Data Mining for Fault Diagnosis in Steel-making Process under Industry 4.0

Zheyuan Chen

School of Engineering

Cardiff University

A thesis submitted to Cardiff University

for the degree of Doctor of Philosophy

September 2022

Acknowledgements

First and foremost, I am deeply indebted to my supervisor, Prof. Ying Liu, who has been extremely supportive and helpful during this journey. Since my first day as a PhD student, he provided me with every bit of guidance, patience, assistance, criticism, and expertise that I needed during my doctoral work. Under Prof. Liu's supervision, I have learnt the fundamentals of conducting scientific research, as well as how to build a proactive researcher mindset and explore the unknown field with novel methods and critical thinking. His passionate attitude, persistent striving for excellence and a strong willingness to embrace new challenges have motivated me to overcome all those difficulties, especially during the pandemic periods, in my PhD study years. To me, what I have emulated from Prof. Liu will benefit not only my future research career but also the rest of my life to be a better person.

My sincere thanks and appreciations also go to my co-supervisor Prof. Agustin Valera-Medina, who has always been supportive and encouraging to me. Also, much appreciated to Cogent Power, Dr. Fiona and Mr. Darren for their data and domain knowledge support. In the meantime, I always feel enjoyable and stimulated to collaborate with senior and junior colleagues in Prof. Liu's research group. Many thanks to Dr. Jian Qin and Dr. Chong Chen, the seniors in our group, for their kind help and encouragement. Also, I am thankful to my sponsor, Tianjin Chison Chemical Co. Ltd and CEO Mr. Denghua Nie, for their financial support of my PhD research.

Finally, I would like to express my eternal gratitude to my parents. The unconditional love and support from my parents have fuelled me with all the confidence and courage throughout this journey.

Abstract

The concept of Industry 4.0 (I4.0) refers to the intelligent networking of machines and processes in the industry, which is enabled by cyber-physical systems (CPS) - a technology that utilises embedded networked systems to achieve intelligent control. CPS enable full traceability of production processes as well as comprehensive data assignments in real-time. Through real-time communication and coordination between "manufacturing things", production systems, in the form of Cyber-Physical Production Systems (CPPS), can make intelligent decisions. Meanwhile, with the advent of I4.0, it is possible to collect heterogeneous manufacturing data across various facets for fault diagnosis by using the industrial internet of things (IIoT) techniques. Under this data-rich environment, the ability to diagnose and predict production failures provides manufacturing companies with a strategic advantage by reducing the number of unplanned production outages. This advantage is particularly desired for steel-making industries. As a consecutive and compact manufacturing process, process downtime is a major concern for steel-making companies since most of the operations should be conducted within a certain temperature range. In addition, steel-making consists of complex processes that involve physical, chemical, and mechanical elements, emphasising the necessity for data-driven approaches to handle high-dimensionality problems.

For a modern steel-making plant, various measurement devices are deployed throughout this manufacturing process with the advancement of I4.0 technologies, which facilitate data acquisition and storage. However, even though data-driven approaches are showing merits and being widely applied in the manufacturing context, how to build a deep learning model for fault prediction in the steel-making process considering multiple contributing facets and its temporal characteristic has not been investigated. Additionally, apart from the multitudinous data, it is also worthwhile to study how to represent and utilise the vast and scattered distributed domain knowledge along the steel-making process for fault modelling. Moreover, state-of-the-art does not address how such accumulated domain knowledge and its semantics can be harnessed to facilitate the fusion of multi-sourced data in steel manufacturing. In this case, the purpose of this thesis is to pave the way for fault diagnosis in steel-making processes using data mining under I4.0.

This research is structured according to four themes. Firstly, different from the conventional data-driven research that only focuses on modelling based on numerical production data, a framework for data mining for fault diagnosis in steel-making based on multi-sourced data and knowledge is proposed. There are five layers designed in this framework, which are multi-sourced data and knowledge acquisition, data and knowledge processing, KG construction and graphical data transformation, KG-aided modelling for fault diagnosis and decision support for steel manufacturing.

Secondly, another of the purposes of this thesis is to propose a predictive, data-driven approach to model severe faults in the steel-making process, where the faults are usually with multi-faceted causes. Specifically, strip breakage in cold rolling is selected as the modelling target since it is a typical production failure with serious consequences and multitudinous factors contributing to it. In actual steel-making practice, if such a failure can be modelled on a micro-level with an adequately predicted window, a planned stop action can be taken in advance instead of a passive fast stop which will often result in severe damage to equipment. In this case, a multi-faceted modelling approach with a sliding window strategy is proposed. First, 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 physics-based approaches, empirical knowledge and data-driven features. Finally, these features were used as inputs for strip breakage modelling using a Recurrent Neural Network (RNN). Experimental results have demonstrated the merits of the proposed approach.

Thirdly, among the heterogeneous data surrounding multi-faceted concepts in steel-making, a significant amount of data consists of rich semantic information, such as technical documents and production logs generated through the process. Also, there exists vast domain knowledge regarding the production failures in steel-making, which has a long history. In this context, proper semantic technologies are desired for the utilisation of semantic data and domain knowledge in steel-making. In recent studies, a Knowledge Graph (KG) displays a powerful expressive ability and a high degree of modelling flexibility, making it a promising semantic network. However, building a reliable KG is usually time-consuming and labour-intensive, and it is common that KG needs to be refined or completed before using in industrial scenarios. In this case, a fault-centric KG construction approach is proposed based on a hierarchy structure refinement and relation completion. Firstly, ontology design based on hierarchy structure refinement is conducted to improve reliability. Then, the missing relations between each couple of entities were inferred based on existing knowledge in KG, with the aim of increasing the number of edges that complete and refine KG. Lastly, KG is constructed by importing data into the ontology. An illustrative case study on strip breakage is conducted for validation.

Finally, multi-faceted modelling is often conducted based on multi-sourced data covering indispensable aspects, and information fusion is typically applied to cope with the high dimensionality and data heterogeneity. Besides the ability for knowledge management and sharing, KG can aggregate the relationships of features from multiple aspects by semantic associations, which can be exploited to facilitate the information fusion for multi-faceted modelling with the consideration of intra-facets relationships. In this case, process data is transformed into a stack of temporal graphs under the fault-centric KG backbone. Then, a Graph Convolutional Networks (GCN) model is applied to extract temporal and attribute correlation features from the graphs, with a Temporal Convolution Network (TCN) to conduct conceptual modelling using these features. Experimental results derived using the proposed approach, and GCN-TCN reveal the impacts of the proposed KG-aided fusion approach.

This thesis aims to research data mining in steel-making processes based on multi-sourced data and scattered distributed domain knowledge, which provides a feasibility study for achieving Industry 4.0 in steel-making, specifically in support of improving quality and reducing costs due to production failures.

Research Achievements

Journal papers:

1. **Z. Chen,** Y. Liu, A. Valera-Medina, F. Robinson, and M. Packianather, "Multi-faceted modelling for strip breakage in cold rolling using machine learning," International Journal of Production Research, pp. 1-14, 2020. **(Impact factor: 8.568)**

2. **Z. Chen,** Y. Wan, Y. Liu and A. Valera-Medina, " A Knowledge Graph-Supported Information Fusion Approach for Multi-faceted Conceptual Modelling, " Information Fusion **(Impact factor: 17.564) (Submitted)**

3. Y. Wan, **Z. Chen**, Y. Liu, M. Packianather, and R. Wang, "Building a Domain-centric iKG through Hierarchy Structure Refinement and Relation Completion, " Advanced Engineering Informatics **(Impact factor: 7.862) (Under review)**

Conference papers:

1. **Z. Chen**, Y. Liu, A. Valera-Medina, and F. Robinson, "Multi-sourced Modelling for Strip Breakage using Knowledge Graph Embeddings" Procedia CIRP, 2021.

2. **Z. Chen**, Y. Liu, A. Valera Medina, and F. Robinson, "A multi-source feature-level fusion approach for predicting strip breakage in cold rolling," presented at the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Virtual, 20-24 August 2020, 2020.

3. **Z. Chen**, Y. Liu, A. Valera-Medina, and F. Robinson, "Characterising strip snap in cold rolling process using advanced data analytics," Procedia CIRP, vol. 81, pp. 453-458, 2019.

4. **Z. Chen**, Y. Liu, A. Valera-Medina, and F. Robinson, "Strip Snap Analytics in Cold Rolling Process Using Machine Learning," presented at the 2019 IEEE 15th International

Conference on Automation Science and Engineering, Vancouver, BC, Canada, 22-26 August, 2019.

5. J. Yang, **Z. Chen**, Y. Liu, and M. Ryan, "Sliding window filter based strip breakage modelling for failure prediction," in 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), 2021, pp. 1461-1466: IEEE.

6. Y. Wan, **Z. Chen**, F. Hu, Y. Liu, M. Packianather, and R. Wang, "Exploiting Knowledge Graph for Multi-faceted Conceptual Modelling using GCN," in ISM 2021 International Conference on Industry 4.0 and Smart Manufacturing.

Table of Contents

List of Tables

List of Figures

List of Symbols

b_i: The output of the LSTM memory cell

b_j: The bias in the fully connected layer

b_ι: The factors of the input gate of the LSTM memory cell

b_Φ: The factors of forget gate of the LSTM memory cell

b_ω: The factors of the output gate of the LSTM memory cell

c_i: The input of the LSTM memory cell

$f()$: The activation function in a fully connected layer

d_i: The difference between the two ranks of each observation

n_c: The number of concordant pairs

n_d: The number of discordant pairs

p_i: The predicted probability

r_{prs}: Pearson correlation coefficient

r_{spm}: Spearman rank correlation coefficient

$\sigma(\cdot)$: The activation function

τ: Kendall rank coefficient

List of Abbreviation

AI: Artificial Intelligence

ANN: Artificial Neuron Network

AUC: The Area Under the receiver operating Characteristic curve

ACC: Accuracy

BF-BOF: Blast Furnace-Basic Oxygen Furnace

BPNN: Back Propagation Neural Network

CNN: Convolutional Neural Network

DD: Data-driven

DRI: Direct Reduced Iron

DT: Decision Trees

EAF: Electric Arc Furnace

EK: Empirical-Knowledge

FAR: False Alarm Rate

GCN: Graph Convolutional Networks

GF: Graph Factorization

GNN: Graph Neural Networks

GPU: Graphics Processing Unit

GRL: Graph Representation Learning

GRU: Gate Recurrent Unit

HRC: Hot-rolled Coil

I4.0/Industrial 4.0: The Fourth Industrial Revolution

IIoT: Industrial Internet of Things

IoT: Internet of Things

KG: Knowledge Graph

PCA: Principal Component Analysis

LSTM: Long-short-term Memory

NB: Naive Bayes

NLP: Natural Language Processing

NN: Neural Networks

PB: Physics-based

PDA: Production Data Acquisition

RDF: Resource Description Framework

ReLU: Rectified Linear Unit

RF: Random Forest

RNN: Recurrent Neural Network

SVC: Support Vector Classification

SVM: Support Vector Machine

TADW: Text-associated Deep Walk

TCN: Temporal Convolution Network

TNR: True Negative Rate

TPR: True Positive Rate

Chapter 1 Introduction

1.1 Background

During the past decade, the term Industry 4.0, which originated in Germany, has been adopted globally. Industry 4.0 technologies are being developed and implemented in a number of countries, and a substantial amount of research has been conducted to develop and implement some of these technologies. The Internet of Things, cloud computing, smart sensors, and other technologies are being integrated into today's factories as part of I4.0 (Cemernek et al., 2017). CPS is considered one of the key enablers of I4.0. By linking real objects, for example, machines and products, with their cyber representatives, CPS enable full traceability of production processes as well as comprehensive data assignments in real-time. Through real-time communication and coordination between "manufacturing things", production systems, in the form of Cyber-Physical Production Systems (CPPS), can make intelligent decisions (Bauernhansl et al., 2014). Specifically, Industry 4.0 focuses on the intelligent networking of machines and processes utilizing embedded networked systems to achieve intelligent control (Lu et al., 2020).

The Smart Factory initiative is one of the major constituent parts of Industry 4.0 (Zuehlke, 2010). In a modern manufacturing plant, the data collection and transfer of information from almost all processes are done electronically using powerful data acquisition systems. Continuous measurements are being conducted at various stages of manufacturing, and the values of these variables are recorded in the databases of organizations. The manufacturing industry generates large volumes of raw data continuously as a result of technological advancements. As a result of the availability of large quantities of data and their increasing quantities, there has been an increased interest in the study of the machine learning concept (Ito et al., 2019). In this sector, data mining applications have been utilized for solving manufacturing problems for nearly two decades. The use of ML methods to perform manufacturing tasks can be exemplified by intelligent systems to support the scheduling of multiple production lines simultaneously and arrangements for the maintenance of machines (Mehrpouya et al., 2019). The detection of manufacturing defects is another example, as well as predicting failures, estimating the energy consumption of machines, and estimating machine energy consumption (Dogan and Birant, 2021).

In addition to the heterogeneous data collected along the steel-making workflow, there is vast existing domain knowledge which has not been fully exploited for data-driven modelling in steel-making. Techniques such as ontology (Lu and Xu, 2017) can be used to develop semantic descriptions of manufacturing resources. As a semantic model, ontology can define and describe a wide variety of entities, features and properties existing in a specified domain (Otero-Cerdeira et al., 2015). However, as ontologies are based on rule representations, they have limited flexibility and adaptability when it comes to describing the semantics of large-scale workshop data. In contrast, KG displays a powerful expressive ability and a high degree of modelling flexibility, making it a promising semantic network (Paulheim, 2017). The feeding of these production data and knowledge into ML methods provides the basis for quality control of highly complex and multidimensional processes.

A typical example of such a complex and high-dimensional manufacturing process is the steel industry. Throughout the industrial sector, steel is a key raw material used in the construction of buildings, bridges, ships, containers, medical instruments, and automobiles (Cemernek et al., 2022). In this context, quality improvements would be highly beneficial to the steel industry with the highly competitive nature of the global steel market since even the smallest variation during the production process causes costly and time-consuming post-processing or scrap (WorldsteelAssociation, 2022).

In steel production, rolling is the main method of metal plastic forming, with its products accounting for more than 90% of metal plastic processing (Mazur and Nogovitsyn, 2018). The rolling process can be classified as hot rolling and cold rolling. Steel produced by cold rolling has closer dimensional tolerances and a broader range of surface finishes than steel produced by hot rolling. In addition, it can be up to 20% stronger than hot-rolled steel due to the use of strength hardening. In this context, the cold rolling process is a primary metal-forming process for the manufacturing of steel strips (Mashayekhi et al., 2011). For the cold rolling process, there are a number of product failures leading to serve consequences. For example, strip breakage, also known as strip snap or strip tearing, is one of the most common quality issues in the cold rolling process (Mashayekhi et al., 2011). This incident resulted in damage to 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 (Iwadoh and Mori, 1992).

For a modern steel-making plant, various measurement devices are deployed throughout this manufacturing process with the advancement of I4.0 technologies, which facilitate data acquisition and storage (Dogan and Birant, 2021). Meanwhile, emerging semantic techniques such as KG and graph representation learning enable an effective pathway to achieve cognitive intelligence by exploiting and aggregating the relationship of events, features, and equipment components (Xia et al., 2022). In this data-rich environment, data-driven approaches can provide a powerful tool to support quality improvement and reduce costs due to production failure. Also, with the merits of better visualisation and reasoning ability, modelling such production failure using semantic approaches can be a promising manner to achieve accurate perception and better understanding for the fault diagnosis in steel-making.

1.2 Motivations

The steel-making industry places a great deal of emphasis on quality control, which is determined by the characteristics of this industry (Kano and Nakagawa, 2008, Zhang et al., 2018). As a process-oriented industry, the steel industry aims to ensure that every production line is produced continuously throughout the entire production process (Tang et al., 2001). In this case, any interruption caused by production failure has a great impact on the consistency of this continuous process. In addition, the scale of production on each production line is relatively large, which means a whole batch of products would be affected by similar quality issues when problems arise. Due to this situation, steel companies seek to eliminate this risk by introducing better quality control approaches and minimising the consequences caused by production failures. There is also a great deal of complexity involved in the steel production process. There are five major stages of the manufacturing process, which are iron-making, steel-making, hot-rolling, cold-rolling, and heat treatment (Missbauer et al., 2009). Within each stage of the steel production process, there are various triggers for a production failure, and these failures can be caused by the former process. For example, in the cold-rolling process, there are numerous parameters that influence the quality of these cold-rolled products, such as the chemical composition of the liquid steel, the rolling speed and the cooling rate (Yan and Li, 2006b, Cui and Zhao, 2013b, Chen et al., 2019, Liu, 2015b, Liu et al., 2014, Takami et al., 2011b, Wang, 2014a, Yun et al., 1998, Lin et al., 2003).

In this context, for the steel-making process, which is characterised by high temperatures and pressures, elevated production speeds, and intense throughput, the early diagnosis of an incoming fault is highly relevant for both safety and economic reasons. The goal of the fault diagnosis method is to determine whether and when a monitored system starts to operate under anomalous conditions and eventually estimate the potential root-cause.

This research is in collaboration with the industrial partner Cogent Orb, a wholly-owned subsidiary of Tata Steel. This company produces Grain Oriented Electrical

Steels (GOES) for use in the cores of all types of transformers. Brittle High Si steels are cold reduced, decreasing the original thickness by 90% to achieve excellent magnetic losses, which are proportional to the final strip thickness. However, under high rolling speed and tension, within the cold-rolling process of such thin (0.23 mm) strip products, strip breakages can occur, which are undesirable. The historical stats indicate that there was an average of 24.3 breakages per 1000 tons manufactured in 2018. In addition to the high frequency of strip breakage in cold rolling, the financial losses associated with this problem are also substantial. For example, a single instance of strip breakage can result in several hours of downtime, repair costs, and lost production, with estimated financial losses ranging from tens of thousands to millions of pounds, depending on the severity of the breakage. Therefore, the ability to accurately diagnose and predict strip breakage and other production failures in the steel-making process is essential for reducing these financial losses and improving the overall efficiency and competitiveness of the industry.

