

# Strip Snap Analytics in Cold Rolling Process Using Machine Learning

Zheyuan Chen, Ying Liu, Agustin Valera-Medina and Fiona Robinson

**Abstract—** Strip snap, also known as strip breakage or belt tearing, is an undesirable quality incident which results in yield loss and reduced work speed in the cold rolling process of strip products. Therefore, it is necessary to reveal a functional relationship between certain selected variables and strip snap event for the aim of quality improvement. In this study, the probability of strip snap occurrence was quantified by a selected measured variable. Several machine learning algorithms were adopted to predict this target probability. To validate this approach, a case study was conducted based on real-world data collected from an electrical steel reversing mill. The excessively good performance indicates several variables which are strongly correlated with the target.

## I. INTRODUCTION

Assumed a plain-strain deformation process, cold rolling compresses and squeezes incoming strip feedstock into thinner strips between the working rolls. With the rapid development of cold rolling processes, a wide variety of sensors are deployed which enable opportunities for quality improvement through data analytics under this data-rich environment. For the cold rolling process of High Silicon strip products, a typical incident of strip snaps frequently occurs. Strip snap, also known as strip breakage or strip tearing, is one of the most common quality issues in the cold rolling process [1]. This incident results in damage on rolls, the steel strip and loss of yield. Therefore, research to identify and determine the causes of strip snap is of great significance in production yield improvement, cost reduction and mill service life extension [2].

The causes of strip snaps have previously been studied in a handful of works [3-5] which focused on retrospective analyses after the occurrence of this incident using conventional metallurgy or mechanical theories. According to these works, causes of strip snap in cold rolling are various: equipment factors, material defects, improper operation, sensor malfunction and production adjustment. Recently, researches of these strip snap cause analysis have been conducted by employing data analytics [6, 7]. The studies carried out based on a subset of selected variables from hundreds of process measurements to analyze the correlations between these selected variables and strip snap. However, before data analytics were conducted in these works, variables had already been selected based on previous domain knowledge and expertise, thus leaving out of other important causes. In addition, these studies investigate the strip snap event at a general batch perspective. The instances are the

extracted from the time span of an entire cold rolling process and each data instance is associated with a single class label that affects its entire behavior. It is assumed in these studies that all temporal observations are equally useful for prediction [8].

The scope of this paper is to predict the strip snap through several selected cold rolling process variables using machine learning algorithms. The strip snap events were represented as the event occurrence probability for each time observation. This probability is assumed to be of continuous values and therefore several regression algorithms were applied.

The subsequent sections are structured as follows. In Section II, a review of both the cold rolling process and strip snap cause analysis using different methods is addressed, followed by a review on cold rolling process characterization. Subsequently, the research problem of this study is stated in Section III, and the methodology is proposed in Section IV, a case study which aims at predicting the probability of strip snap through regression is revealed in Section V. Finally, the findings and further works are concluded in Section VI.

## II. LITERATURE REVIEW

### A. Cold Rolling Process and Research on the Cause of Strip Snap

The rolling process has an essential role in the manufacture of a wide variety of products because of its high accuracy, efficiency and production rate. Sheets and strips can be rolled either in the single stand or tandem mills [1]. As one of the main processes in electrical steel strip production, cold rolling enhances strip properties by changing the microstructure and thickness of the steel. These enhanced properties include yield strength, tensile strength, surface smoothness and hardness [6].

Like other metal forming processes, the final product of cold rolling can exhibit some mechanical defects. Various defects were observed in industrial metal forming processes, including plane strain rolling. Based on 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 tearing requires special consideration, because it does not only significantly increase production costs but can also cause serious damage to rolls and mill accessories [9].

As one of the most common production incidents in the cold rolling process, plenty of research has been conducted on the causes of strip snap. It has been summarized [4] that strip

Zheyuan Chen and Ying Liu are with Institute of Mechanical and Manufacturing Engineering, School of Engineering, Cardiff University, Cardiff, UK(corresponding author e-mail: chenz57@cardiff.ac.uk).

