post-analysis, such as in determining what countermeasures should be used.

For real cold rolling practice, even if we considered all the causes of strip breakage beforehand, occurrence of this failure may not always be avoided. This limitation is due to information such as unexpected sudden changes, an undetected internal material defect or, like most cases, from an unknown reason not conveyed in the current dataset. Therefore, this approach is more practical for breakages with a divinable manifestation in rolling process variables, such as breakages caused by chatter.

In further work, first, the algorithm performance in terms of ACC and AUC can be continuously improved. Furthermore, with more studies on strip breakage cause analysis and further domain expert assistance, future work would include more domain-based features to expand the scope of this proposed multi-faceted approach. Finally, the data collected in this study were under the same material grade. In this context, strip breakages caused by material defect, which is another critical issue for strip breakage, was not within the scope of this work. Therefore, to improve the breadth of collected information regarding strip breakage, data recorded about this production failure from different sources, such as material data, need to be incorporated to generate collective values.

REFERENCES

[1] M. Mashayekhi, N. Torabian, and M. Poursina, "Continuum damage mechanics analysis of strip tearing in a tandem cold rolling process," *Simulation Modelling Practice and Theory,* vol. 19, no. 2, pp. 612-625, 2011.

[2] K. Yan and H. Li, "Causes and Countermeasures of Cold Strip Break in Masteel," *Anhui Metallurgy,* no. 4, pp. 36-37, 2006.

[3] X. Cui and A. Zhao, "Analysis the Causes of Deviation Breaking of Cold Rolling Strip," *Xinjiang Steel,* no. 02, pp. 44-47, 2013.

[4] D. Bhattacharya, A. Mishra, G. P. Poddar, and S. Misra, "Case study of severe strip breakage in rolling mill of Thin Slab Casting and Rolling (TSCR) shop of TATA Steel, Jamshedpur," *Case Studies in Engineering Failure Analysis,* vol. 5, pp. 15-22, 2016.

[5] Z. Chen, Y. Liu, A. Valera-Medina, and F. Robinson, "Characterizing Strip Snap in Cold Rolling Process Using Advanced Data Analytics," *Procedia CIRP,* vol. 81, pp. 453-458, 2019.

[6] J. Xu, "Cause Analysis for Strip-breaking of Non-oriented Silicon Steel," *JIANGXI METALLURGY,* no. 05, pp. 26-28, 2015.

[7] B. Liu, "Study and application of decoupling control of roll tilt and rolling force difference to inhibit strip break during cold rolling," *Metallurgical Industry Automation,* no. 06, pp. 40-44, 2015.

[8] B. Liu, C. L. Tang, and T. Q. Gu, "A Method to Avoid Strip Breakage for Thin Strip Steel in Cold Rolling," in *Advanced Materials Research*, 2014, vol. 1004, pp. 1211-1215: Trans Tech Publ.

[9] Y.-H. Kuo and A. Kusiak, "From data to big data in production research: the past and future trends," *International Journal of Production Research,* vol. 57, no. 15-16, pp. 4828-4853, 2019.

[10] K. Takami, J. Mahmoudi, E. Dahlquist, and M. Lindenmo, "Multivariable data analysis of a cold rolling control system to minimise defects," *The International Journal of Advanced Manufacturing Technology,* vol. 54, no. 5, pp. 553-565, 2011.

[11] C. Wang, "Research on Fault Diagnosis of Belt Tearing in the Cold Rolling Process," *Instrumentation Technology,* no. 09, pp. 16-20, 2014.

[12] C. Da Cunha, B. Agard, and A. Kusiak, "Data mining for improvement of product quality," *International journal of production research,* vol. 44, no. 18-19, pp. 4027-4041, 2006.

[13] M. R. Lopes, A. Costigliola, R. Pinto, S. Vieira, and J. M. Sousa, "Pharmaceutical quality control laboratory digital twin–A novel governance model for resource planning and scheduling," *International Journal of Production Research,* pp. 1-15, 2019.

[14] J. C. B. Gamboa, "Deep learning for time-series analysis," *arXiv preprint arXiv:1701.01887,* 2017.

[15] S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan, and K. Litsiou, "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail," *International Journal of Production Research,* pp. 1-16, 2020.

[16] Z. Wang, H. Ma, H. Chen, B. Yan, and X. Chu, "Performance degradation assessment of rolling bearing based on convolutional neural network and deep long-short term memory network," *International Journal of Production Research,* pp. 1-13, 2019.

[17] A. Katona *et al.*, "Time to Die: Death Prediction in Dota 2 using Deep Learning," *arXiv preprint arXiv:1906.03939,* 2019.

[18] E. Orowan, "The calculation of roll pressure in hot and cold flat rolling," *Proceedings of the Institution of Mechanical Engineers,* vol. 150, no. 1, pp. 140-167, 1943.

[19] S. Oh and S. Kobayashi, "An approximate method for a three-dimensional analysis of rolling," *International Journal of Mechanical Sciences,* vol. 17, no. 4, pp. 293-305, 1975.