For the diagnosis of faults such as strip breakage, methods can be roughly classified into two broad categories: model-based and data-driven methods. The first relies on the availability of a physical model of the system under analysis, the derivation of which can be a complex and time-consuming task. Moreover, the conventional approach is not able to handle production failures with highly complexity and multidimensionality (Cui and Zhao, 2013b, Xu, 2015, Liu, 2015b). 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.

Compared with conventional approaches, data-driven approaches have been widely employed with the advancement of I4.0 technologies (Kuo and Kusiak, 2019). 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 backward 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 (Chen et al., 2019, Takami et al., 2011a, Wang, 2014b). 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. Firstly, these works were conducted with the aim of quality characterisation (Da Cunha et al., 2006), which is the primary step for quality improvement rather than quality prediction (Lopes et al., 2019). Secondly, the rationale for determining the variables for breakage modelling was not explained and justified. Thirdly, the granularity of the data used in these works cannot match the temporal characteristic of strip breakage, which occurs instantaneously.

Moreover, the existing data-driven research on fault diagnosis in steel-making only considers the numerical production data. However, among the heterogeneous data surrounding the production failures in steel-making, a significant amount of data consists of rich semantic information, such as technical documents and production logs generated through the steel-making process. Also, as a conventional manufacturing process with a long history, there exists vast domain knowledge regarding the production failures in steel-making. Hence, in terms of modelling for fault diagnosis in steel making, proper semantic technologies are desired for the processing of semantic data and domain knowledge in steel-making.

Recently, Knowledge Graphs (KG) has attracted significant research interest due to their ability to store structured as well as unstructured knowledge elicited from heterogeneous domains and query them for the purpose of answering questions (Li et al., 2021). KG exhibits remarkable expressive ability and excellent modelling flexibility, which has been described as a graph model for representing information in a manner that can be understood by a broad audience (Paulheim, 2017). Considering it as a medium for conveying information in some web-based services, the majority of present studies have focused on the performance of KG itself, for example, KG-enable industrial products and services development. Specifically, there are three common research purposes of KG: its customization, enhancement and integration, such as building KG based on multiple sources and forms of industry records and developing algorithms based on semantics and topology to conduct knowledge deduction in multiple contexts. Besides, the advantages of KG also mitigate or even effectively resolve some industry pain points in product and service development, including multidisciplinary knowledge extraction and fusion, comprehensive solution searching, explainable knowledge recommendation, risk detection and prediction, and information distillation. As a result of flexibility in knowledge representation and advanced deduction approaches, KG alleviates multiple industry challenges in lifecycle stages and demonstrates its potential in various industrial products and services.

1.3 Research Questions and Objectives

Following the background and motivations, under the context of I4.0, this research aims to investigate data mining for fault diagnosis in steel-making based on multi-sourced data and knowledge. The following research questions have been formulated to achieve this goal:

1. *With the development of emerging technologies such as IIoT, CPPS, and KG in the context of I4.0, the steel-making industry has shifted to a data-rich and knowledge-intensive environment. In this case, what is an appropriate data mining framework for fault diagnosis in steel-making with the full exploitation of these resources?*

2. *Fault diagnosis is important for quality improvement in steel-making, and a data-driven approach has been widely employed. However, existing fault diagnosis research was conducted within the schema of quality characterisation, and the rationale for determining the variables for fault modelling was not explained and justified. Therefore, how to achieve fault modelling in a predictive manner with the justification of facets surrounding the production failure?*

3. *Besides the numerical production data in the steel-making process, it is widely acknowledged that the semantic data and existing domain knowledge are of great significance to fault diagnosis. Meanwhile, research on knowledge graphs has been receiving more and more attention because of the great expressive power of graphs. In this context, how to manage and exploit these resources in steel-making using KG with the consideration of their heterogeneity?*

4. *With the advancement of graph representation learning and GNN techniques, KG can be applied for machine learning analysis with the unique advantage of capturing intra-features relationships. Therefore, how can KG be utilised for the computational modelling of fault diagnosis using multi-sourced steel-making data?*

With the identification of the research questions, the research objectives following these research questions are listed below:

1. *To propose a framework of data mining for fault diagnosis in steel-making with full utilisation of multi-sourced data and knowledge.*

2. *To propose a data-driven approach that identifies features with solid rationality and models the production failures of steel-making in a predictive manner.*

3. *To study a semantic approach that can utilise semantic data and existing knowledge using KG.*

4. *To investigate a KG-aided approach to facilitate the integration of multi-sourced data for fault modelling in steel-making.*

The details of this research will be reported in Chapters 3, 4, 5 and 6.

1.4 Thesis Outline

In **Chapter 1**, a broader context and background are provided as to the motivation and significance of this research.

Chapter 2 presents a detailed review of the existing literature on related topics. It is divided into five parts: (1) An overview of the steel-making process with a specific review of cold-rolling; (2) the application of data mining in the steel-making industry, (3) knowledge graph and graph representation learning; and (4) the studies of information fusion.

In **Chapter 3**, a framework is proposed for data mining in steel-making under 4.0 with a focus on fault diagnosis. There are five layers designed in this framework, which are multi-sourced data and knowledge acquisition, data and knowledge processing, KG construction and graphical data transformation, KG-aided modelling for fault diagnosis and decision support for steel manufacturing.

Chapter 4 reports a multi-faceted modelling approach, which can characterise and model a typical production failure of steel-making in a predictive manner using machine learning. In this approach, historical multivariate time-series data of a cold rolling process are extracted in a run-to-failure manner, and a sliding window strategy is adopted for data annotation. Then, fault-centric features are identified from three facets — physics-based approaches, empirical knowledge, and data-driven features. Finally, these features are used as inputs for strip breakage modelling using Recurrent Neural Networks (RNNs), which are specialised in discovering underlying patterns embedded in time-series data. The experimental results using real-world data revealed the effectiveness of the proposed approach.

In addition to the production data in steel-making, **Chapter 5** aims to exploit the domain knowledge and semantic data surrounding the production failure for further modelling. A semantic approach to construct a domain knowledge graph, which serves as a semantic organisation to elicit, fuse, and utilise numerous entities and relationships embedded in manufacturing processes, is proposed. Firstly, multi-source information is extracted and integrated to build a knowledge base containing entities and relations for underlying ontology design. Then, an ontology design framework based on hierarchy structure refinement is deployed to improve the reliability in constructing domain-centric ontologies. Lastly, the missing relations between each couple of entities were inferred based on existing knowledge in KG, with the aim of increasing the number of edges that complete and refine KG. The validation of the proposed approach by a case study is also reported.

In **Chapter 6**, KG is further exploited to facilitate the information fusion for the modelling of multi-faceted concepts in steel making. Based on the construction of the concept-centric knowledge graphs, as stated in Chapter 5, multivariate time-series data is transformed into a temporal graph representation of the data sequence. Then, a Graph Convolutional Networks (GCN) model is applied to extract features from these temporal graphs, and these features are fed into a Temporal Convolution Network (TCN) for fault concept modelling. The experimental results show the merits of the KG-aided fusion approach.

Chapter 7 concludes the thesis, and a summary of its achievements is presented. There is a discussion of the restrictions and future work. As a final note, the main contributions to the body of knowledge resulting from this research are summarised.

1.5 Research Contributions

This thesis makes several contributions to the wider body of knowledge.

1. As part of Industry 4.0, a research framework is imperative for supporting the study of fault diagnosis in steel-making. This framework outlines a technical path towards Industry 4.0 levelled fault diagnosis of steel-making based on multi-sourced data in steel-making and existing domain knowledge.

2. Fault diagnosis in a predictive manner is important to quality management in steel-making. With the aim of establishing a production failure prediction model, a multi-faceted data-driven approach is proposed to integrate sliding windows and deep learning techniques. Prediction of the failures can bring significant benefits to the cold rolling industry in terms of contingency mitigation and quality improvement.

3. Since the semantic data and existing domain knowledge also can have an impact on fault diagnosis, it is challenging to utilise these resources and exploit them for further modelling. A KG construction framework is proposed as an approach that aims to design a reliable ontology and complete relations in constructing the domain-centric KG for knowledge management in steel-making. Also, this approach can facilitate the exploitation of knowledge for further computational modelling.

4. It has become progressively more evident that a single data source is unable to comprehensively capture the variability of a multi-faceted fault. Meanwhile, KG can aggregate the relationships of multiple aspects by semantic associations, which can be exploited to facilitate multi-faceted modelling. In this case, a KG-aided data fusion approach is proposed for multi-faceted modelling.

Chapter 2 Literature Review

2.1 Introduction

As discussed in the preceding, this chapter reviews the related works and previous relevant research regarding five main sections: cold rolling, data mining applications, KG and its applications, and information fusion. The cold rolling in the steel-making process was examined through three main aspects, including the steel-making process, cold rolling process, and strip breakage. The relevant overview and studies were demonstrated in Section 2.2. Section 2.3 reviewed data mining and its applications in the steel-making industry. Concisely, as the core part of the steel-making industry, the specific tasks and the applied techniques were involved in this section. In Section 2.4, the relevant studies of KG were introduced concerning KG construction and graph representation learning. Section 2.5 investigated the studies on the strategies and KG-aided techniques under an information fusion context, and Section 2.6 summarised this chapter.

2.2 Steel-making Process

As a fundamental industry sector, the steel industry is of great importance to the economy. A number of important industries rely on the production of iron and steel by providing raw materials, making it one of the largest industries in the world. Meanwhile, cold rolling in the steel-making industry is recognised as an important process in the production of electrical steel strips because of its advantages with regard to accuracy, efficiency, and output rate. Presently, cold rolling contributes to the improvement of the properties of steel strips on changes both in the microstructure and thickness of the steel. Since the properties that have been improved include surface smoothness, tensile strength, yield strength and hardness, cold-rolled products usually have superior mechanical properties, small dimensional tolerances and high-quality surfaces (Wu et al., 2021). Furthermore, as science and technology continue to advance, the quality requirements for steel strip products from cold rolling processes are becoming more detailed and demanding. Therefore, it is imperative increasingly to analyse and monitor the quality of cold-rolled products. This section conducted a literature review on the cold rolling process in the steel-making process.

2.2.1 Overview of Steel-making Process

Due to its commercial and emissions significance, the steel industry is an essential element of many national economies, as well as an essential material for the modern world. Among its many applications, it can be found in construction, military and defence, as well as manufacturing (such as automobiles) (Fan and Friedmann, 2021). In 2021, crude steel production exceeded 1,953 million tons in the world, representing an increase of 3.8% over 2020 (WorldsteelAssociation, 2022).

Generally, a multi-stage process of steel making that transforms iron ore, scrap, and other input factors into steel products, such as plates and tubes, can be roughly divided into three phases (Missbauer et al., 2009). In the first stage, the production of molten iron (termed hot metal) is formed from iron ore, coke and a fluxing agent, which refers to ironmaking. Subsequently, the steel with a well-defined chemical composition is prepared from the hot metal, followed by solidifying the steel into cuboids (termed slabs). It is referred to as steelmaking-continuous casting. The final stage is to conduct the production of finished products through a variety of processes, such as hot rolling, cold rolling etc.

In terms of ironmaking, three dominant production processes contributed to 99.6% of steel hot metal production: blast furnace-basic oxygen furnace (BF-BOF), electric arc furnace (EAF), and direct reduced iron (DRI) (Fan and Friedmann, 2021). Iron ore is reduced to pig iron in the blast furnace, which is the predominant steel production route in the iron and steel industry. It is necessary to charge hot iron into a basic oxygen furnace in order to manufacture hot steel metal (BOF steel making). EAF utilizes an electric arc to heat materials, such as pig iron, steel scrap, and DRI. In contrast to a BF-BOF plant, EAF steel production takes place in a batch mode rather than in a continuous process. By using DRI, iron ore is directly reduced to a solid state at a temperature below the melting point of iron. Despite the fact that DRI production is more energy efficient than pig iron production from BF, additional processing (typically EAF) is required to upgrade DRI sponge iron for the market. It is important to note that these processes operate with different feedstocks. Raw iron ore is converted into pig iron and then into hot steel metal through the BF-BOF pathway, whereas steel scrap and sponge iron are converted into steel hot metal through the EAF pathway. Sponge iron is created by DRI by converting raw iron ore into porous, permeable, and highly reactive sponge iron that must be treated with EAF before it can be sold.

As a consecutive casting process, continuous casting is also called strand casting. The molten steel is injected into a mould and sprayed onto the surface, and then solidified by a cooling process alongside the caster (Lee et al., 2020). As part of the entire manufacturing process, it consists of the production of intermediate materials, such as blooms, billets, and slabs, before some additional processing and final steelmaking procedures are carried out (Wang et al., 2005). This process is widely used for manufacturing intermediate casting products due to its excellent characteristics, including massive productivity, high quality, and cost-effectiveness (Ha et al., 2001, Louhenkilpi, 2014). The most representative of this process is the slabs used to make steel plates (Song et al., 2019). As thick plates, slabs are produced in thick-plate mills or hot-rolled steel plates in hot-rolling mills.

After the steelmaking-continuous casting process, various processes are conducted to gain finished products in the steel industry, such as the rolling process, pickling, annealing process etc. The Rolling of steel is one of the most important manufacturing processes in the steel-making industry (Oduguwa and Roy, 2006, Altınkaya et al., 2014). Usually, it is the first step in the manufacturing process of steel after it is made, either as an ingot or continuous cast product, in a steel melting shop. It has been widely used with its products, accounting for more than 90% of steel plastic processing because of its lower cost and high productivity (Hu et al., 2021b). Rolling is the process of plastically deforming steel by passing it between rolls. The rolling process involves reducing the cross-sectional area of the steel piece being rolled or forming the steel products in general through the use of rotating rolls. The friction between the rolls and the surface of the steel workpiece during rolling causes high compressive stresses on the steel workpiece. As a result of the compressive forces between two continuously rotating rolls, the workpiece is plastically deformed. Consequently, these forces reduce steel thickness and alter its grain structure. Steel stock is passed through one or more pairs of rolls in order to reduce its thickness, make it uniform, and impart a desired mechanical property to enable the steel-making process. By rotating the rolls, the metal is continuously deformed, resulting in a change in size and shape while simultaneously improving its structure and material properties. It is generally accepted that rolling is classified according to the temperature at which the metal is rolled, including the hot rolling process and cold rolling process (Kumar et al., 2019). A hot rolling process occurs when the temperature of the metal exceeds its recrystallization temperature. Unlike hot rolling, a cold rolling process occurs when the metal is rolled at a temperature below its recrystallization temperature. Normally, cold rolling, as an essential process in the metal processing of sheets and strips, is implemented after a hot rolling process.

Specifically, regarding the industrial partner, the process used in the production of electrical steels is illustrated in Figure 2.1. After casting into ingots, the steel is reheated to around 1400°C and then rapidly cooled after hot rolling. The steel is side trimmed, annealed, descaled, and pickled to refine the metallurgical structure of the hot-rolled coil and make it suitable for cold rolling. The coil is reduced to an intermediate thickness of around 0.6 mm and is returned to the anneal and pickle line for an intermediate anneal. After a final cold reduction, the material is brought to the finished gauge of 0.23 to 0.50 mm. The decarburising anneal line removes relatively high levels of carbon by annealing in an atmosphere of moist hydrogen and nitrogen at about 840°C. The HTCA process is carried out in an atmosphere of dry hydrogen at about 1200°C for 4-5 days. This produces secondary recrystallisation of well-oriented grains, and MgO on the steel surface reacts with the silica and Fayalite to form an electrically insulating glass film made mostly of Forsterite. The strip is washed to remove unreacted magnesium oxide powder, coated with a phosphate solution, and cured at a temperature of approximately 800°C. The edges of the coil are trimmed before the coil is sent to be packaged or slit into several widths.

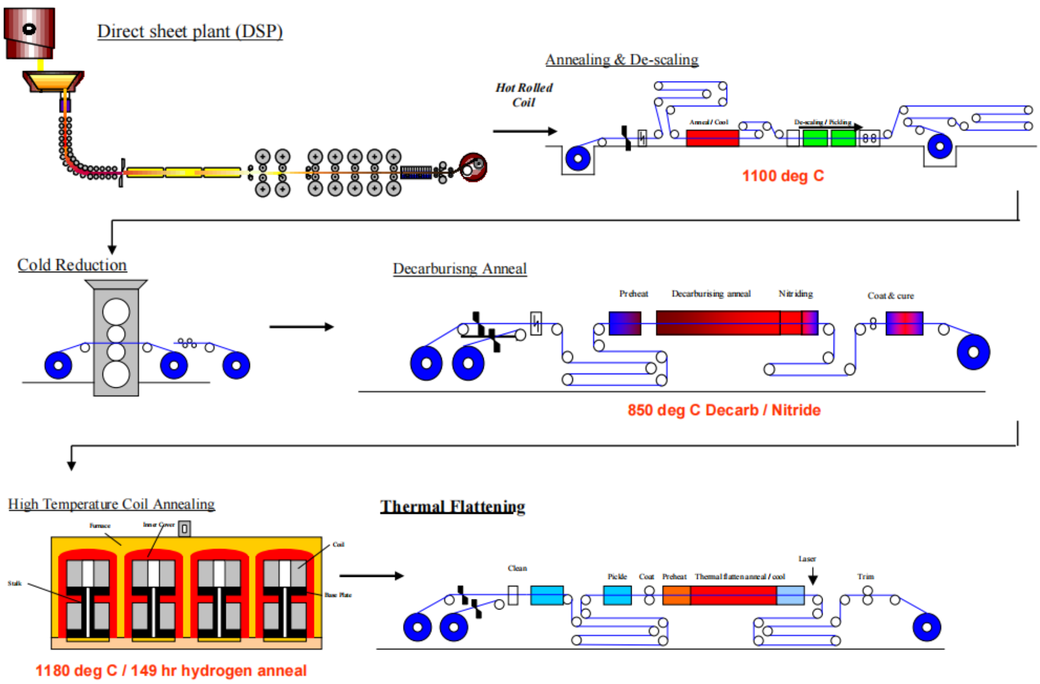

Figure 2.1 A schematic of manufacturing process for electrical steel (Provided by Cogent)

2.2.2 Cold Rolling Process and Typical Faults

Cold rolling is an essential process in the steel processing of sheets and strips due to its high accuracy, efficiency, and production rate (Hou et al., 2007). Normally, cold rolling can be conducted using a single stand in a reverse manner or continuous stands (Mashayekhi et al., 2011). As one of the primary metal forming processes in the production of steel strips, cold rolling is implemented to decrease the thickness of steel and strengthen its pliability and properties of yield strength and tensile strength. Specifically, compared to hot rolled products, cold rolled products offer the following advantages: greater dimensional accuracy, hardness improved by up to 20%, increased yield strength, increased tensile strength, improved surface finish, improved straightness etc. (Takami et al., 2011a). In consequence, the cold rolling process requires special attention since the improved mechanical properties increase the value of the final product, whether it is a sheet, a strip, or a coil.