Agustin Valera-Medina is with School of Engineering, Cardiff University, Cardiff, UK.

Fiona Robinson is with Cogent Power Ltd, Orb Electrical Steels, Newport, UK.

snaps could result from material defect, equipment malfunction, improper operation and improper rolling parameter settings.

Among these possible causes, equipment factor has been analyzed as the primary one. In a real case study [5], servo-valve malfunction resulted in high-pressure fluctuation, thus leading to inter-frame tension deviations, crushing the strip on one side. Other equipment malfunctions such as the piston rod protrusion of HGC (Hydraulic Gap Control) and tension meter detection accuracy have also been discussed by [10].

Apart from equipment factors, inappropriate operation and parameter settings also account for the occurrence of strip snap. Several operating parameters related to the working roll such as diameter disparity between top and bottom working rolls, levelness of the bottom working roll and convexity degree of the working rolls have been discussed to be significant strip snap causes. Apart from working rolls, levelness and perpendicularity of the deflector rolls have also been proved to generate strip snaps as well [3].

### B. Characterizing the Cold Rolling Process

Regarding the modelling of the cold rolling process, many different models have been developed and presented over the past decade. 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. A model developed by Orowan was one of the most comprehensive among these cold rolling process models. Conventional rolling force formulas, however, provide not more than reasonably accurate approximations [11, 12].

Recognized to be a desirable approach to investigate the design of mill equipment and rolling operation practices, the mathematical modelling of the cold rolling process is conducive to productivity and quality improvement [13].

However, many factors such as friction conditions, roll flattening, deflection of the rolling mill and temperature make the theoretical analysis of the rolling process very complicated and time-consuming. Since the exact values of these variables cannot be measured during the rolling process, there are many parameters needed for better accuracy of the mathematical model [14]. For example, strip snap has been analyzed using a strip deformation model. Nevertheless, the strip deformation models are dependent on parameters which are determined from a combination of approximations to existing rolling theory [15].

### III. PROBLEM STATEMENT

This paper deals with an undesired event - the occurrence of strip snap during a grain-oriented electrical steel cold rolling process. According to the domain experts, this type of steel is fragile due to a high silicon content which leads to a higher rate of strip snaps during cold reduction to 10% of the starting thickness. The occurrence of this incident can result in yield loss due to the failure to achieve the final target thickness. In addition to this, when the cold rolling production resumes from this incident, an altered rolling performance may occur due to the unexpected disruption caused by strip snaps. This disruption may result in a variation of strip thickness, thus influencing the magnetic loss which is proportional to strip thickness. It is therefore of great significance to reveal a

functional relationship between certain selected measured variables and strip snap incident regarding production yield improvement and cost reduction for cold rolling processes.

From hundreds of measured variables, our previous work was conducted through a data analytics approach to choose and determine the most relevant variables affecting snap occurrence. The result of the selected variables provides a reference for input variables in this study.

However, in this previous work, each data instance in the training dataset represents the mean value of a time span covering a rolling pass since the strip snap incidents were marked by the shop floor engineers with the most specific to the pass number of a broken coil. In this context, the previous work assumes that each data instance (represented by a time series of a rolling pass) is associated with a single class label that affects its entire behavior. That is, it is assumed that all temporal observations are equally useful for prediction.

In terms of monitoring and detecting an event such as a strip snap incident, the above manner of labelling the whole temporal observations in a pass is general and inappropriate. In practical, the collected data were recorded in a continuous time series manner under a sampling rate of 100hz. Meanwhile, the occurrence of a strip snap was momentary. This incident can take place at any stage in a rolling pass and it is impractical to locate the exact start time point for this incident. For a rolling pass, it is very common that the mill was in the normal condition in the early stage and a strip snap occurred at the later stage. In this case, only the temporal observations in later stage should be labelled as a snap rather than labelling the whole temporal observations of the entire pass as a snap.