[20] J. Son, D. Lee, I. Kim, and S. Choi, "A study on genetic algorithm to select architecture of a optimal neural network in the hot rolling process," *Journal of Materials Processing Technology,* vol. 153, pp. 643-648, 2004.

[21] E. J. M. Geddes and I. Postlethwaite, "Improvements in product quality in tandem cold rolling using robust multivariable control," *IEEE transactions on control systems technology,* vol. 6, no. 2, pp. 257-269, 1998.

[22] W. Johnson and A. Mamalis, "A survey of some physical defects arising in metal working processes," in *Proceedings of the Seventeenth International Machine Tool Design and Research Conference*, 1977, pp. 607-621: Springer.

[23] F. Hou, J. Zhang, J. Cao, and X. Shi, "Review of chatter studies in cold rolling," *Gangtie Yanjiu Xuebao(Journal of Iron and Steel Research),* vol. 19, no. 10, pp. 6-10, 2007.

[24] Y.-J. Lin, C. Suh, R. Langari, and S. Noah, "On the characteristics and mechanism of rolling instability and chatter," *Journal of manufacturing science and engineering,* vol. 125, no. 4, pp. 778-786, 2003.
[25] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 7, pp. 5525-5534, 2018.
[26] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," in *International Conference on Extending Database Technology*, 1996, pp. 1-17: Springer.
[27] M. J. Zaki, "SPADE: An efficient algorithm for mining frequent sequences," *Machine learning,* vol. 42, no. 1-2, pp. 31-60, 2001.
[28] J. Pei *et al.*, "Mining sequential patterns by pattern-growth: The prefixspan approach," *IEEE Transactions on knowledge and data engineering,* vol. 16, no. 11, pp. 1424-1440, 2004.
[29] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu, "Sequential pattern mining using a bitmap representation," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 429-435: ACM.
[30] H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen, and U. Dayal, "Multi-dimensional sequential pattern mining," in *Proceedings of the tenth international conference on Information and knowledge management*, 2001, pp. 81-88: ACM.
[31] E. Bautista-Thompson and R. Brito-Guevara, "Classification of data sequences by similarity analysis of recurrence plot patterns," in *2008 Seventh Mexican International Conference on Artificial Intelligence*, 2008, pp. 111-116: IEEE.
[32] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht, "Mining recent temporal patterns for event detection in multivariate time series data," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 280-288: ACM.
[33] Y. Mejova and P. Srinivasan, "Exploring feature definition and selection for sentiment classifiers," in *Fifth international AAAI conference on weblogs and social media*, 2011.
[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.
[35] A. Graves, "Supervised sequence labelling," in *Supervised sequence labelling with recurrent neural networks*: Springer, 2012, pp. 5-13.
[36] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM Sigkdd Explorations Newsletter,* vol. 6, no. 1, pp. 1-6, 2004.
[37] J. D. Storey, "The positive false discovery rate: a Bayesian interpretation and the q-value," *The Annals of Statistics,* vol. 31, no. 6, pp. 2013-2035, 2003.
[38] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition,* vol. 30, no. 7, pp. 1145-1159, 1997.
[39] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters,* vol. 27, no. 8, pp. 861-874, 2006.
[40] H. Li *et al.*, "Edge-drop control behavior for silicon strip cold rolling with a sendzimir mill," *Metals,* vol. 8, no. 10, p. 783, 2018.

[41] Y. Yan, Q. Sun, J. Chen, and H. Pan, "Effect of processing parameters on edge cracking in cold rolling," *Materials and Manufacturing Processes,* vol. 30, no. 10, pp. 1174-1178, 2015.

[42] V. Panjković, R. Gloss, J. Steward, S. Dilks, R. Steward, and G. Fraser, "Causes of chatter in a hot strip mill: observations, qualitative analyses and mathematical modelling," *Journal of Materials Processing Technology,* vol. 212, no. 4, pp. 954-961, 2012.

[43] P.-H. Hu, H. Zhao, and K. Ehmann, "Third-octave-mode chatter in rolling. Part 3: Stability of a multi-stand mill," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* vol. 220, no. 8, pp. 1293-1303, 2006.

[44] Y. Shahar, "A framework for knowledge-based temporal abstraction," *Artificial intelligence,* vol. 90, no. 1-2, pp. 79-133, 1997.

[45] J. Pei *et al.*, "Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings 17th international conference on data engineering*, 2001, pp. 215-224: IEEE.

[46] P. Fournier-Viger *et al.*, "The SPMF open-source data mining library version 2," in *Joint European conference on machine learning and knowledge discovery in databases*, 2016, pp. 36-40: Springer.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[48] A. Paszke *et al.*, "Automatic differentiation in pytorch," 2017.

[49] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, no. 3, pp. 18-22, 2002.

[50] S. Jun, S. Lee, and H. Chun, "Learning dispatching rules using random forest in flexible job shop scheduling problems," *International Journal of Production Research,* vol. 57, no. 10, pp. 3290-3310, 2019.

[51] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters,* vol. 9, no. 3, pp. 293-300, 1999.

[52] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of machine learning research,* vol. 12, no. Oct, pp. 2825-2830, 2011.

[53] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, no. 8, pp. 1798-1828, 2013.