Cold rolling processes can encounter certain defects with regard to the final product. Technical reports indicate that various types of defects exist in steel strips from cold rolling production. The most common defects encountered in the sheet metal rolling process include edge cracking, burrs in the centre, surface defects, and buckling. Especially electrical steel is an iron alloy containing high percentages of silicon. Alloys containing a high silicon content have a low magnetisation loss as a result of the high electrical resistivity. As a result of a high silicon concentration, the strip becomes brittle, resulting in breaks during cold rolling. As the most serious defect, strip breakage needs to be paid special attention to as it has severe consequences. Specifically, strip breakage has damaged rolls and mill accessories badly, not only an increase in production costs (Johnson and Mamalis, 1977). Hence, various studies have been conducted on strip breakage in cold rolling.

Research on strip breakage has typically been conducted in a retrospective manner which focuses on root cause analysis. By summarising related studies on the cause analysis of strip breakage, the causes can be concluded into four different facets as follows. The first type of possible breakage cause is material-related issues. The hot-rolled coil (HRC) is the feedstock of the cold-rolling process. The undesired physical or chemical properties of HRC can result in a breakage (Hongzhao et al., 2010). To be specific, previous work has discovered that there is a higher possibility for coils to break if any non-metallic material, such as protective slag or oxide scale, is included in steel during the hot rolling process (Runchang et al., 2017). For the impurity of the strip, it was proved that the impurity has a negative impact on the homogeneity of the steel strip, which can contribute to a breakage (Rusnák et al., 2020). Another work discovered that the material hardness and hardening through the deformation of cold rolling have an impact on yield stress, which is an essential parameter when considering breakage (Ilin and Baranov, 2020). In terms of the incoming HRC, apart from the chemical and physical properties, the surface condition, shape, and flatness of the strip derived from the roll gap model are the potential causes of strip breakage as well (Wang et al., 2017). Other than HRC, the emulsion, which acts as the coolant and lubrication, also plays an important role in the occurrence of strip breakage according to the friction model, which describes the friction between the roll surface and strip using parameters such as strip speed, roll and strip surface roughness and lubrication (Tan et al., 2008). Moreover, the conditions of stability and reliability of the hot rolling process are also proven to be influential with possible strip breakage (Mazur and Nogovitsyn, 2018). Specifically, for electrical steel, the strips are annealed and pickled before cold rolling. For this process, hot-rolled coils anneal, followed by water quenching to control the precipitation of grain growth inhibitors. In addition, shot blasts and pickles will be conducted to remove the scale of the strip, which will have an impact on the strip surface condition as well (Ros-Yáñez et al., 2004).

Secondly, equipment malfunction, especially in the rolling mill, is proved to be another facet of breakage causes. In a previous case study (Yan and Li, 2006a), the strip was broken and crushed to the other side due to an inter-frame tension deviation resulting from mill malfunction. Another research (Cui and Zhao, 2013a) discover that the levelness and verticality of the steering roll of the uncoiler and the piston rod elongation of the hydraulic gauge control (HGC) system are potential causes of strip breakage. In addition, strip breakage can be caused by an unexpected high servo valve adjustment resulting from the defects of the backup roll bearing (Xu, 2015). Under this unexpected adjustment, the pressure fluctuations on both the entry and exit sides are different, which results in tension deviation, which is a significant cause of breakages.

Thirdly, rolling operations such as inappropriate parameter settings were analysed to be the representative causes for strip breakage in some recent works. It has been concluded that the opposite effect will occur with the reduction of rolling speed, which can increase the risk of strip breakage (Rusnák et al., 2020). In a related study (Xu, 2015), the authors discovered that an inappropriate tension match between the entry and exit sides leads to a large deformation on one side of the strip. Another study discovered that inappropriate tension and roll separating force setting caused by unreasonable HGC control is the main cause of breakage (Xiao, 2013). Moreover, the variation of specifications such as maximum gauge, width and yield stress should be compensated during the rolling operation; otherwise, breakage is more likely to occur (Han et al., 2018).

Fourthly, the rolls, such as the work rolls and backup rolls of the mill, are proven to be relevant with breakage in related works (Takami et al., 2011a, Song et al., 2018, Cui and Zhao, 2013a). The roll wear applies an adverse effect on the shape of strips, which can further result in strip breakage (Takami et al., 2011a). According to the roll wear model, which calculates the time-dependent thermal contours of the rolls (Song et al., 2018), the roll contour and roll temperature have an impact on the roll wear. In terms of the roll contour, the bending model, which describes the roll bend, roll contour and flattening between the work roll and support roll, should all be taken into consideration for the calculation of bending. In addition, both the convexity degree and diameter disparity of the work rolls have been discovered to be possible causes for strip breakage (Cui and Zhao, 2013a). Another research discovered that imperfections of the backup rolls and working rolls could result in uncontrolled mill resonance, which is the main cause of strip breakage (Gérard et al., 2007).

2.2.3 The Industrial Partner and Its Manufacturing under Industry 4.0

Cogent Power is a global supplier of electrical steels for large rotating machines, generators, and transformer cores. As a wholly-owned subsidiary of Tata Steel, the company has manufacturing and service operations in the UK, Sweden, and Canada. Cogent Power is a world-renowned brand in electrical steels and is recognized as a leader in the manufacture of thin-gauge electrical steels for use in high-frequency machines. In Section 2.2.3 of Chapter 2, a detailed summary is provided of the company's I4.0 maturity, challenges, requirements, and constraints. Specifically, the company has made investments in I4.0 technologies, including the installation of a Supervisory Control and Data Acquisition (SCADA) system, an Operational Control System (OCS), and a variety of sensors for real-time data collection and analysis. However, the company faces several challenges related to the implementation and adoption of I4.0 technologies, such as the integration of data from multiple sources and the complexity of the steel-making process. In terms of requirements, the company has identified the need for scalable and adaptable solutions, data security and privacy, and effective communication and collaboration across different teams and departments. These requirements and constraints are critical considerations for the successful implementation of the proposed approach for fault diagnosis and prediction in the steel-making process.

The concept of Industry 4.0 (I4.0) is highly relevant to this work on fault diagnosis and prediction in the steel-making process, as it provides a framework for leveraging advanced technologies such as cyber-physical systems (CPS), the industrial internet of things (IIoT), and data analytics to improve the efficiency and effectiveness of manufacturing processes. By using these technologies to collect and analyze data in real-time, steel-making companies can gain greater visibility into their production processes and make data-driven decisions to optimize their operations. In the case of the steel-making industry, this is particularly important due to the complex and multi-faceted nature of the manufacturing process, which can make it difficult to identify the root causes of production failures. By leveraging I4.0 technologies, we can harness the power of data to better understand the steel-making process and improve our ability to diagnose and predict production failures such as strip breakage in cold rolling.

The industrial partner of this research has installed a Supervisory Control and Data Acquisition (SCADA) system, an Operational Control System (OCS), and a variety of sensors to acquire data from the GOES material and processing units prior to cold reduction and the cold reduction process. These hardware components, combined with software tools for data management and analysis, provide the company with a robust infrastructure for real-time data collection and analysis. This means that the company is well-positioned to benefit from our proposed approach for fault diagnosis and prediction, as it has already invested in the necessary hardware and software to support data-driven decision-making. Additionally, the company's strategy for I4.0 involves leveraging data analytics and machine learning to improve the efficiency and effectiveness of its operations, which aligns closely with the goals of our research.

Given the hardware and software of the industrial partner, the data used in the study was provided by a cold-rolled silicon electrical steel manufacturing line. The data was collected from four main sources across the line, including hot rolled coils (HRC), annealing and pickling (A&P), emulsion, and cold rolling process (CRP). The HRC data contained critical information on chemical properties and strip shape parameters, while the A&P data provided details on the physical properties affected by annealing and the surface condition affected by pickling. The emulsion data recorded the lubrication and cooling effects on the strip-roll friction and thermodynamics. The CRP data provided direct and real-time measurements of the operations before the occurrence of breakage. The collected data included labels indicating whether each coil was either "break" or "good." To be specific, the HRC data consisted of 47 variables that recorded the physical and chemical properties of each incoming feedstock hot-rolled coil. The A&P data comprised 18 variables that recorded the real-time annealing and pickling process on each incoming hot rolled coil at a frequency of 50Hz. The emulsion data were recorded daily in the steel plant and contained eight variables. The CRP data was extracted from a production data acquisition (PDA) system installed on-site. Cold rolling process variables were continuously sampled and recorded at a frequency of 100Hz.

In order to ensure a comprehensive and accurate analysis of the data, the expertise of domain experts was incorporated into the research process. These experts were consulted for their knowledge in feature identification, modelling of strip breakage prediction, and construction of domain ontology. By considering the specific domain knowledge of the dataset, a combination of physical-based and empirical knowledge features was identified and selected for further study. This collaboration with domain experts helped to ensure that the research was grounded in established principles and best practices within the field.

2.3 Data Mining Applications in Steel-Making Industry under Industry 4.0

The rapid growth in data mining has led many industries to use it to discover hidden patterns in their systems, which could then be used to design new models to enhance production quality, productivity, optimum cost and maintenance. Additionally, it is essential for steel producers to continuously improve their steel production processes in order to avoid quality deficiencies and increase production yields (Iwasaki and Matsuo, 2011). Consequently, data mining has become a useful tool for acquiring knowledge in the steel industry (Umeshini and PSumathi, 2017).

2.3.1 Tasks of Data Mining in Steel-Making

Manufacturing organizations are required to employ a variety of techniques and tools to achieve their foundational goals in the steel industry. Data mining is regarded as an excellent solution to address this challenge. Data mining is defined as a technique to find patterns or interesting information in a large amount of data (Faleiro et al., 2013). It is a mathematical method and technique for solving problems through the analysis and evaluation of data that has already been gathered and stored on a computer system. In steel-making, it is well established that data mining techniques are used because intelligent analysis of data can lead to valuable insights and provide a competitive edge.

Concisely, the process of data mining should be followed in a few steps in order to achieve optimum results. The first step is to gain an understanding of the background of a particular field. A second step that needs to be carried out is to perform an in-depth analysis and understanding of the data. Thirdly, it is important to prepare the data for the missing data so thereby it can be used in the data mining process. The fourth step is to determine which method will be the most effective for achieving the data mining goal so that a better result can be obtained. Subsequently, the method will be tested and compared with the previous data in order to evaluate the method. In the final step of the deployment process, all of the results and progress will be reviewed, and a final report will be produced detailing how the entire process went (Witten and Frank, 2002, Villacampa, 2015).

In general, tasks of data mining in steel-making can be categorized into four main categories: scheduling task(Tang et al., 2001), monitoring task (Cemernek et al., 2022), quality task (Li et al., 2018a), and failure task (Chen et al., 2021b). Steel-making scheduling refers to the process of organizing, controlling and optimizing production and manufacturing in the steel industry. In this process, plant and machinery resources are allocated, human resources are planned, production processes are planned, and materials are purchased. The purpose of the monitoring task is to avoid key performance indicator (KPI) value deviations and to increase the visibility of manufacturing systems in the steel-making process, including decision support systems (DSS) and process monitoring (Qu et al., 2017, Syafrudin et al., 2018). A quality task can be described as a process for predicting and improving the quality of steel-making products, concerning quality monitoring and quality diagnosis etc. In failure tasks, abnormal situations or faults are detected and predicted in steel-making manufacturing, such as product failures, equipment failures, and process failures.

Production scheduling is used to gain the processing timing and sequence in each manufacturing process while ensuring that manufacturing conditions and delivery dates are met. Through a variety of processes, including blast furnaces, converters, continuous casting, rolling, annealing, and surface treatment, steel products are manufactured from iron ore, coal, and other raw materials in response to the demands of customers in various industries. In the specification of a product, there are a series of requirements which are based on the use of the product, such as the quality and strength of the materials (such as strength and toughness), the grade of the inside and outside surfaces of slabs, and the size (such as thickness and width). Depending on the product type, specifications can range from several thousand to tens of thousands. Additionally, a number of factors contribute to the manufacturing conditions of steel-making products, including molten steel components, rolling size, annealing temperature, and plating type. As with product specifications, there is a wide variety of manufacturing conditions. Particularly, large-lot production, during which steel-making products are continuously manufactured under the same conditions, has proven to be advantageous from a quality and cost point of view, resulting in the type of production that has been targeted. It has been explained above that conditions for manufacturing products vary from process to process, as well as delivery dates. In this context, it is necessary to examine scheduling tasks while balancing quality, costs, and delivery dates in steel-making manufacturing. Specifically, quality, costs, delivery dates, and other performance metrics must be considered in a comprehensive manner (Ito et al., 2019). Therefore, various technologies for supporting steel-making scheduling were developed in various manufacturing processes, such as raw materials, steelmaking, hot rolling, and logistics. For optimizing oxygen and nitrogen usage in the steel industry, a two-stage predictive scheduling method was proposed, utilizing a long-term prediction model based on Granular Computing to predict the requirements of these two gas. The results of the experiments and the online application demonstrated that the proposed method is capable of providing satisfactory prediction accuracy and scheduling performance (Han et al., 2016). For the Linz–Donawitz converter gas (LDG) system, a scheduling approach based on a three-layer causal network was presented regarding generation and consumption uncertainties (Jin et al., 2021). A model for optimizing the operation of a gas–steam–power conversion system was developed. This model is designed to analyse the by-product gas system of an iron and steel enterprise, taking into account the operational status of equipment, fluctuations in the cost of gas holders, the cost of fuel, the income generated by external power transmission, and the cost of environmental pollution (Hu and He, 2022).

Increasing customer demands and highly competitive conditions force steel companies to develop innovative and flexible concepts in order to remain competitive. Particularly in, the steel industry is technology-intensive; even the slightest variation in production can lead to expensive and time-consuming post-processing. The result will be a direct reduction of costs in addition to any improvements in quality. Therefore, continuous monitoring, control, and assessment of the implementation are required to achieve this improvement. In the steel industry, a number of studies are providing the foundation for continuous monitoring and control of highly complex and high-dimensional processes. Based on self-organizing maps, a new method for state monitoring and strip quality prediction was proposed in a hot rolling process (Cser et al., 1999). A detecting model for clogs in continuous steel casting was proposed, which can be used to eliminate the effect of confounding disturbances. Using a causal graphical model that incorporates field knowledge, it exploited statistical independence and invariance properties. The experimental results indicated the effectiveness of the casual graphical model based on real-world cases (Yang et al., 2021).

There is a great deal of responsibility placed upon the quality control of products in the steel industry, which is determined by the characteristics of the industry. The steel industry is a process-oriented industry, where each production line is continually produced during the production process. Moreover, the production scale of each production line is relatively large, which means that a whole batch of products would be affected by similar quality problems if a quality problem occurs. If this occurs, severe losses will be incurred by the economy. Hence, a better quality-control solution is needed in the steel industry, which has recently been referred to as intelligent manufacturing (Kano and Nakagawa, 2008). For example, data mining technology was applied to quality management, including quality control and quality analysis. By using the rough set attribute reduction theory, an improved Apriori algorithm was proposed for identifying quality association rules. It combines the merits of rough set and association rule excavation methods, allowing users to gain non-redundant decision rules whose confidence and support valve values can be pre-determined (Cai-yan and You-fa, 2009). In steel manufacturing processes, where environmental conditions can influence image quality, quantum machine learning technology was applied in surface quality supervision as a branch of quantum computing. Compared with conventional deep learning, there is substantial potential for utilizing this technology in application cases, primarily due to the speed of the physical quantum engine (Villalba-Diez et al., 2022).

Steel manufacturing is characterized by extreme working conditions, such as high temperatures, high pressure, speedy production, and high throughput. Since the overall production process involves a high level of economic and energy investment, an intensive and costly preventive maintenance program is needed to prevent breakdowns. It would be beneficial for the steel-making process to have a predictive maintenance module which can detect incoming faults through the analysis of data. By using two different methodologies (static and dynamic), a two-step scheme was designed to detect faults in rolling mills in steel production plants. Firstly, a preliminary fault detection phase was conducted to detect faulty samples, which is regarded as an effective and efficient computational approach. Following the robust distances obtained in the previous phase, an additional method is applied to confirm the fault detection and provide additional information on the probability of transitioning between latent states. Case studies proved the effectiveness of the proposed methodology in the experiment section (Sarda et al., 2021). A novel analysis of variables in a cold rolling mill was presented using statistical, computational and numerical methods. The co-variation and correlation of variables were determined using principal component analysis (PCA) (Takami et al., 2011b). A study compared visual defects-detection techniques in the steel industry. According to the fundamental concepts of image processing, detection methods are classified as statistics, filtering, modelling, and machine learning. Various approaches are described, along with their fundamental concepts, benefits, and disadvantages, so that researchers can select the best approach for their particular application (Mordia and Verma, 2022).