In this context, the label denotes an event should be associated with a specific time point or a time interval in the instance, not necessarily in the entire time sequence. In the cold rolling domain, a rolling pass may be normal at first, then develops a tendency of strip snap, then a snap occurs or goes back to normal condition again. Consequently, the context in which the prediction is made should be local and affected by the most recent behavior of the monitored event.

### IV. METHODOLOGY

The objective of this study is to discover a functional relationship between certain measured variables recorded in the form of multivariate time series and the occurrences of strip snap. The measured variables are determined by both domain experts and previous work. The strip snap events are represented as the probability of strip snap occurrence within a time window denoted as  $P_{s,t}$  which is assumed to be of continuous values.

To achieve the objective stated above, a machine learning approach is proposed. This approach aims to label an unseen instance (a time interval of a certain length) from selected variables with  $P_{s,t}$ .

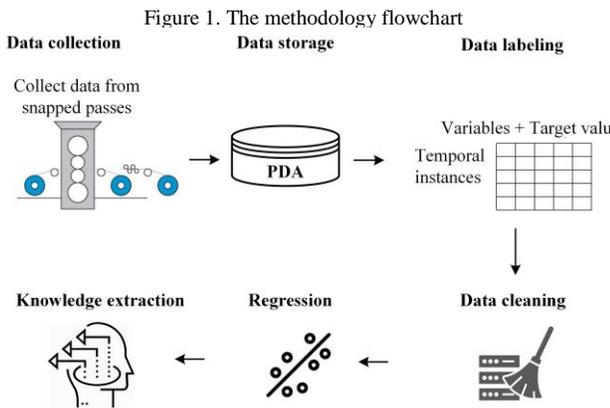
The first step of this machine learning approach is to collect the PDA time series data of snapped passes with selected variables.

Following this, with the aims of transforming the problem into a supervised learning task and being specific to the

detailed temporal observation length, the labels which is assumed to be of continuous values in this study are assigned to each time interval of certain length according to a selected index which is highly related with the event.

Subsequently, data cleaning is conducted to filter out the noise and to discard the temporal observations of idle states in a cold rolling process.

Regression algorithms are then adopted with the aim of establishing models to reveal the relationships between selected variables and the strip snap incident. Further investigations regarding the performance of the regression algorithms are finally conducted.



## V. CASE STUDY

### A. Experimental Setup

This case study was conducted based on the historical data collected from the PDA system of a reversing mill in an electrical steel company. This company produces strip coil products which are cold rolled five passes back and forth

through a reversing mill. With the silicon exceeding 3% and the thickness decreasing by up to 90%, these fragile strips often break undesirably during the cold rolling process.

The following experiment is set up with the aim of detecting a strip snap event by establishing a regression model to reveal a functional relationship between several selected input variables and output  $P_{s,t}$  in each time window. If  $P_{s,t}$  can be predicted in a short time window, the steel plant can set a predictive manner to stop the mill before this incident takes place.

To conduct this supervised learning experiment, the first step is to label the time window in the collected PDA time series with  $P_{s,t}$ . In this steel company, the occurrences of a strip snap are recorded by the shop floor engineers in an event log. By tracking this event log, the passes of each strip coil can be marked with a good pass or a snap pass. According to the domain experts, from various monitoring variables in the PDA system, Measured Slip (MS) is highly relevant with the strip snap event and therefore it can be used to track and locate an approximate scope of a strip snap event on the timeline. This measurement indicates the relative displacement between the strip coil and the working roll. Under normal operation conditions, the value of MS is usually negative or slightly fluctuating around 0. As shown in Figure 2, when a strip snap takes place, the MS value increases sharply to a peak as a response to this incident and then plummets to the vicinity of 0. The strip snap event occurs before the plummeting event of MS and there is an unmeasurable delay between these two events. Hence, for a snap pass, the time series after a time point (noted as time point B) can be deduced with 100% of  $P_{s,t}$  only when the MS increases to the maximum at this time point. However, it is impractical and irrational to locate the exact start time point of a strip snap occurrence. To discover the difference between good passes and snap passes regarding MS value, based on the collected data of 1010 good passes and 149 snap passes, statistics analysis was conducted the max MS value. The statistical results are shown in Figure 3 and Figure 4.