2.3.2 Data Mining Techniques Applied in Steel-Making

In the application of data mining, data-driven approaches perform significantly on data analysis and pattern discovery. Computers can model based on experience and accurately predict future events with data-driven approaches. The success and the associated rise of data-driven techniques in recent years have led to the deployment of a variety of tools and techniques to address challenges in steel-making. Presently, in this data-rich environment, data-driven approaches to investigating steel-making have previously been applied in a handful of works (Chen et al., 2019, Takami et al., 2011a, Wang, 2014b).

Generally, depending on different theories, data mining techniques applied in steel-making are classified into three main types: supervised learning, unsupervised learning, ensemble learning and deep learning. The three types of data mining techniques have been detailed below.

A supervised learning process is defined to learn the mapping between inputs and outputs. Usually, with only one output variable, many input variables are used in a supervised algorithm. It is logical to conclude that the number of samples available for learning will proportionally affect the predictability of a supervised learner. There are two types of supervised learning in steel-making: classification and regression. In general, classification is used to predict discrete or nominal or categorical values, whereas regression is used to predict continuous or numerical values. Various algorithms are available for serving these purposes with their own advantages and disadvantages, including decision trees (DT), neural networks (NN), support vector machines (SVM), and naive Bayes (NB). For example, to fulfil the requirements for accuracy and efficiency, an SVM-based predictor for the by-product gas flow in the steel industry was built by real-time optimization for the width of the Gaussian kernel and the regularization factor (Zhao et al., 2012). Neural networks were introduced to predict molten steel temperature in a continuous casting process (Sobaszek et al., 2017).

In data-driven techniques, unsupervised learning identifies patterns and dependencies in unlabelled data by identifying regularities. It is the goal of all the tasks mentioned above to produce an effective representation of the inner data structure without explicitly labelling it. In steel-making, it is highly possible to encounter data with class labels, which is why unsupervised learning studies are relatively few. The most representative unsupervised learning methods include clustering and association rule mining. According to their similarities, clustering divides instances into different groups. Depending on the degree of similarity or distance between the data, clusters are formed based on similarity or distance measurements. A rule-based machine learning method called association rule learning is used to discover interesting relationships between variables in large databases. With the use of some measures of interestingness, it aims to identify strict rules that have been discovered in databases. A set of association rules is used in specific domains involving a wide variety of items to discover how or why certain items are related. For example, a DC-ML model based on clustering was proposed to classify samples collected from a steel plate rolling mill, and then the datasets were fed into supervised learning models. Experimental results illustrated the superior over four supervised models, including RF, gradient boosting regression (GB), gaussian process regression (GP), and conditional linear gaussian (CG) (Park et al., 2020).

By analysing all types of highly validated steel-making research studies, it is found that ensemble learning and deep learning are employed mostly to enhance the performances of data-driven models. An ensemble learner consists of a group of machine learning methods that combine a committee of classifiers in order to perform a classification or regression exercise. In the case of homogeneous ensemble learners, the same algorithms are applied in different arrangements that form the committee, or different training sets are generated from the original dataset. In contrast, its heterogeneous counterpart consists of various types of classifiers. As an alternative to ensemble learning, deep learning is a technique that processes data in several connected layers, where its structure is non-linear and complex. For example, in order to predict casting quality, a weighted random forest (WRF) algorithm was developed based on the analysis of the relationship between multiprocess parameters and casting billet quality. By weighting the decision tree results, this algorithm effectively solved the sample imbalance problem and accurately identified negative samples. As a result of real-time billet data collected during the casting process, the proposed method was proven to be effective (Ye et al., 2018). In order to predict productivity at the granularity, a multivariate and multifrequency Long Short-Term Memory model (mmLSTM) was proposed. The mmLSTM model incorporates equipment status and order as new supporting factors and utilizes a multivariate LSTM to determine their relationship to productivity. In addition, the mmLSTM model incorporates a multi-level wavelet decomposition network to capture the multi-frequency features of productivity in a comprehensive manner. An evaluation of performance using productivity data for nearly two years is conducted using the proposed method in a real-world steel factory. It has been shown that our method is effective in improving the accuracy and granularity of industrial productivity prediction (Zhang et al., 2020b).

In recent years, ensemble learning and deep learning have achieved great success in a wide range of applications due to their better generalization. Due to its ability to improve the performance of weak learners, ensemble learning has become increasingly popular in steel-making over the past few years. Compared to ensemble learning, deep learning is more frequently used in steel-making. In addition to their separate implementation, some studies have also combined ensemble and deep learning in order to compare their strengths and weaknesses. Due to its significant contributions to performance improvement, deep learning studies in steel-making are exponentially increasing.

Although the above three types of techniques have already been extensively used in the steel industry, which accounts for most of all applications, graph representation learning has received less attention. By transforming the knowledge into latent vector space representations, graph representation learning enable an easy way to discover useful patterns in complicated and interconnected data and improve the modelling performance by introducing connected features (Myklebust et al., 2019).

2.4 Knowledge Graph and Graph Representation Learning

As the latest technology in AI to represent and organise knowledge, KG is a unique type of content retrieval method for graph databases, which was proposed by Google in 2012. At present, KG has been energetically developed as it elicits, fuses, and utilises numerous entities and relationships embedded in manufacturing processes and products. Furthermore, KG displays a powerful expressive ability and a high degree of modelling flexibility, making it a promising semantic network (Paulheim, 2017). In this context, KGs have been utilised widely in many scenarios, such as knowledge recommendation and knowledge visualisation. It has become a foremost objective of the research on KG. In this section, state-of-the-art literature on KG is summarised, focusing on KG construction and graph representation learning.

2.4.1 Knowledge Graph Construction

KG has been referred to as a graph model to represent information in an understandable way. It is composed of interconnected sets of entities in which different entities are connected using semantic links. As a factual reflection of human knowledge, it is now widely accepted that knowledge graphs are useful in solving various domain-specific problems in industry and academia (Ji et al., 2021). It has been shown that the KG paradigm can be applied to a wide variety of domains due to its incorporation of graph technology and the availability of an abundance of graph datasets, thus making KGs applicable to a variety of problems in a variety of different areas (Abu-Salih et al., 2021). KGs are typically divided into domain-specific KGs and general-purpose KGs. In addition to containing high-quality domain-specific knowledge, KGs provide substantial benefits for tackling domain-specific problems and maximising the value of domain corpora (Kejriwal, 2019). Over the years, continuous efforts have been made to develop KGs that capture various domains of knowledge, and the generation of KG using ontologies has gained considerable popularity.

Depending on the application scenarios, KGs have been divided into two types normally: general KGs and domain KGs. For general KGs, it is important that the information is not limited to a particular field. The information on these KGs is comprehensive, reasonable, and common sense. However, in terms of general KGs, it has been emphasised that the knowledge requirements need more broad than precise (Chen et al., 2021a). For example, the general KG has been usually constructed as an intelligent search engine to achieve the question-answering system on the Web. At present, beyond the generic KGs, the majority of KG research has focused on the construction of domain-specific KGs regarding certain ontologies in their fields (Abu-Salih, 2021, Singhal, 2012, Li et al., 2020b). Unlike general KGs, domain KGs are regarded as vertical KGs, which describe specific particular domains (Zhou et al., 2021). Although the description scopes of domain KGs are very limited, the depth of knowledge has been emphasised in building a given domain KG (Zhao et al., 2019, Kejriwal, 2019). In this context, domain-specific KG provides a promising mechanism to fuse more sophisticated knowledge and structure from multiple sources.

The construction of KGs involves an iterative engineering process that can be applied to many different techniques and tools. Existing approaches can be grouped into two categories: top-down and bottom-up. The most widely used KG construction is the top-down method, which originates from the modelling process in database construction. Generally, five steps are implemented to construct KG for a specific scenario. The first step is to identify a subject domain, followed by a list of research requirements. In the second step, a conceptual model will be developed in order to gather the entities of interest, their interrelationships, and the categories. Furthermore, by using logical and physical models, entities and relationships will be represented logically, and assertions will be made about those entities and relationships. A fourth consideration is the appropriate coding language (for example, RDF and OWL), serialization formats (for example, RDF/XML, Turtle, and JSON-LD), and the KG development platforms (for example, Protégé and DOGMA). The last step in the development process is to deploy the KG as a service so the community can utilize it and provide feedback.

Designing an ontology is the most vital aspect of constructing a knowledge graph (Pietranik and Nguyen, 2014, Dou et al., 2018). Ontology is defined as a model describing structured and unstructured information through entities, properties, and the way they relate to each other. As a semantic model, ontology can define and describe a wide variety of entities, features and properties existing in a specified domain (Otero-Cerdeira et al., 2015). Therefore, ontology is expressed as a formal representation of knowledge by a set of concepts within a domain and the relationship between these concepts. In other words, the knowledge graph is the manifestation of the ontology of the specific content. Even though there are differences between ontology and knowledge graph, an ontology that serves as a framework to model the content of multiple data sources can be applied to create a knowledge graph. Recently, ontology has been used as a solid tool to construct a knowledge graph in a lot of works. For example, a large-scale knowledge graph about drug-drug interactions was built on the corresponding ontology with data from multiple sources (Karim et al., 2019). A research work combined computer vision algorithms ontology to construct a knowledge graph that can automatically detect hazards to address the ‘semantic gap’(Fang et al., 2020). Four large and heterogeneous ontologies were applied to build a knowledge graph along with storing the data and semantics in the biomedical domain (Cardoso et al., 2020). An ontology for risk management of hazardous chemicals was designed to construct KG in the chemistry industry (Zheng et al., 2021).

In addition, information in KGs generally constructed over raw data is incomplete and erratic, such as missing information. It is essential to refine or complete such KGs before they can be applied in real scenarios properly (Paulheim, 2017, Pujara et al., 2013). In this context, entity and relation completions have been proposed as a process of completing KG by discovering and adding the missing and implicit entities and relations. For example, existing knowledge is utilised to infer potential relations between each couple of entities for KG completion (Wang et al., 2021c, Wang et al., 2021b, Ren et al., 2022). In other words, relation completion augments KG in some respects by increasing the edge number in order to enhance reliability.

As mentioned previously, KG has been referred to as the advanced technology of knowledge representation, which provides an important way to integrate multi-source information. However, the construction of KG in industrial scenarios has certain limitations. For instance, developing a reliable KG requires extensive participation from domain experts, which is a time-consuming and labour-intensive process. Moreover, the construction of KGs from raw data can also be inadequate in situations where information is missing, such as relationships that are not included in the KG.

2.4.2 Graph Representation Learning

On the basis of the integration of multi-source information, constructing knowledge graphs has allowed obtaining the collective and comprehensive information for modelling multi-faceted phenomena considerably in many industrial applications. In this way, real-time capabilities of multi-faceted modelling can be achieved with the consideration of connections and relations across different facets. For multi-faceted phenomenon modelling, machine learning is one type of the most prevailing techniques. However, KG, which represents structured data effectively, is challenging to manipulate by conventional machine learning algorithms due to such triples' underlying symbolic nature (Chen et al., 2017). When it comes to inputs of conventional machine learning models, they usually take feature vectors representing objects in terms of tabular attributes (such as numeric attributes and categorical attributes) as inputs (Nickel et al., 2015). In this context, conventional machine learning approaches, such as Convolution Neural Networks (CNN) and Recurrent Neural Networks (RNN), cannot be directly applied in the graph domain.

Therefore, it is challenging to propose an expressive and informative representation of the graphs to bridge these semantic gaps. In other words, given the knowledge graphs, the representation learning on these graphs is another challenging problem. Firstly, the number of nodes in a graph can be variable, which poses a great challenge for traditional machine learning models that can only have fixed-sized input. Secondly, the graph has the isomorphism problem, meaning that the same graph can have factorially many different expressions by simply permuting the nodes, which brings additional challenges to distinguishing graphs. Thirdly, the graph topology contains rich information important for the learning tasks, yet it is extremely hard to extract and learn. Hence, it has been developed various graph representation learning techniques that convert raw graph data into a high-dimensional vector while preserving the intrinsic properties of the graph. There is also a term for this process called graph representation learning, which is becoming a hot topic and a challenge in recent years.

Techniques of graph representation learning are applied for two purposes: machine-learning tools can be effectively utilized to perform downstream tasks; multi-faceted modelling can be achieved directly to discover hidden patterns. Depending on the different theories, four types of graph representation learning methods were defined and have formed the most prevailing taxonomy, which are dimension reduction-based methods, random walk-based methods, matrix factorization-based methods, and neural network-based methods (Hamilton et al., 2017b, Chen et al., 2020a).

The dimension reduction method is proposed as a method for representing high-dimensional graph data in a low-dimensional representation while preserving the desired properties of the original data. Specifically, a graph with D-dimensional vector space is converted into a d-dimensional vector space, where $d \ll D$. There are two types of dimension-reduction-based methods: linear and nonlinear. The linear approaches are implemented under the linear assumption, including Principal Component Analysis (PCA) (Jolliffe and Cadima, 2016), Linear Discriminant Analysis (LDA) (Ye et al., 2004), Multidimensional Scaling (MDS) (Ye et al., 2004) etc. However, there is a possibility that linear methods will fail if the underlying data are highly nonlinear (Saul et al., 2006). In this case, nonlinear dimensionality reduction approaches have been deployed to learn the nonlinear topology automatically (Saul et al., 2006), such as isometric feature mapping (Isomap) (Samko et al., 2006), locally linear embedding (LLE) (Roweis and Saul, 2000), kernel method (Harandi et al., 2011) etc.

In random walk-based methods, numerous paths are generated in a graph by sampling walks from random initial nodes. The semantic context between connected nodes is demonstrated from these paths. As a result of the randomness of these walks, it is possible to explore the graph and capture the global and local structural information by walking through neighbouring nodes. On the randomly sampled paths, probability models, such as skip-grams (Guthrie et al., 2006) and bag-of-words (Zhang et al., 2010), are employed in order to learn the representation of nodes. Various approaches based on the random walk have been proposed and applied in many specific scenarios, such as DeepWalk (Perozzi et al., 2014) and node2vec (Grover and Leskovec, 2016).

The matrix factorization-based method is also regarded as Graph Factorization (GF). It has been widely used to handle the sparsity of graph data. Specifically, the matrix factorization techniques have been introduced to discover an approximation matrix for the original graph. A graph's adjacency matrix is factorized to gain the embeddings in GF. The process corresponds to a reduction of dimensionality that preserves structure. As summarized, there are several matrix-factorization-based methods and their variants, such as graph Laplacian eigenmaps (Belkin and Niyogi, 2003), node proximity matrix factorization (Singh and Gordon, 2008), text-associated DeepWalk (TADW) (Yang et al., 2015), GraRep (Cao et al., 2015), and HOPE (Ou et al., 2016) etc.

Inspired by the success of CNNs, graph representation learning based on neural networks has been developed, which is termed Graph Neural Networks (GNNs). According to different theories, GNNs are classified into four categories: Recurrent GNNs (RecGNNs), Convolutional GNNs (GCN), Graph Autoencoders (GAEs), and Spatial-temporal GNNs (STGNNs) (Wu et al., 2020). Owing to the promising performance and good interpretability, the application of GNNs is very beneficial and a widely studied topic in many fields for different tasks (Liu and Zhou, 2020), such as knowledge graphs, recommender systems, and visual reasoning. Consistent with different outputs, these tasks are classified into three levels: node level, edge level, and graph level (Wu et al., 2020). In other words, with the graph structure and node content information as inputs, the output of GNNs can focus on different graph analytics with these three levels, such as node classification, link prediction, graph classification etc.

For node labels, a training node augmentation was proposed, which enlarges the training set using the labels predicted by existing GNN models. As self-enhanced GNN improves the quality of the input graph data, it is general and can be easily combined with existing GNN models. Experimental results on three well-known GNN models and seven popular datasets show that self-enhanced GNN consistently improves the performance of the three models (Yang et al., 2020). At the edge level, a novel method for anonymising data, model training, explainability and verification was proposed to predict links in Master Data Management (Ganesan et al., 2020). In terms of graph level, an adaptive structural coarsening module was introduced to produce a series of coarsened graphs and then construct the convolutional network based on these graphs. In this work, a Multi-level Coarsening-based GCN (MLC-GCN) was proposed to learn the graph representations that preserve the local and global information of graphs for graph classification (Xie et al., 2020).

In graph analysis, following the idea of representation learning and the success of word embedding, three different methods (dimension-reduction-based methods, random walk-based methods, and matrix-factorization-based methods) mentioned above apply the different theories to gain patterns of graph representation. Similar variants of them, such as LINE and TADW, also achieved breakthroughs. However, these methods suffer from two severe drawbacks. First, no parameters are shared between nodes in the encoder, which leads to computational inefficiency since it means the number of parameters grows linearly with the number of nodes. Second, the direct embedding methods lack the ability to generalisation, which means they cannot deal with dynamic graphs or be generalised to new graphs (Liu and Zhou, 2020).

Encouraged by CNN and graph embedding, GNNs provide a series of deep learning-based methods to manipulate graph analytics. Compared with the other methods, GNNs learn the representations of a graph both from the general graph structure and node features. In other words, GNNs are designed to learn hidden patterns from graph-structured data in an end-to-end manner that escapes hand-engineered feature extraction (Chami et al., 2020). Especially, GCN tackles this problem by defining a convolution operator on a graph (Kipf and Welling, 2016). The model iteratively aggregates the embeddings of neighbours for a node and uses a function of the obtained embedding and its embedding at the previous iteration to obtain the new embedding. Aggregating embedding of the only local neighbourhood makes it scalable, and multiple iterations allow the learned embedding of a node to characterise the global neighbourhood. Deep learning methods can model a wide range of functions following the universal approximation theorem; given enough parameters, they can learn the mix of community and structural equivalence to embed the nodes such that reconstruction error is minimised. In addition, GraphSAGE has gained recognition in graph representation learning. Compared with the original GCN model, the GraphSAGE algorithm has been proposed as a comprehensive improvement, which is an inductive representation learning for node embedding (Hamilton et al., 2017a). The main idea of GraphSAGE is to adhere to GNN and aggregate the neighbours' information by embedding them into each node.