Figure 2. Quantify  $P_{s,t}$  with MS value in a snap pass

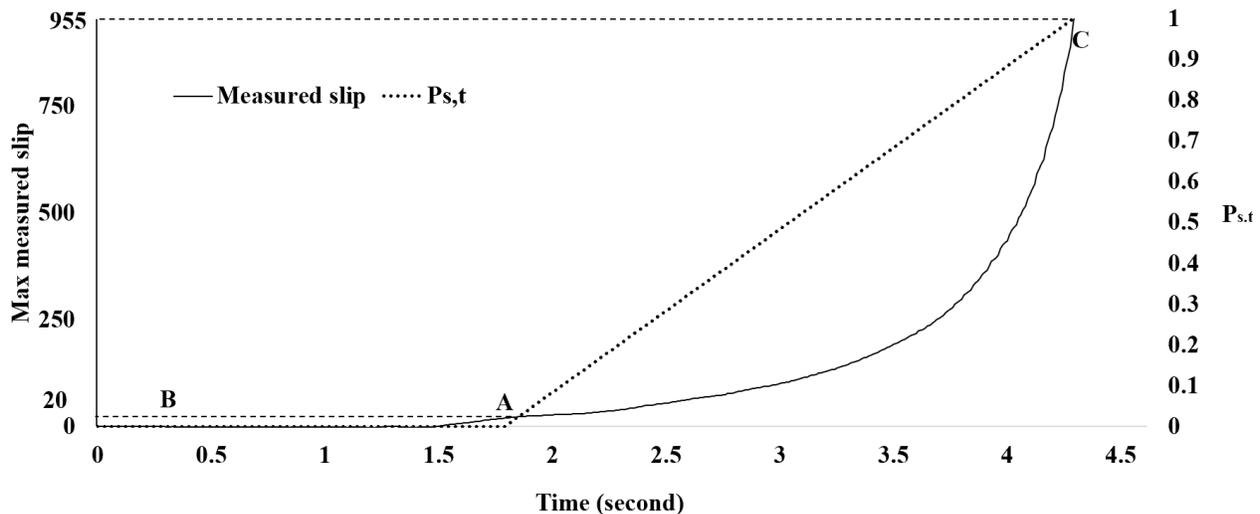


Figure 3. Plot box of Max MS values of 149 snap passes

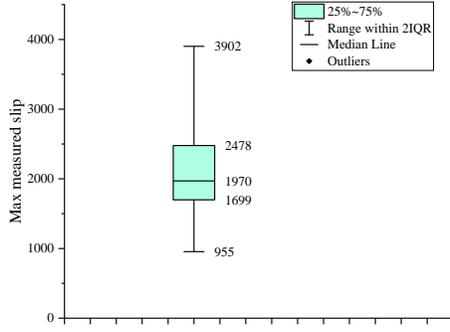
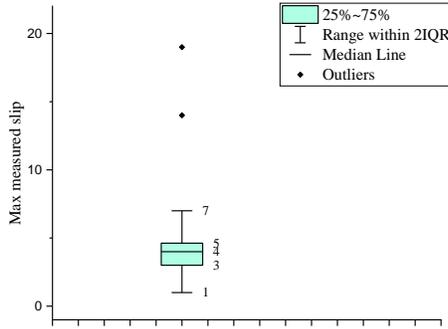


Figure 4. Plot box of Max MS values of 1010 good passes



For the collected dataset, MS values of all good passes are fluctuating under 20; while the max MS values of all snap passes vary from 955 to 3902 which is far above 20. This obvious difference on the MS values regarding good and snap passes can be used as an index for strip snap.