2.5 The Studies of Information Fusion

Studies on data integration or fusion represent a renewed growing field, which started as a result of the continuous development of instrumental techniques and sensor devices and a paradigm change in the research of complex systems towards holistic, data-driven approaches (Kitchin, 2014, Martens, 2015). On the one hand, the availability of technological development in instrumentation has increased the data acquisition speed, the coupling of different instrumental modalities, their portability, and the data storage capacity successfully, thus giving rise to an enormous explosion of available data for analysis. On the other hand, the paradigm shift is driven by data-intensive statistical exploration and data mining for knowledge discovery (Cocchi, 2019b). Achieving data fusion assumes that the result will be more comprehensive and informative than any outcome obtained by the distinct analysis of each single data source. In other words, the complementarity or synergy among different data acquisition modalities contributes to a unified and enhanced view of the system under study or improved modelling or reformative decision-making and understanding of the system or phenomenon. Therefore, a complex challenge lies in the fusion and mapping of various distributed and heterogeneous data in arbitrarily feature representation spaces. It is vital to propose a valid integration approach to cover the versatility of a phenomenon. This section investigates studies on information fusion, which involves two aspects: strategies of information fusion and KG-aided information fusion.

2.5.1 Strategies of Information Fusion

Information fusion is defined as a framework, fit by an ensemble of tools, for the joint analysis of data from multiple sources (modalities) that allows achieving information/knowledge not recoverable by the individual ones, although the degree of generalisation, formalisation, and unification of the methodology is distinctive (Cocchi, 2019b). Unlike the data collected from a single source, the data fused from multiple sources usually assembles comprehensive and informative information about different phenomena (Kong et al., 2020). Depending on the different objectives of fusion, three levels of information fusion were defined and have formed the most prevailing taxonomy, which is low-level, mid-level, and high-level, respectively (Hall and Llinas, 1997).

Information fusion at a low-level focus on integration algorithms that operate directly on raw data blocks, which is the simplest form and is known as the observational level. In other words, low-level fusion refers to the concatenation of two or more data matrices in such a way that the observations are in the shared mode. This type of fusion may be undertaken either by using a variety of methods that directly operate on several data blocks, joined or coupled, by decomposition, i.e., multiblock, multiset, tensor, and coupled matrix-tensor factorisation methods, or simply by concatenating the different data blocks. The characteristic of low-level fusion is to model the data (either for exploratory or predictive purposes) after fusion coupling. Therefore, a major advantage is the possibility of interpreting the results directly in terms of the original variables collected in each data block (Cocchi, 2019a). Meanwhile, the main disadvantage of low-level fusion is that typically data sets are obtained in which the number of observations is much smaller than the number of variables (Cocchi, 2019a). As a consequence, many multivariate data analysis techniques are not directly applicable to such data, and some form of regularisation needs to be applied due to the curse of dimensionality.

For mid-level, information fusion methods are applied to feature extraction from each data set, which is called feature level or state-vector level. The data is usually collected and integrated from many scenarios. To be specific, the feature sets collected from multiple sources are integrated to represent the input by generating a new high-dimension feature vector. In other words, the information fusion methods at the mid-level are typically designed to accomplish the most informative feature vector through feature selection algorithms (Gravina et al., 2017). This type of fusion implies a first modelling step before fusion aimed at extracting features from each data block. Mainly two approaches are used, either decomposition (or resolution) techniques or variables selection methods. The obtained features are then fused in a ‘new’ data set, which is modelled to produce the desired outcome. Therefore, the major advantage of mid-level fusion lies in reducing the dimensionality of each data matrix separately before attempting to link them by means of data fusion (Cocchi, 2019a). However, in terms of outcome interpretation of models, mid-level fusion builds a link between the prominent features in the final model and the original variables of the corresponding patterns, thus leading to the most conspicuous limitation (Cocchi, 2019a).

High-level information fusion can be regarded as decision level or information level, where decisions of model outcomes from the processing of each data block are fused. In other words, information fusion at a high level focuses on the final result, such as the correct prediction of the class or attribute of each sample. Generally, the role of each data block and its original variables is not investigated because a fused model in the strict sense is not obtained but only a ‘fused decision’. Therefore, in this type of integration, the adequate scaling of each platform is not an issue because only the results are combined. Compared with the low-level and mid-level, the main merit of high-level fusion is to improve the final performance by statistic-mathematical integration of information from the different analytical sources, thus giving rise to reducing the overall uncertainty among the single experimental uncertainties (Cocchi, 2019a). Therefore, greater confidence in the combined performance is achieved by this type of fusion. Meanwhile, another obvious advantage of high-level fusion is the possibility of handling missing information because high-level fusion does not necessarily require the measurement of signals for all the samples (Cocchi, 2019a). However, in the context of the fusion process at a high level, there is an obvious drawback that does not provide further insight into relevant analytical features (variables), as well as the correlation and common variation between the considered analytical sources. In addition, the selection, optimisation, and calibration of classification or regression models for each analytical block can be troublesome and time demanding.

In addition, conventional multi-faceted modelling approaches concatenate feature vectors to fuse different facets, not considering varying distance metrics across facet boundaries. In this case, one fundamental challenge in fusing disparate facets lies in bridging the semantic gaps between them. Meanwhile, as a large and complex representation, graph-structured data comprises rich relational information among elements, which has received considerable attention in recent years. The graph-structured data is the natural target to cover the domain information in many applications, such as social networks and molecular structures (Bronstein et al., 2017). A knowledge graph, which is represented as a typical type of graph-structured data, is composed of entities (nodes) and edges. Edge is regarded as a fact that shows a specific relational connection between two entities. Specifically, an edge is demonstrated in the form of a triplet, including a head entity, relation, and tail entity. Therefore, as an abstraction to encode knowledge in a specific domain, constructing knowledge graphs based on graph-structured data provides a promising way to integrate rich relational information through connecting entities represented by features.

2.5.2 KG-aided Information Fusion

As reviewed above, KG is a promising and emerging information technology with the potential to integrate heterogeneous multi-source data. In the meantime, there is still a research challenge regarding using KG to fuse high-dimensional and heterogeneous data. In this context, a KG-supported information integration approach was proposed with the representation of temporal data and transforming such data into temporal graphs for further modelling. Through such transformation, both the temporal dependency of the manufacturing events and the attribute interactions among the multi-faceted concepts are considered.

Although there has been a notable increase in the number of efforts to construct large-scale KGs, the process of harvesting meaningful information from heterogeneous data sources is not easy. Integrating data from different sources provides users with a unified view of data by combining data from different sources. In most enterprises, relational databases house a significant amount of data (Abu-Salih, 2021). In order to integrate data across multiple databases, one approach relies on a global schema which indicates how the items within these databases are interrelated (Ji et al., 2021). However, the result of a large number of tables and attributes, establishing a global scheme can be a very challenging task as knowledgeable experts who created the databases are usually unavailable and owing to a lack of documentation, it can be challenging to interpret the data as well.

In light of the difficulty of creating a global schema, it is convenient to convert the relational data into a database that follows a generic triple schema, i.e., a knowledge graph. An attribute mapping is created based on specific business needs, for example, in response to a specific business question, and this mapping can be represented in a knowledge graph. A recent study (Wang and Zhang, 2018) proposed a scheme for calculating non-linear distributions of IoT data using deep learning. As a result of the fusion of multi-source heterogeneous data sets, the accuracy of the recognition of data sets has been significantly improved. Although redundant data and dynamic data flow can be fused, high accuracy cannot be achieved through the fusion of redundancy and dynamicity. It has been shown in another study (Jabbar et al., 2018) that different data models for unstructured and heterogeneous raw data formed by the internet of things were analysed in real-time, but the analysis and study of text data did not take place. In a similar work (Li et al., 2020a), the topological and semantic similarities between multiple sources of knowledge using two knowledge graphs are analysed. With the aim of implementing semi-automatic linkage among nodes and merging the relations between two graphs, four concept-knowledge operators were provided. In order to reduce the dimensionality of the associated data, a ternary data fusion algorithm based on reinforcement learning was proposed (Ng et al., 2021).

Data fusion tasks using GNN and their variants are showing promise in a variety of applications due to the advancement of GNN. The convolutional neural network, for example, is limited to processing only grid structures rather than general domains, while the recurrent neural networks fail to take into account spatial relations between sensors or suffer from long-term dependency learning. It has been proposed by (Wang et al., 2021a) which combine the strengths of graph convolutional networks for spatial learning with the strengths of temporal convolutional networks for sequential learning to address these problems. As a result of these methods, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), sensory data are analysed only in terms of their temporal information, while the intrinsic spatial relations between the sensors are ignored. An adaptive spatiotemporal graph convolutional neural network (ASTGCNN) is proposed by (Zhang et al., 2020a) in order to address this issue. GNNs can be used to perform data fusion, but they have some limitations in practice, such as being complicated to use in practice and requiring a large amount of computation time.

2.6 Summary

In summary, cold rolling in the steel-making process was reviewed in Section 2.2, concerning an overview of steel-making, research of the cold rolling process, and a survey of the strip-breakage phenomenon. During the past decades, data mining has grown rapidly in popularity, allowing many industries to use it to discover hidden patterns in their systems, enabling them to create new models to improve production quality, productivity, optimum cost, and maintenance. Meanwhile, as a result of the development of IoT, the situation of a data-rich environment has led to the deployment of a variety of tools and techniques for data mining. Therefore, the study of steel making based on data mining deserves more attention. In Section 2.3, data mining and its applications were investigated. With the capability of processing various types of data, it can be seen that data mining has become prevailing and essential in industrial analytics. Moreover, fusing information from multiple sources is of great significance to obtaining more comprehensive and informative knowledge than that of each single data source. As a promising way, KG encodes knowledge for information integration in a specific domain that comprises rich relational information among elements. In this context, the relevant studies of KG were introduced, which are classified into KG construction and graph representation learning. It is evident that KG-aided is an important topic in the era of IIoT. However, from the literature, it can be seen that KG has not gained sufficient attention in the steel industry, and very few existing studies have focused on KG-aid data mining in the steel-making process. Therefore, Section 2.5 examined the studies on information fusion concerning the strategies of information fusion and KG-aided information fusion techniques, and Section 2.6 summarised this chapter.

By reviewing these related works, the importance of Industry 4.0 technologies for improving the efficiency and effectiveness of manufacturing processes, particularly in the complex steel-making industry, was highlighted. Through the analysis of various data mining techniques and applications in steel-making, we have identified the challenges and opportunities associated with the use of data-driven approaches for fault diagnosis and prediction. Additionally, the review of knowledge graphs and graph representation learning techniques has provided insights into how we can represent and utilize the vast and scattered domain knowledge in steel-making for fault modelling. Finally, the exploration of information fusion strategies, particularly the

KG-aided information fusion approach, has shed light on how we can integrate multi-sourced data and knowledge to improve the accuracy and effectiveness of fault diagnosis and prediction in steel-making. Overall, this literature review has provided a solid foundation for our proposed approach to fault diagnosis and prediction in the steel-making process using data mining techniques under Industry 4.0 and has identified several key areas for future research and development in this field.

Chapter 3 A Framework for Data Mining in Steel-Making under Industry 4.0

3.1 Introduction

Industry 4.0 refers to the fourth revolution of the industry, which mainly focuses on intelligence and automation (Qin et al., 2016). Emerging technologies such as AI, cloud computing and IoT have boosted the development of Industry 4.0. The Internet of Things, cloud computing, smart sensors, and other technologies are being integrated into today's factories as part of I4.0 (Cemernek et al., 2017). In this case, data-driven approaches to investigating production failure in steel-making have previously been applied in a handful of works, as reviewed in Section 2. However, there are no existing study reports on how to conduct fault diagnosis under the context of Industry 4.0. It is imperative that the industry explore a framework that can assist with fault diagnosis in steel-making in the next generation. In this chapter, a framework is designed for fault diagnosis based on the multi-sourced data and existing domain knowledge in steel-making. The data and techniques relevant to the proposed framework are introduced in Section 3.2.

3.2 A Framework for Data Mining in Steel-Making under Industry 4.0

According to the literature review in section 2, besides the numerical production data in the steel-making process, it is widely acknowledged that the semantic data and existing domain knowledge are of great significance to fault diagnosis. Existing studies mainly focus on modelling based on numerical steel-making production data, while semantic data and existing domain knowledge are not fully exploited. With the development of IIoT and knowledge management techniques, it is possible to conduct data and knowledge concerning production failure in steel-making. Hence, it is of great significance to introduce multi-source data and domain knowledge for data-driven fault diagnosis.

Meanwhile, KG exhibits remarkable expressive ability and excellent modelling flexibility, which has been described as a graph model for representing information in a manner that can be understood by a broad audience (Nguyen et al., 2020). It has been widely used and shown merits. In our case, the aim was to capture and represent the complex relationships and dependencies among various factors contributing to strip breakage in the steel manufacturing process. With the ability to represent entities and relationships in a rich and flexible manner, knowledge graphs provide a more suitable solution for our needs than ontologies or taxonomies. While other techniques, such as ontologies and taxonomies, are useful for organizing and classifying information, they are limited in their ability to capture the complexity and variability of real-world relationships. Knowledge graphs, on the other hand, allow for more nuanced and context-dependent representations of knowledge, which is particularly relevant for our research, where the relationships between various factors influencing strip breakage can be highly dynamic and context-specific. Furthermore, knowledge graphs can be easily extended and modified as new data and knowledge become available, whereas ontologies and taxonomies often require more effort to update and maintain.

In this context, exploring how to combine different techniques also needs to be considered in the framework. The proposed framework is illustrated in Figure 3.1. The proposed framework includes the following stages: multi-sourced data and knowledge acquisition, data and knowledge processing, KG construction and graphical data transformation, KG-aided modelling for fault diagnosis and decision support for steel manufacturing. The framework was designed to refer to the concept of data mining and graph representation learning, which was detailed in Appendix A1. Firstly, a multitude of fault-related data from different sources is collected, as well as fault-centric domain knowledge such as empirical knowledge and existing literature. Secondly, data and knowledge processing are conducted using different techniques based on their modality. Thirdly, a domain KG is constructed, and it serves as the backbone for the population of multi-sourced data. Then, by graph feature extraction using GRL techniques, the embeddings of KGs are fed into the ML pipeline. Lastly, the modelling results from ML are used for decision support regarding fault diagnosis, contingency measures design and quality improvement in steel-making.

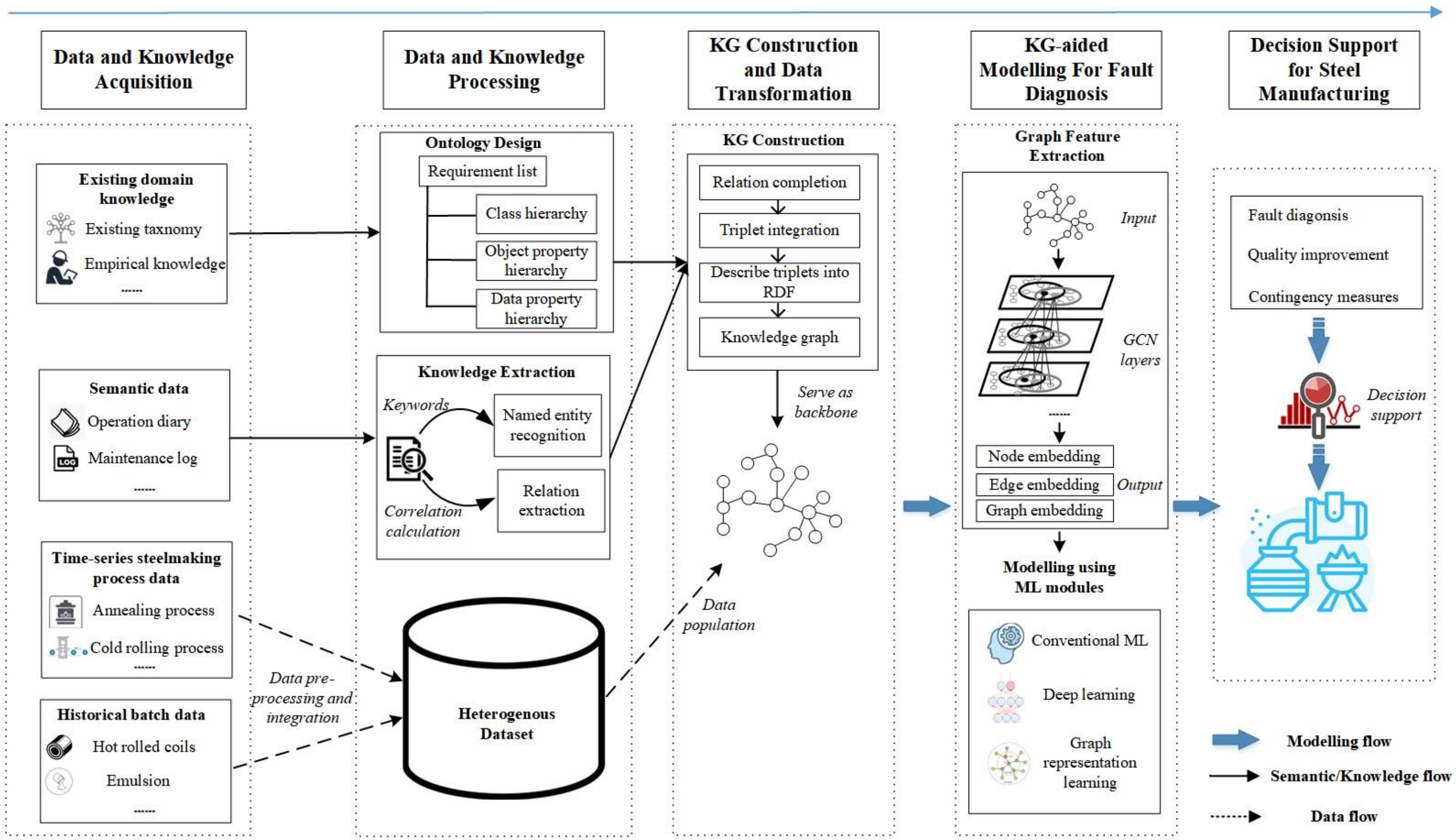

Figure 0.1 A Framework for Data Mining in Steel-Making under Industry 4.0

3.2.1 Knowledge and Multi-sourced Data Acquisition

As reviewed in Section 2.3 and suggested by the domain experts, the preparatory processes before the occurrence of production failure should be deemed as potentially relevant. For example, the failure in the cold rolling process and its relevant data should be considered in the cold rolling and its former processes. It should be noted that regarding the collection of cold rolling data, which is typically recorded in a multivariate time-series manner, the concept of recency should be incorporated. Since most of the failures are instantaneous, the temporal observations that extend far from the breakage point into the past are believed to be less informative than breakage-recent observations [27]. In this context, data are collected in a run-to-failure manner, from the occurrence time point backwards in time to obtain the most recent observations. Figure 3.2 is an illustration of data collection along the cold rolling process.

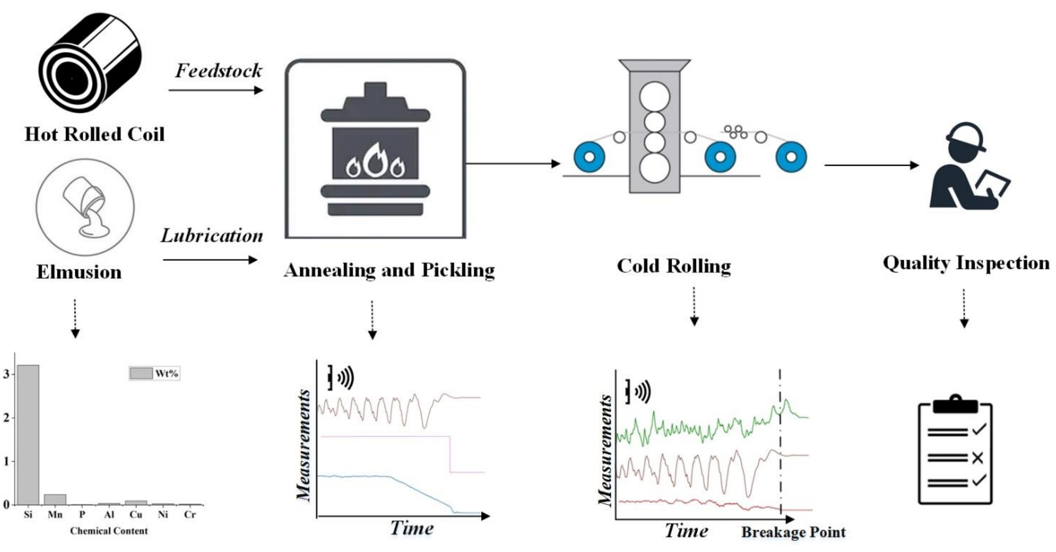

Figure 0.2 Illustration of data collection along the cold rolling process

Besides the cold rolling process data, the annealing and pickling data are also collected. This process affects the occurrence of failure from the feedstock aspect, such as strip roughness. In most cases, it is not easy to construct the backbone of KG under a domain environment without the collaboration of domain experts (Jin, 2018). In the domain-centric knowledge representation, each triplet is described using the RDF language utilising open-source platforms. Therefore, these platforms are responsible for building and storing domain-centric knowledge bases. The fault-centric KG was constructed based on the refinement of the hierarchy structure and the completion of the relationships. A summary of related studies on strip breakage and the cause analysis can be categorised into four different facets, namely material-related issues, equipment malfunctions, rolling operations, and the rolls pushing the strips.

3.2.2 Data and Knowledge Processing

The multi-source data are different in type and granularity. In order to use multi-source data for modelling, the first step is data mapping. For example, the cold rolling process data is collected at a high frequency and a sampling rate that can be up to the scale of a millisecond, while the batch data, such as hot-rolled coil data, are collected per batch. Hence, the data need to be mapped into the same granularity before it is used for modelling.

Knowledge repositories are typically fragmentary, with scattered knowledge distribution in industries, resulting in extensive participation of knowledge alignment from domain experts. For the steel-making industry, the processes are strongly correlated and complex. Processing such knowledge using conventional approaches is usually time-consuming and labour-intensive. Moreover, it is important that information extraction and entity resolution are both performed accurately for the development of knowledge management, but they are often insufficient (Getoor and Machanavajjhala, 2012, McCallum, 2005). In this context, information is often missing and noisy, such as missing relations. In other words, such a relationship extracted from KG needs to be refined or completed in order to be used effectively in an application.

3.2.3 Knowledge Graph Construction and Graphical Data Transformation

For KG construction, the first key step is to determine the applicable ontologies. As semantic data models, ontologies are mainly utilised to describe the relationships between concepts in a given domain and provide standardised, clear and unambiguous definitions that can be shared. Specifically, although the general types of things that share certain properties are modelled in domain-centric ontologies, these models do not contain information about specific individuals from domains. As each industrial scenario has distinct characteristics, the construction method of general KG cannot be applied to the establishment of industrial domain-centric KG. It is necessary to determine a suitable ontology in advance when establishing a KG in an industrial environment.

Typically, a two-step section comprising ontology design and knowledge extraction is conducted. After clarifying the request for the domain-centric KG construction, the class hierarchy structure is designed to conform to the requirements thoroughly. Based on that, the property hierarchy is confirmed in accordance with the class hierarchy, including the object property hierarchy and the data property hierarchy. Meanwhile, the keywords and correlation calculation methods are developed to recognise the entities and relations in extracting knowledge, respectively. In this context, these three types of hierarchy are described in OWL language for the documentation of the designed ontology. In the second stage, the GRL model is built to predict potential relations for KG completion. Specifically, the edges are divided into two types: existing edges and non-existing edges. The relations represented by the two entities' embeddings are classified in the pre-trained GNN model. Finally, the third stage contains KG construction and KG visualisation. The overall triplets are described by the RDF format based on the triplet integration. Subsequently, the open-source platform is used to fuse the triplets for KG construction, and the visualisation tool is deployed to present the generated KG in a graphic format.

With the aim of mining implicit knowledge from a graph, triples in the knowledge graph are transformed into their corresponding low-dimensional vector embedding using either a knowledge representation learning model or a graph convolution model. By populating multi-sourced data into knowledge graphs, an embedding approach is necessary in order to transform the data from these graphs into information that can be used for multi-source conceptual modelling. In this study, as a convenient way to accomplish the embedding process, GRL techniques can be used to extract the connected features in an end-to-end manner.

3.2.4 KG-aided Modelling for Fault Diagnosis

KG enables the merging and organisation of time series data knowledge in order to predict faults in subsequent devices. Multiple sensors generate and collect data during the manufacturing process. With the transformation of multivariate time-series data into the stack of temporal graphs, the output feature can be fed into the ML model for further modelling. Meanwhile, ML algorithms such as temporal convolution networks, a method of processing time-series data, utilise dilated causal convolution and residual connections in order to address the problems discussed above. Dilated causal convolutions are used only for elements that precede the current element, while CNN performs convolution on elements adjacent to the current element.

3.2.5 Decision Support for Steel Manufacturing

Prediction of this failure can bring significant benefits to the cold rolling industry in terms of contingency mitigation and quality improvement. For severe production failure which occurs instantaneously in the steel-making process, a micro-level prediction can minimise the occurrence and impact of such 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.

3.3 Summary

With the leverage of connectivity and intelligence in the era of Industry 4.0, PdM has become an essential key. In a data-rich environment, the surrounding data and knowledge relevant to the production failure in steel-making can be collected and used for fault diagnosis. A framework has been designed for fault diagnosis based on the multi-sourced data and existing domain knowledge in steel-making.

Chapter 4 Multi-faceted Modelling for Fault Diagnosis in Steel-making

4.1 Introduction

Due to its high efficiency and accuracy, the cold rolling process is a primary metal-forming process for the manufacturing of steel products (Mashayekhi et al., 2011). 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 (Yan and Li, 2006a). This failure has serious consequences, such as yield loss due to unplanned stops of the rolling mill, extended downtime caused by severe damage to work rolls and altered rolling performance for subsequent rolling when production resumes following a strip breakage (Cui and Zhao, 2013a, Bhattacharya et al., 2016, Chen et al., 2019).

Numerous studies on this product failure 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 (Cui and Zhao, 2013a, Xu, 2015, Liu, 2015a, Liu et al., 2014), the causes of strip breakage vary 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 (Kuo and Kusiak, 2019). 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 (Da Cunha et al., 2006), which is the primary step for quality improvement rather than quality prediction (Lopes et al., 2019). 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 in this chapter. Recurrent Neural Networks (RNNs) (Lipton et al., 2015) 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 we should select features. In this chapter, 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 strips, 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 annotates whether a strip will break within the next time window (denoted as the 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 (Gamboa, 2017). Among various deep learning architectures, recurrent neural networks (RNNs) retain the recent memories of input patterns, which makes them suitable for time-series processing (Punia et al., 2020). 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 (Wang et al., 2019). 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 (Katona et al., 2019) 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 chapter is structured as follows. Section 4.2 outlines the flowchart of the proposed methodology of multi-faceted modelling on strip breakage. Section 4.3 reports an experimental study using real-world cold rolling data to demonstrate the effectiveness of the method, followed by results analysis and discussion in Section 4.4. Section 4.5 summarises this chapter.

4.2 Methodology

In this chapter, we propose a machine learning-based approach for fault diagnosis in steel-making that consists of three main stages, as shown in Figure 4.1.

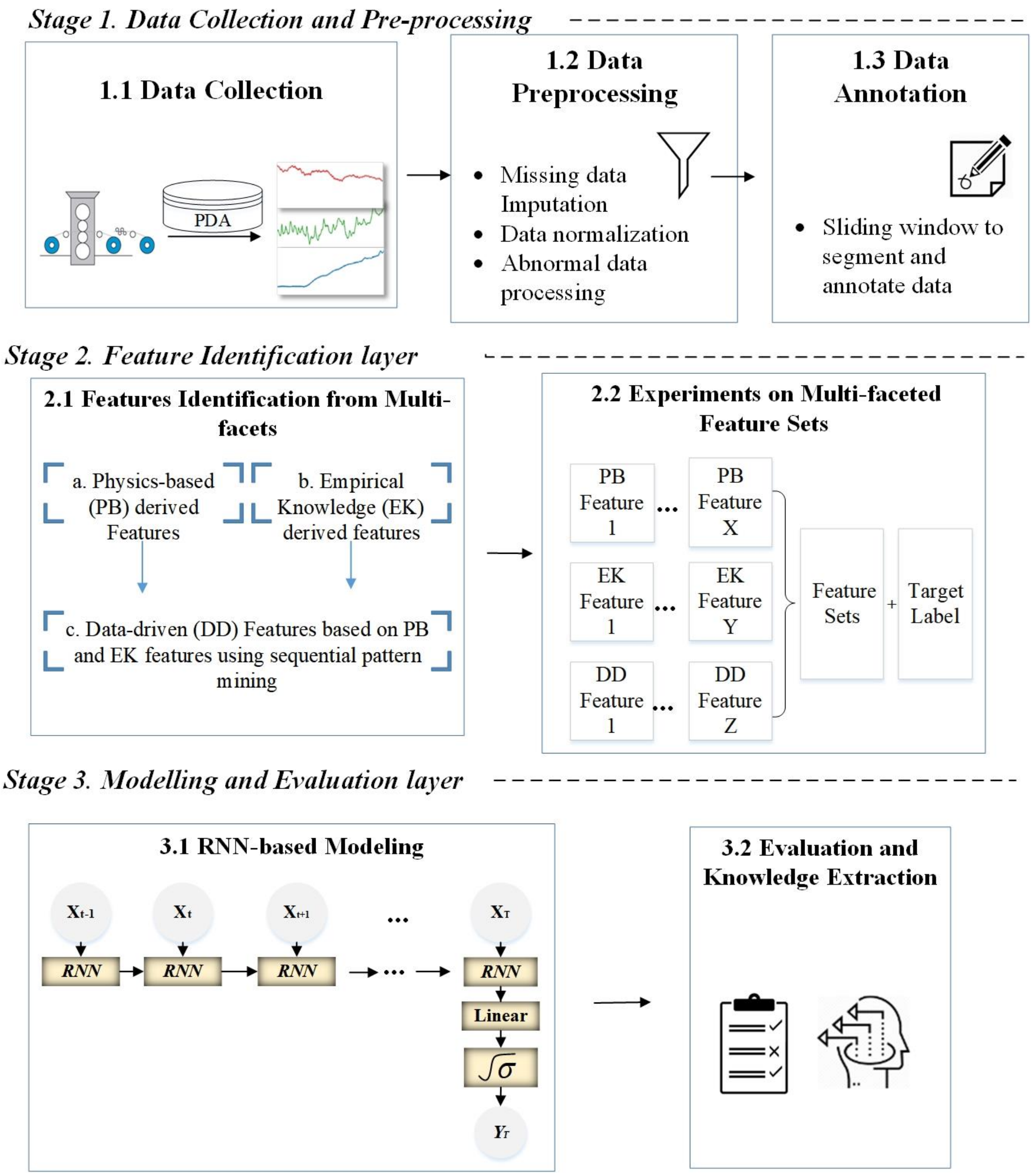

Figure 0.1 Illustration of the proposed multi-faceted modelling approach

The methodology illustration figure depicts the multi-stage approach followed in this study. The first stage involves the collection and pre-processing of data from a cold-rolled steel strip plant. In stage 2, three different facets are considered to generate sets of candidate features. The three feature sets are then combined to present different scenarios for modelling strip breakage. This approach helps to account for the complexity and variability of factors that can contribute to strip breakages, such as the mechanical properties of the strip, the processing conditions, and the operating parameters of the plant. In stage 3, in the case of time-series data, RNNs are often preferred over other techniques, such as traditional regression models, due to their ability to capture the temporal dependencies in the data. RNNs are able to take into account not only the current input but also the historical context, which is critical for accurately modelling time-series data. In addition to their ability to handle time-series data, RNNs also offer advantages over other techniques in terms of their flexibility and ability to model complex relationships. For example, RNNs can be used for tasks such as sequence prediction. Therefore, a sequence-to-vector RNN architecture is applied for modelling.

4.2.1 Data Pre-processing and Time Window Processing

The collection of cold rolling process monitoring data was first conducted. In another related study (Chen et al., 2020c), we proposed a data fusion approach, which focuses on the fusion of multiple data sources to predict strip breakage. In this chapter, we focus on the multivariate time-series cold rolling process data to model this instantaneous production failure. Since strip breakage is an incident that occurs suddenly, 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 (Batal et al., 2012). Following this, data was collected from the time point where strip breakage occurs backwards in time so that the most recent observations are obtained. Under this run-to-failure manner, we applied a sliding window strategy to segment the raw time-series data into a collection of pieces due to the high correlation of neighbouring data, an illustration of this strategy is shown in Figure 4.2. 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.

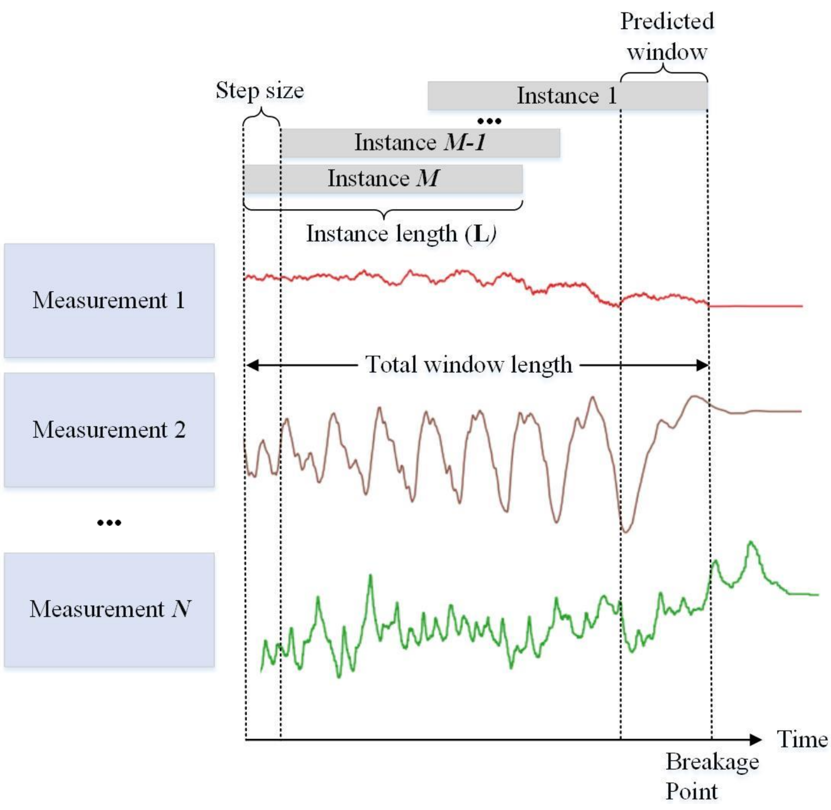

Figure 0.2 Illustration of the proposed sliding window strategy.

To be specific, the multivariate time-series cold rolling process data was segmented into two-dimensional matrixes following a certain step size. The total window length is segmented into ***M*** instances by sliding the window backwards in time from the breakage point. Each instance is a matrix that contains ***L*** sampling point and ***N*** attributes. The interval between the last sampling point and the strip breakage point determined the label of each instance. The matrix instance was labelled as 0 if the interval was wider than the predefined predicted window length, which means the strip would not break in the next predicted window. Otherwise, the matrix instance is labelled as one if the interval is within the predefined predicted window length, which represents a future breakage.

4.2.2 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. In terms of feature identification, certain features work better either for a specific domain or in a non-domain-dependent dataset (Mejova and Srinivasan, 2011). 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.
- 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.