To quantify  $P_{s,t}$  through the MS value, two assumptions were made: Firstly, the variation manner of the MS value for good passes is the same with this manner for snap passes. Under this assumption, for a snap pass,  $P_{s,t} = 0$  when MS value is under 20. The time point when MS=20 is noted as time point A. The strip snap takes place between the time point A (MS=20) and B (MS=maximum). Secondly, between time point A and B,  $P_{s,t}$  is linearly distributed.

Under these two assumptions, Figure 2 illustrates how a snap time series is labelled with  $P_{s,t}$  value with reference to MS value.

The following data preprocessing procedures were coded and carried out using Python of version 3.7.2 [16]. The regression experiments were coded and carried out through the open-source package Scikit-learn of version 0.20.3 [17].

### B. Data Collection and Preprocessing

This study focuses on the cold rolling process rather than on the incoming material. Variables related to materials are not in the scope of this research. In this context, the collected 149 snap passes were rolled with the same grade of feedstock material.

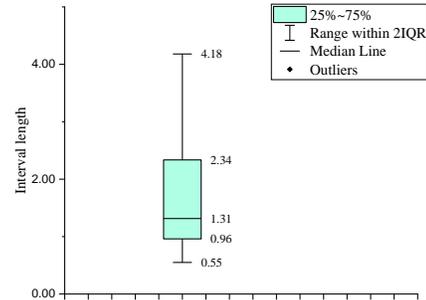
Table I illustrates the 11 selected variables according to domain experts and previous work. It should be noted that within this dataset, there are negative values in variable ‘‘Stand RPM’’ and ‘‘Mill speed’’, thus indicating the vector direction related to the reversing mill process. The absolute value of these negative values was taken before a Min-Max normalization was conducted on this training dataset.

TABLE I. TABLE OF INPUT VARIABLES

No.	Input variables from domain experts	Description	Unit
V1	Stand RPM	Rotational velocity of the stand	Rpm: revolutions per minute
V2	Servo total load feedback	-	T: ton
V3	Mill speed	Rolling speed	Mpm: meters per minute
V4	RR tension	Tension applied to the right reel	T: ton
V5	Stand current	-	p.u (per unit)
<b>Input variables from previous work</b>			
V6	Bottom 750 lb Hydraulic pressure feedback	Pressure feedback	PSI: Pounds per square inch
V7	Servo back position B	-	Um: micrometer
V8	DS Capsule Servo Input Supply Pressure Fb	Pressure feedback	Bar: pressure metric unit
V9	Rate of Change of Total Load Reference	Number of times	-
V10	Gap Control on Permits	The number of gap control operations	-
V11	Back Capsule force	Force	T: ton
V12	Mean trim 1-4	-	Um/scan

In the event detection setting, the event-recent training temporal instances closer to the event are usually more predictive than distant measurements [8]. In this context, the training dataset are extracted from the time sequence recent to the time point B. In order to train a classifier which can learn both the patterns of good and snap passes, training temporal instances are also extracted from point A forward to a time point noted as C. Considering the class balance issue [18], the time interval length between time point A to C (noted as  $L_{AC}$ ) is set to 1.5 seconds with reference to the length between time point A to B (noted as  $L_{AB}$ ). The Figure 5 illustrates the distribution of  $L_{AB}$  regarding the collected 149 snap passes.

Figure 5. Plot box of  $L_{AB}$  regarding 149 snap passes.



For the preprocessing of time series data, it is usually desirable to develop representation techniques that can reduce the dimensionality of time series [19]. In this study, raw data was collected from the PDA system at a sampling frequency of 100Hz by taking the mean value, which means the raw data are consist of time windows with the size of 0.01 second. The raw data was already represented using a Piecewise Aggregate Approximation (PAA) transformation which produces a piecewise constant approximation of the original sequence [20]. Considering the represented data is low in dimension and follows the distribution of  $L_{AB}$ , it can be used for further modelling.