- The **data-derived** (DD) features are binary features derived from sequential patterns based on PB and EK features.

4.2.3 Multi-Faceted Modelling for Strip Breakage using Machine Learning

Under the sliding window strategy in Section 3.1, the strip breakage task is transformed into a binary classification task classifying whether the strip will break within the next predicted time window. In this context, we proposed a supervised machine learning approach using a sequence-to-vector RNN architecture to conduct this classification task.

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 4.3. 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 (Hochreiter and Schmidhuber, 1997). 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 different output layer, depending on whether the required task is related to regression or classification (Graves, 2012).

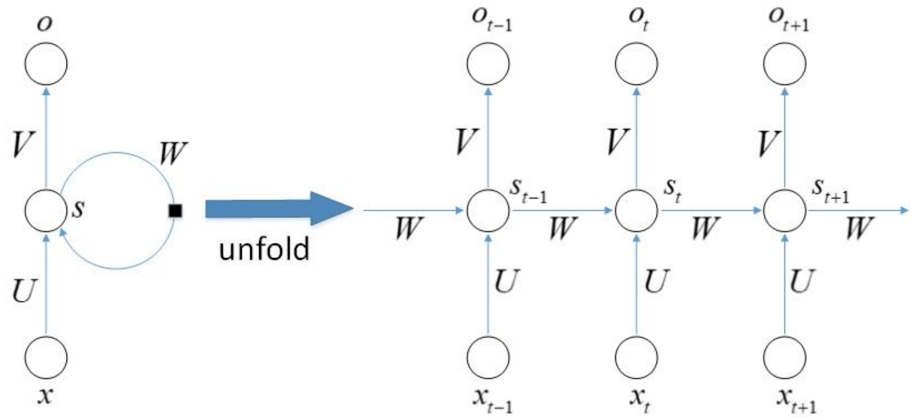

Figure 0.3 Illustration of the RNN structure

Specifically, a two-dimensional matrix instance entering the proposed RNN architecture contains $\boldsymbol{L}$ sampling points and N attributes. This instance represents a time segment as a matrix ($\boldsymbol{L} \times \boldsymbol{N}$) and is fed into the recurrent layers. As a sequence-to-vector architecture, only the output from the last neuron is fed into the linear layer; the other outputs are ignored. Subsequently, an activation function is embedded into a linear layer to make binary predictions.

4.3 Experimental Setup

We conducted this experimental study using the historical data provided by a cold-rolled 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 (Li et al., 2018b). 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. With a higher silicon content, the electrical steel strips are lower in toughness compared with the general low-carbon steel (Yan et al., 2015). In this case, there is a higher chance of strip breakage during the production 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 the attention of 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.

4.3.1 Data Acquisition and Preparation

The data used for this research was obtained from multiple sources relevant to strip breakage, namely hot rolled coils (HRC), annealing and pickling (A&P), emulsion, and cold rolling process (CRP). Specifically, the on-site production data acquisition (PDA) system, continuously monitored and recorded CRP variables such as tension, speed, and roll gap position at a frequency of 100 Hz. The raw data was of high resolution to ensure the most accurate and detailed information possible. The data was pre-processed by identifying the momentary manner of strip breakage, followed by data cleaning to address abnormal and missing values. The dataset contained 33 broken strip coils marked with non-material causes under the same incoming material grade, covering three months of production, and was divided into training and testing sets. The sliding window strategy was implemented to obtain a manageable dataset size while still considering recency.

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. The training set consisted of 27 coils, and the test set contained six coils. The granularity of the data was high, enabling the identification 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 at 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. Specifically, 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. The division of training and testing data was made to prevent overfitting and to evaluate the performance of the developed models on previously unseen data. The selection of the test set was based on its similarity to the training set, as both sets shared the same incoming material grade and contained coils marked with non-material causes. The training and testing sets were carefully selected to ensure that the models were trained and tested on a representative sample of the data, which increased the confidence in the generalizability of the results.

Table 0.1 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

Regarding the absence of validation data, it is worth noting that the models were not validated in this study to prevent overfitting. The absence of validation data might raise concerns about the reliability of the models. However, the use of a sliding window strategy to generate instances of data for each coil, the careful selection of the training and testing sets, and the use of appropriate evaluation metrics, such as precision, recall, and F1-Score, were used to assess the performance of the models. These measures ensured that the models' generalizability and reliability were evaluated and that the absence of validation data did not significantly impact the results.

4.3.2 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.

Table 0.2 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 results in periodic variation in the roll gap
11	Measured slip (%)	The displacement ratio between the strip coil and the working roll

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 4.2.

Second, in addition to the features identified from EK, which cover specific strip breakage causes, features identified from PB models were considered as the second facet to be included in general causes of strip breakage. The diameter disparity between the top and bottom working rolls as well as the left and right deflector rolls, was first considered (Cui and Zhao, 2013a). 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 (Panjković et al., 2012), 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 4.3.

Table 0.3 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	The vertical vibration frequency of top work rolls (Hz)	The spectrum of the top work roll position signal
4	The 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 (Hu et al., 2006). 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 Tables 4.2 and 4.3. First, to handle the complex time series, temporal abstraction (Shahar, 1997) was conducted to transform the numeric time-series variables into a high-level qualitative 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 10%, 25%, 75% and 90% 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 (Pei et al., 2001) sequential pattern mining algorithms were applied to 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 sequential pattern mining was then conducted for all 27 coils. Sequential pattern mining was conducted using the open-source package SPMF (Fournier-Viger et al., 2016). SPMF is an open-source software and data mining library which is specialized in pattern mining. For the PrefixSpan algorithm, the Min support was set to 0.4 and the 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 is shown in Table 4.4.

Table 0.4 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. Adam was adopted as an optimiser (Kingma and Ba, 2014). 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 (Paszke et al., 2017). Details of the parameters for the training process are specified in Table 4.5.

Table 0.5 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) (Liaw and Wiener, 2002, Jun et al., 2019), support vector classification (SVC) (Suykens and Vandewalle, 1999), 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 the time domain and frequency domain, were designed and crafted, as shown in Table 6. The Scikit-learn (Pedregosa et al., 2011) with the default setting were applied for the benchmark tests of the conventional algorithms. 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, the 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 the 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 4.5. 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 0.6 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.

4.3.3 Evaluation Metrics

Typically, we can use the following two metrics from the confusion matrix to evaluate the classification performance of positive and negative classes independently (Storey, 2003):

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

$$TPR = \frac{TP}{TP+FN} \tag{4.1}$$

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

$$TNR = \frac{TN}{FP+TN} \tag{4.2}$$

Since classification intends to achieve good results for both positive and negative classes, neither of these measures is adequate for evaluating the classification performance. The Area Under the receiver operating characteristic Curve (AUC) provides a single measure of a classifier's performance to evaluate which model is better on average (Fawcett, 2006). In this context, AUC is selected to evaluate the performance of the proposed modelling methodology.

4.4 Experimental Results

4.4.1 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 4.7 under the metrics of AUC evaluated using the test dataset.

Table 0.7 AUCs of models with the best performance 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 4.7, 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. Additionally, it is worth noting that rerunning the deep learning modelling on strip breakage experiments may not necessarily give the same results. This is because there are several factors that can influence the performance of machine learning models, such as the specific initialization of the model parameters, the random shuffling of the training data, and the choice of hyperparameters. These factors can result in slight variations in the performance metrics of the model, such as accuracy or F1-score. However, the meaningfulness and significance of the experiment should not be affected by these variations, as the main contribution of the study lies in demonstrating the effectiveness of the proposed approach in predicting strip breakage using multi-source data. The study also provides insights into the importance of feature engineering and model selection in improving the accuracy of the predictions.

4.4.2 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 4.4.1, to gain insights into 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.

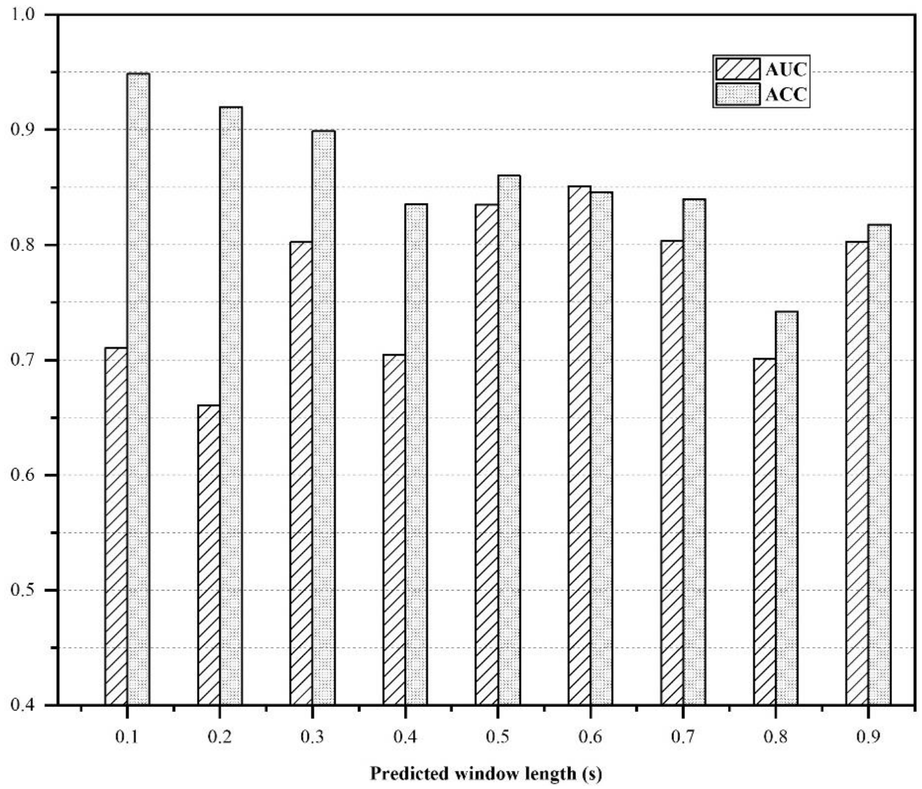

Figure 0.4 Performance of models based on different predicted time windows in terms of ACC and AUC.

Based on different window lengths, the performance was compared when the algorithm converged. The relationship between the algorithm performance and different predicted window lengths in terms of AUC and accuracy (ACC) is shown in Figure 4.4. 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.

4.5 Discussion

In this Chapter, the results of the experiments using multi-faceted feature sets and different time window lengths were presented. The ability to predict strip breakage on a micro-level with an adequately predicted window can provide significant benefits in terms of decision-making for the occurrence of strip breakage.

It is noted that the LSTM network shows merit in multi-faceted strip breakage modelling and outperforms conventional algorithms when only single feature sets are being applied. This is likely due to the limitations of hand-crafted features used to represent raw time-series data, which may limit modelling 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 (Bengio et al., 2013). 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.

Additionally, the study also highlights the performance advantage of the inclusion of the knowledge-based (KB) and statistical process monitoring (SPM) feature sets for the primary feature set. This finding indicates the importance of incorporating domain-specific knowledge and process monitoring data in strip breakage prediction models. Moreover, the trade-off between the predicted time window length and prediction performance is also discussed. It is noted that while ACC generally decreases with the increment of predicted window length, the best AUC is achieved with a predicted window of 0.6 seconds. This finding suggests that the selection of a predicted window length will affect the data balance and that a wider predicted window does not necessarily lead to better prediction performance.

In conclusion, the discussion provides a critical reflection on the experimental results and highlights the strengths and limitations of the study. Future research directions are also suggested, such as exploring more sophisticated deep learning architectures and incorporating additional domain-specific features. Overall, this section adds depth and insight to the research findings and demonstrates the author's ability to evaluate the significance and implications of the study. Fault diagnosis in a predictive manner is of importance to the steel-making industry. 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 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.

4.6 Summary

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 thesis 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. Through the modelling of this instantaneous failure at the event level, the rolling mill operator can improve and optimise their contingency mitigation strategies. For instance, a planned stop action can be taken in advance to avoid further damage from an unplanned fast stop according to the predicted information. Secondly, the post-analysis of this production failure can benefit from understanding the likelihood of strip breakage in the near future. For real cold rolling practice, even if we considered all the causes of strip breakage beforehand, the 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, in most cases, from an unknown reason not conveyed in the current dataset. Therefore, this approach is more practical for breakages with a definable manifestation in rolling process variables, such as breakages caused by chatter.

Chapter 5 Fault-centric KG Building by Structure Refinement and Relation Completion in Steel-making

5.1 Introduction

For the modelling of a multi-faceted concept, no single data source can capture the complexity of all the relevant factors. In Chapter 4, modelling strip breakages in the cold rolling process has proved to be a multi-faceted task as there are multitudinous factors contributing to this fault. However, among the heterogeneous data surrounding multi-faceted concepts in steel-making, a significant amount of data consists of rich semantic information, such as technical documents and production logs generated through the steel-making process. Also, as a conventional manufacturing process with a long history, there exists vast domain knowledge regarding the production failures in steel-making. In this context, in terms of multi-faceted modelling for fault diagnosis in steel-making, proper semantic technologies are desired for the processing of semantic data and domain knowledge in steel-making.

Techniques such as ontology (Lu and Xu, 2017) can be used to develop semantic descriptions of manufacturing resources. However, as ontologies are based on rule representations, they have limited flexibility and adaptability when it comes to describing the semantics of large-scale workshop data. In contrast, knowledge graphs are conceptual databases of structured semantic knowledge, which have become increasingly applicable to IoT semantic collaboration (Li et al., 2019). Depending on the application scenarios, KGs have been divided into two types normally: general KGs and domain KGs. In terms of general KGs, it has been emphasised that the knowledge requirements need more broad than precise (Chen et al., 2021a). Unlike general KGs, domain KGs are regarded as vertical KGs, which describe specific particular domains (Zhou et al., 2021). Although the description scopes of domain KGs are very limited, the depth of knowledge has been emphasised in building a given domain KG (Zhao et al., 2019, Kejriwal, 2019). In this context, domain-specific KG provides a promising approach to handling semantic data and domain knowledge. By using the knowledge concerning fault analysis, a proper fault-centric domain ontology ensures the formalisation of knowledge, making knowledge readable for humans and computational methods (Regulski et al., 2014). Therefore, it is of great importance to construct a fault-centric KG regarding the production failures in steel making.

In this context, this chapter proposed a framework for domain-centric KG construction. In the steel-making industry, knowledge repositories are typically fragmentary with scattered knowledge distribution, resulting in extensive participation of knowledge alignment from domain experts for a KG establishment (Hu et al., 2021a, Yu et al., 2020). In addition, information is often missing and noisy in KGs constructed from raw data, such as missing relations (Getoor and Machanavajjhala, 2012, McCallum, 2005). With the aim of addressing the above-mentioned issues, a framework of domain-centric KG construction has been presented to avoid the massive participation of domain experts, as well as to refine KGs and discover the missing relations in this study. This proposed KG construction framework serves as a model that aims to design a reliable ontology and complete relations in constructing the domain-centric KG for information management and knowledge sharing in steel-making.

The remainder of this chapter is structured as follows. In Section 5.2, a detailed technical roadmap has been demonstrated and explained. Section 5.3 reports the case study using data and empirical knowledge from a real-world cold rolling plant. The results are discussed in Section 5.5. Finally, Section 5.6 summarises this chapter.

5.2 Methodology

In this chapter, we propose a framework of domain-centric KG construction through hierarchy structure refinement and relation completion. The choice of techniques and methods used in this framework is based on the need for accurate and comprehensive knowledge representation and extraction in a domain-centric manner, as well as their effectiveness in processing graph-structured data and predicting potential relations for KG completion. Figure 5.1 illustrates the process of knowledge extraction, which is a crucial step in the first stage of the proposed framework. Keywords and correlation calculation methods are employed to extract knowledge from unstructured or semi-structured data sources. The keywords are used to recognize entities, while the correlation calculation methods are used to identify relations among entities. Specifically, the correlation between each pair of variables is calculated using three different rank correlation coefficient methods, including the Pearson correlation coefficient, the Kendall Tau rank correlation coefficient, and the Spearman rank relational coefficient. These methods are selected based on their effectiveness in accurately and efficiently extracting knowledge from data sources. The extracted knowledge is then used to design the ontology for the KG construction, which is a formal, machine-readable representation of domain knowledge. Overall, the knowledge extraction process plays a critical role in the accurate and comprehensive representation of knowledge in a domain-centric KG.

To be specific, for the first stage, a two-step process comprising ontology design and knowledge extraction is conducted. The choice of using an ontology-based approach for the KG construction is that ontologies provide a formal, machine-readable representation of domain knowledge that can be used to improve information retrieval, reasoning and decision-making. To design the ontology, we clarified the request for the domain-centric KG construction and designed a class hierarchy structure that conforms to the requirements thoroughly. Based on that, the property hierarchy was confirmed in accordance with the class hierarchy, including the object property hierarchy and the data property hierarchy. Meanwhile, we developed keywords and correlation calculation methods to recognize entities and relations in extracting knowledge, respectively. These methods are selected due to the fact that they enable the accurate and efficient extraction of knowledge from unstructured or semi-structured data sources. In this context, these three types of hierarchy are described in OWL language for the documentation of the designed ontology.

In the second stage, a two-layer Graph Neural Network (GNN) model is built to predict potential relations for KG completion. GNN is applied given the fact that it can effectively capture the complex dependencies and interactions among entities and relations in a graph structure. Specifically, the edges are divided into two types: existing edges and non-existing edges. The relations represented by the two entities' embeddings are classified in the pre-trained GNN model.

Finally, the third stage contains KG construction and KG visualisation. The overall triplets are described by the Resource Description Framework (RDF) format based on the triplet integration. The choice of using the RDF format is due to the fact that it provides a flexible and standardized way to represent and exchange structured data on the web. Lastly, the open-source platform is used to fuse the triplets for KG construction, and the visualization tool is deployed to present the generated KG in a graphic format.