### C. Evaluation Metrics

Many regression techniques can be used to conduct this task. In this work, Naïve Bayes (NB) [21], Decision Tree (DT) [22],  $k$ -Nearest Neighbors ( $k$ -NN) [23] and Random Forest (RF) [24] are tested respectively.

The regression performance was evaluated under 10-fold cross-validation.

In order to reveal the regression performance of the selected algorithms, the root mean square error (RMSE) is used. It is a scale-dependent metric which measures the difference between the prediction value and the actual value [25]. RMSE is 0 if the prediction value equals to the actual values. The expression of RMSE is:

$$RMSE = \sqrt{\frac{\sum_i (p_i - a_i)^2}{n}} \quad (1)$$

In the formula,  $p_i$  is the prediction value,  $a_i$  is the actual value,  $n$  is the number of training data instances

Regarding correlation coefficient, Model correlation coefficient [26] is used to measure the correlation between two variables. MCC is used to measure the correlation between two variables, and can be expressed mathematically as [27]:

$$MCC = \frac{S_{PA}}{\sqrt{S_P S_A}} \quad (2)$$

where,

$$S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n - 1}; S_P = \frac{\sum_i (p_i - \bar{p})^2}{n - 1};$$

$$S_A = \frac{\sum_i (a_i - \bar{a})^2}{n - 1};$$

In the formula,  $\bar{p}$  is the average of the prediction value,  $n$  is the number of training data instances.

### D. Results and Discussions

Using all these four algorithms, both the RMSE and MCC was excessively high as shown in Table II below.

TABLE II. REGRESSION PERFORMANCE ON FULL SET

	NB	DT	$k$ -NN	RF
RMSE	0.093	0.009	0.009	0.008
MCC	0.961	0.999	0.999	0.999

This could be caused by several input variables which are strongly related to  $P_{s,t}$ . Therefore, regarding these 11 input variables, Correlation-based subset feature selection (CFSSub) [28], a feature selection technique using filter approach is carried out with the aim of measuring the correlation between variables and class concept. This filter approach focus on the evaluation of every single attribute [29]. As shown in Table III, seven input variables were selected as strongly related variables.

TABLE III. STRONG CORELATED VARIABLES USING CFSSUB

No.	Variable name
V1	Stand RPM
V3	Mill speed
V4	RR tension
V5	Stand current
V6	Bottom 750 lb Hydraulic pressure feedback
V7	Servo back position B
V8	DS Capsule Servo Input Supply Pressure Fb

According to the feature selection result, these seven variables could be strongly related with  $P_{s,t}$ . With the removal of these seven variables, the same regression experiment was conducted for comparison based on the remaining five variables and the result is shown in Table IV. The training dataset of these five variables is noted as training dataset **B**.

TABLE IV. REGRESSION PERFORMANCE AFTER REMOVING STRONG RELATED VARIABLES

	NB	DT	$k$ -NN	RF
RMSE	0.255	0.167	0.148	0.157
MCC	0.646	0.870	0.897	0.883

The MCCs of the different algorithms decrease to a reasonable interval while the RSMes are still desirable. To reveal the impact of each strong correlated variable on MCC, a set of comparative experiments were conducted. The results are shown in Table V.

TABLE V. MCCS OF MODELS ESTABLISHED ON DIFFERENT CONFIGURED DATASETS

Dataset	NB	DT	$k$ -NN	RF
<b>B</b>	0.646	0.870	0.897	0.883
<b>B+V1</b>	0.893	0.995	0.987	0.997
<b>B+V3</b>	0.899	0.997	0.992	0.998
<b>B+V4</b>	0.775	0.991	0.983	0.994
<b>B+V5</b>	0.646	0.959	0.917	0.968
<b>B+V6</b>	0.645	0.982	0.979	0.989
<b>B+V7</b>	0.918	0.980	0.966	0.988
<b>B+V8</b>	0.722	0.989	0.984	0.993

As shown in Table V, adding any of the selected seven strongly correlated variables, the MCCs regarding the models based on different algorithms increased obviously, which indicates that these seven variables are strongly correlated with the target  $P_{s,t}$  which is represented by MS. However, from the perspective of domain knowledge and statistical result shown in Figure 2, this MS value is irrelevant with these seven variables since MS is relatively stable regardless of the variation of these seven variables under normal condition. However, the MS value in this study is used to represent the probability of strip snap occurrences. The strong correlation between the MS and these seven variables indicates that these seven variables are strongly related with the strip snap events as well.

## VI. FUTURE WORK AND CLOSING REMARKS

This study proposed an approach to label the temporal instances with the probability of strip snap occurrence through statistics and assumptions on a measured variable MS. In practice, the relative displacement which is measured as MS is a consequence and manifestation of strip snap. Lag exists between the MS measurement and strip snap occurrence. However, the seven variables which are proved to be strongly correlated with the probability of strip snap occurrence can be used to represent the probability of strip snap occurrence more informative and timely compared with the current approach. Sensitivity analysis can also be applied to analyse how these seven selected variables will impact the probability of strip snap occurrence in future.

## ACKNOWLEDGMENT

The authors wish to thank Darren Reeve from Orb Electrical Steels of Cogent Power Ltd for their technical support and suggestions.

## 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] S. Iwadoh and T. Mori, "Effect of work roll materials and progress of manufacturing technology on cold rolling and future developments in Japan," *ISIJ international*, vol. 32, no. 11, pp. 1131-1140, 1992.
- [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] J. Xu, "Cause Analysis for Strip-breaking of Non-oriented Silicon Steel," *JIANGXI METALLURGY*, no. 2015 年 05, pp. 26-28, 2015.
- [5] 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.
- [6] 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.
- [7] C. Wang, "Research on Fault Diagnosis of Belt Tearing in the Cold Rolling Process," *Instrumentation Technology*, no. 09, pp. 16-20, 2014.
- [8] 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.
- [9] 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.
- [10] K. Yan and H. Li, "Causes and Countermeasures of Cold Strip Break in Masteel," *Anhui Metallurgy*, no. 4, pp. 36-37, 2006.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] H. Furumoto, K. Yamada, and J. Yanagimoto, "Effect of the number of work-roll surface division on prediction of contact length in coupled analysis of roll and strip deformation during sheet rolling," *ISIJ international*, vol. 42, no. 7, pp. 736-743, 2002.
- [16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-2830, 2011.
- [17] W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc.", 2012.
- [18] R. Longadge and S. Dongre, "Class imbalance problem in data mining review," *arXiv preprint arXiv:1305.1707*, 2013.
- [19] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 275-309, 2013.
- [20] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263-286, 2001.
- [21] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, vol. 3, no. 22, pp. 41-46: IBM New York.
- [22] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660-674, 1991.
- [23] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?," in *International conference on database theory*, 1999, pp. 217-235: Springer.
- [24] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18-22, 2002.
- [25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679-688, 2006.
- [26] M. McCann, Y. Li, L. Maguire, and A. Johnston, "Causality challenge: benchmarking relevant signal components for effective monitoring and process control," in *Causality: Objectives and Assessment*, 2010, pp. 277-288.
- [27] R. Boddy and G. Smith, *Statistical methods in practice: for scientists and technologists*. John Wiley & Sons, 2009.
- [28] M. A. Hall, "Correlation-based feature subset selection for machine learning," *Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato*, 1998.
- [29] L. Talavera, "An evaluation of filter and wrapper methods for feature selection in categorical clustering," in *International Symposium on Intelligent Data Analysis*, 2005, pp. 440-451: Springer.