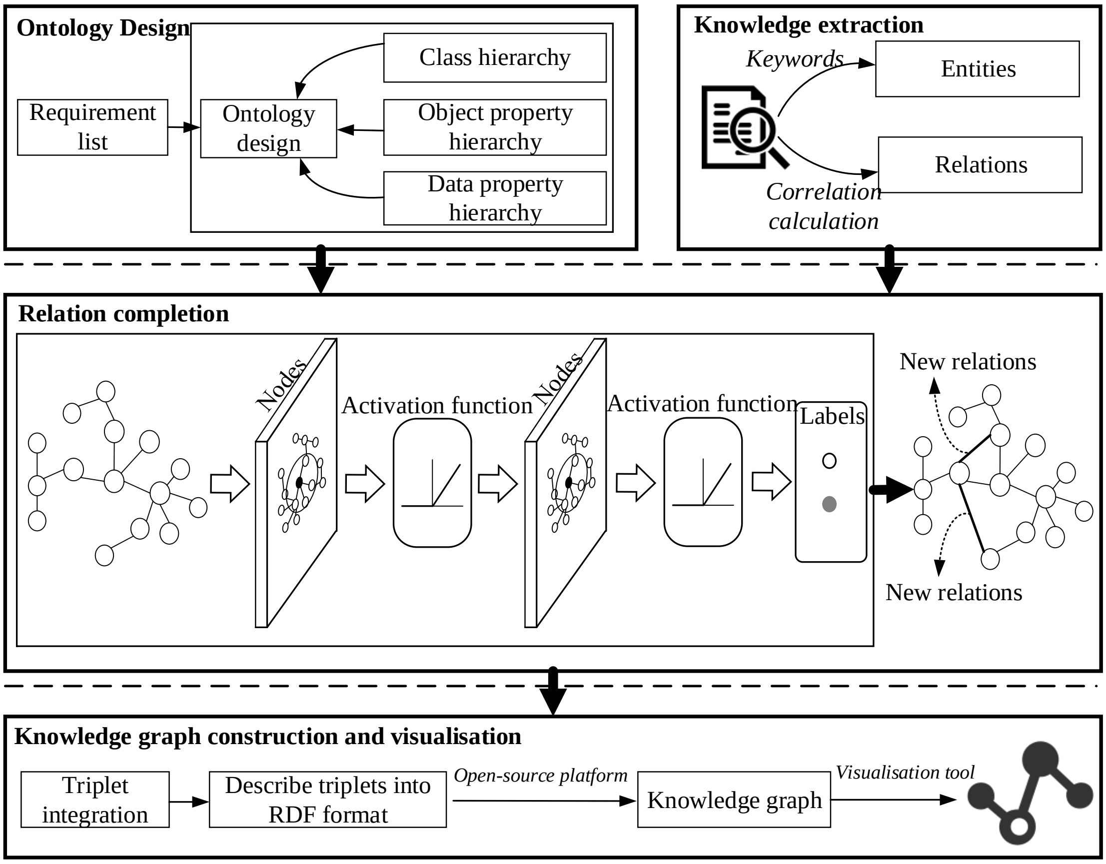

Figure 0.1 The overall framework of the domain-centric KG construction.

5.2.1 Ontology Design

For KG construction, the first key step is to determine the applicable ontologies. As semantic data models, ontologies are mainly utilised to describe the relationships between concepts in a given domain and provide standardised, clear and unambiguous definitions that can be shared. Specifically, although the general types of things that share certain properties are modelled in domain-centric ontologies, these models do not contain information about specific individuals from domains. As each industrial scenario has distinct characteristics, the construction method of general KGs cannot be applied to the establishment of industrial domain-centric KGs.

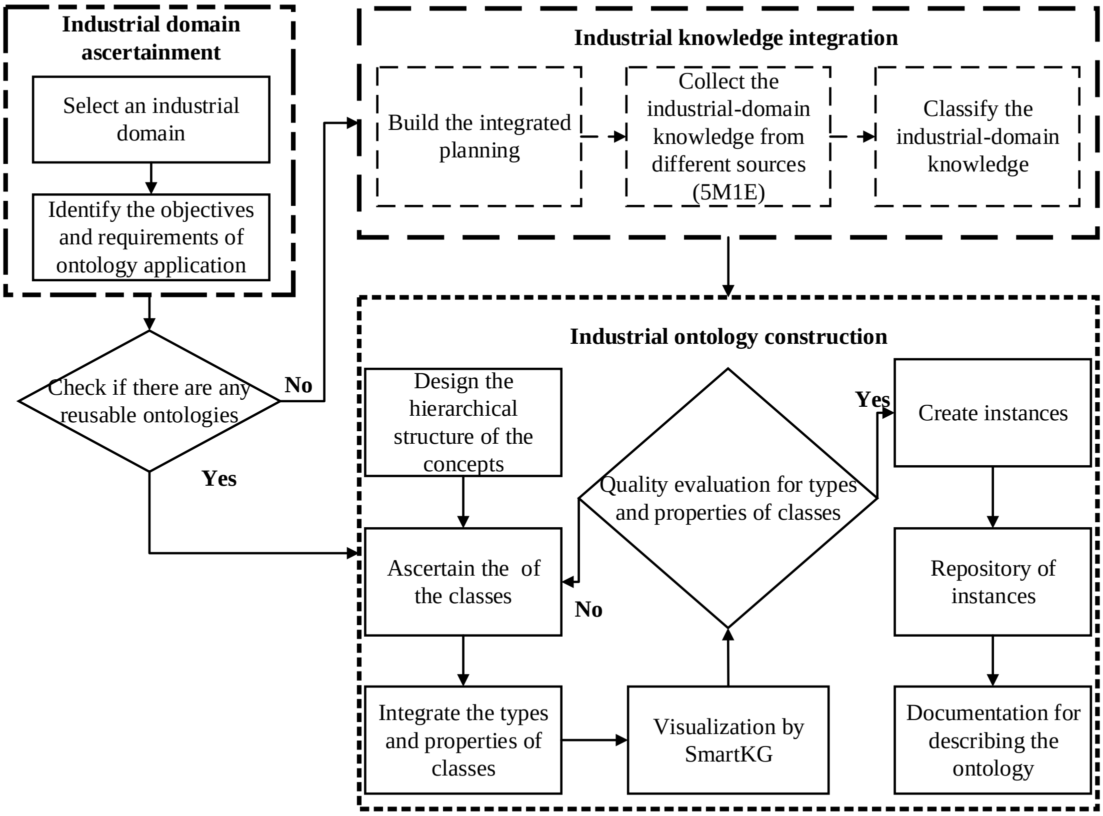

Figure 0.2 The detailed flowchart of industrial ontology design in industrial scenarios.

Figure 5.2 provides a detailed flowchart of the industrial ontology design methodology, which is a critical aspect of knowledge graph construction. The first step of this methodology involves determining the specific industrial scenario, which is essential in identifying the objectives and requirements of ontology applications. This step is crucial in ensuring that the ontology design meets the needs of the target domain. Once the specific industrial scenario has been identified, the next step involves checking if there are any reusable ontologies for the selected domain. If reusable ontologies are available, ontology construction can be based on them to reduce the time and effort required for constructing a new ontology from scratch.

In cases where no reusable ontologies are available, industrial knowledge integration is necessary before constructing ontology. For industrial knowledge integration, three regular procedures are applied to integrate the fragmented domain knowledge from six different sources (5M1E) comprehensively. The collected information is then classified into different and incompatible subsets for the next step. In the last step, the hierarchical structure of concepts is designed in the context of the given domain. The top-level classes are depicted as the root of concepts in the hierarchical structure, which have properties and interrelationships, and constraints.

To ensure the quality of the ontology, an implementation of the SmartKG framework, developed by Microsoft, is made to visualise the developed hierarchy structure quickly. This framework enables quick determination of the hierarchy structure's appropriateness and completeness. If the classes are of high quality, the instances are created and collected into a repository for ontology construction. The ontology is then documented in a file, as it is essential to achieve a knowledge graph. In the case of incomplete or inappropriate classes, the properties and interrelationship constraints of the classes need to be ascertained and fused again, ensuring the quality of the ontology.

The rationality behind these techniques, including the use of the SmartKG framework and integration of domain knowledge, ensures the quality of ontology design and construction, ultimately leading to the development of a robust and effective knowledge graph. This methodology provides a step-by-step approach that ensures ontology design meets the specific industrial scenario's objectives and requirements, leading to the development of a comprehensive and effective knowledge graph.

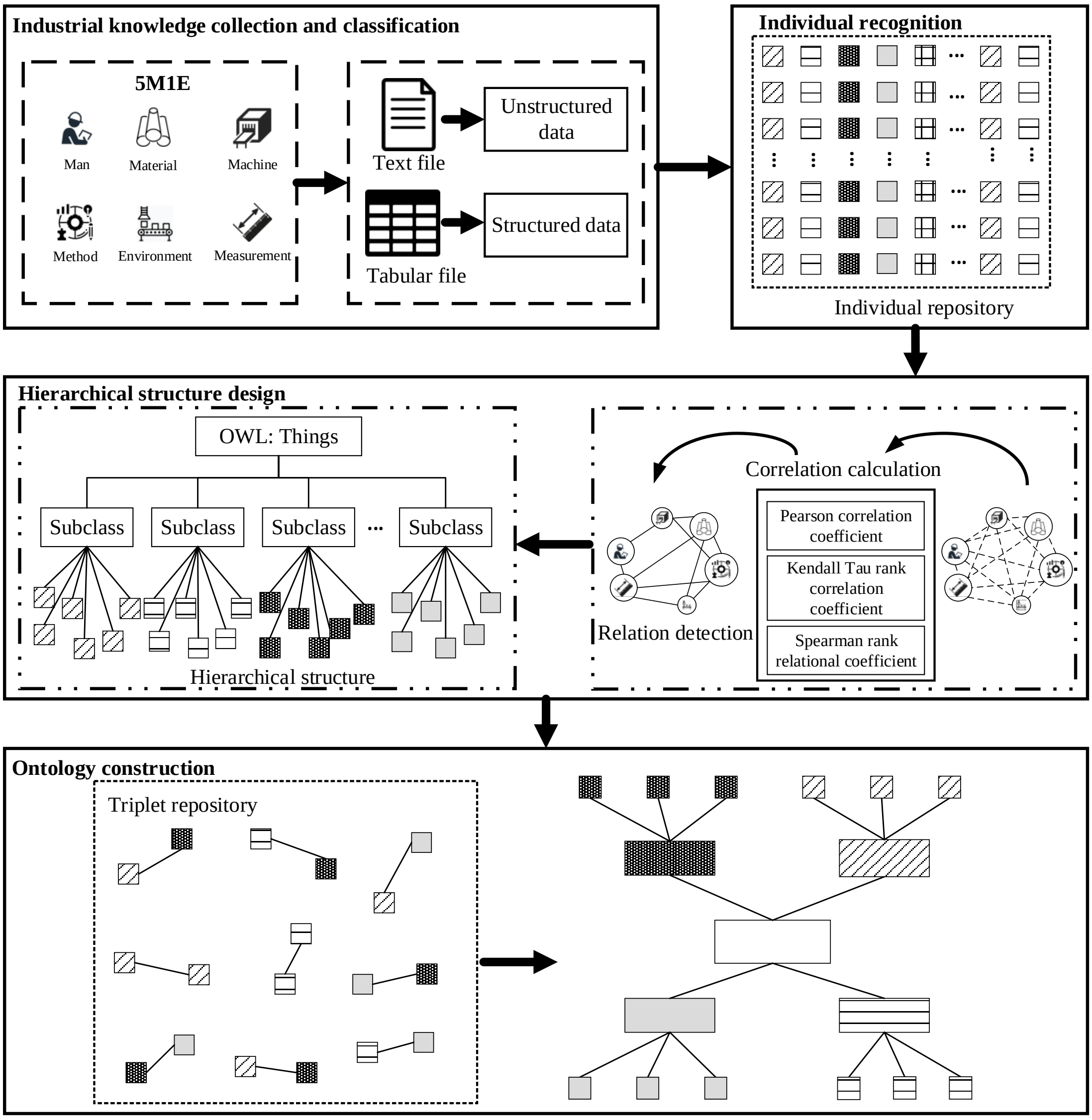

Figure 0.3 The illustrative framework of ontology design and construction.

Figure 5.3 illustrates the framework of industrial ontology design and construction, which consists of four parts. Firstly, the related knowledge of a domain-centric phenomenon has been collected after determining the scope, such as the six resources (5M1E) (YING et al., 2020). The extracted data can be classified into structured data and unstructured data. Secondly, individuals are recognised and extracted to construct the individual repository based on the data resources. Thirdly, the overall hierarchical structure has been designed through a combination utilisation of top-down and bottom-up methods. Specifically, the existing constraints and interrelationships among subclasses have been detected to assist the hierarchical structure accomplishment. For relation detection, the correlation between each pair of variables has been calculated. Inspired by Ref. (Wang et al., 2021a), three different rank correlation coefficient methods, which are the Pearson correlation coefficient, the Kendall Tau rank correlation coefficient, and the Spearman rank relational coefficient, are applied. Lastly, based on the hierarchical structure, the relations of each pair of two entities have been used to connect the individuals. After that, triplets have been integrated to construct specific domain-centric ontologies in industries.

Pearson's correlation coefficient represents the covariance of the two variables divided by the product of their standard deviations, as shown as follows:

$$r_{prs} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (5.1)$$

where: n is the sample size; x_i, y_i denotes the individual sample points indexed with i; and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ represents the sample mean, and analogously for $\bar{y}$.

Kendall rank coefficient is usually regarded as a statistic used to measure the ordinal association between two measure quantities. Specifically, the calculation formula is described as follows:

$$\tau = \frac{n_c - n_d}{n_0} \quad (5.2)$$

where: $n_0 = \frac{n(n-1)}{2}$ and n is the sample size, n_c is regarded as the number of concordant pairs, n_d is the number of discordant pairs.

Spearman's rank correlation coefficient is a nonparametric rank correlation which assesses how well the relationship between two variables. For a sample of size n, X_i, Y_i are converted to ranks $R(X_i), R(Y_i)$. If all ranks are distinct integers, Spearman's rank correlation coefficient has been computed as follows:

$$r_{spm} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{5.3}$$

where: $d_i = R(X_i) - R(Y_i)$ is the difference between the two ranks of each observation.

5.2.2 Relation Completion

Compared with the original GCN model, the GraphSAGE algorithm has been proposed as a comprehensive improvement, which is an inductive representation learning for node embedding (Hamilton et al., 2017a). The main idea of GraphSAGE is to adhere to GNN and aggregate the neighbours' information by embedding them into each node. Generally, the GraphSAGE-based link prediction method involves propagating GraphSAGE networks forwards and propagating GraphSAGE networks backwards. The specific main procedures are demonstrated as follows:

A graph is represented by $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ denotes a set of nodes, and $\mathbf{E}$ is a set of edges between each pair of nodes. Specifically, the i_{th} node is indicated by $v_i \in \mathbf{V}$, and the features of all nodes are defined as $\mathbf{X}_v, \forall v \in \mathbf{V}$. Meanwhile, the adjacency matrix $\mathbf{A} \in \mathbf{R}^{|n|\times|n|}, A_{\boldsymbol{ij}} \in \{0,1\}$ is usually used to describe $\mathbf{E}$, which is a $|n| \times |n|$ square matrix. If an edge exists between node v_i and node v_j, then $A_{\boldsymbol{ij}} = 1$, otherwise $A_{\boldsymbol{ij}} = 0$.

Table 5.1 illustrates the embedding generation process in the GraphSAGE algorithm for all nodes. For starters, each node $v \in \mathbf{V}$ aggregates the representations of its neighbourhoods into a single vector $\{\mathrm{h}_u^{k-1}, \forall u \in N(v)\}$. This aggregation step relies on the representations learned at the previous iteration of the outer loop, and the first step starts at $k = 0$, which is regarded as the original input node features. After that, the node's current representation h_u^{k-1} has been concatenated with the aggregated neighbourhood vector $\mathrm{h}_{N(v)}^{k-1}$. Then, this combined vector is fed into a fully connected layer with nonlinear activation function σ, which transforms the representations to be used in the next step. The final representations are outputted at depth K as $\mathbf{z}_v \equiv$

$\mathrm{h}_v^u, \forall v \in \mathbf{V}$. The step of neighbours' representation aggregation has been accomplished by various aggregator architectures. The mean aggregator has been applied:

$$\mathrm{h}_v^k \leftarrow \sigma(\mathbf{W} \cdot MEAN(\{\mathrm{h}_u^{k-1}\} \cup \{\mathrm{h}_u^{k-1}, \forall u \in N(v)\ \})) \tag{5.4}$$

Table 0.1 The pseudocode of the GraphSAGE algorithm (Hamilton et al., 2017a).

Algorithm: GraphSAGE embedding generation (i.e., forward propagation) algorithm

Input: Graph $G = (\mathbf{V}, \mathbf{E})$; input features matrix $\{\mathbf{X}_v, \forall v \in \mathbf{V}\}$; depth K; weight matrix $\mathbf{W}^k, \forall k \in \{1, \ldots, K\}$; non-linearity σ ; differentiable aggregator functions $AGGREGATE_k, \forall k \in \{1, \ldots, K\}$; neighbourhood function $N: v \to 2^V$

Output: Vector representations $\mathbf{z}_v$ for all $v \in V$

$\mathrm{h}_v^u \leftarrow \mathbf{X}_v, \forall v \in \mathbf{V}$;
for $k = 1, \ldots, K$ **do**
  for $v \in$ **V do**
    $\mathrm{h}_{N(v)}^k \leftarrow AGGREGATE_k, (\{\mathrm{h}_u^{k-1}, \forall u \in N(v)\})$;
    $\mathrm{h}_v^k \leftarrow \sigma(\mathbf{W}^k \cdot CONCAT(\mathrm{h}_u^{k-1},\ \mathrm{h}_{N(v)}^k))$
  end
  $\mathrm{h}_v^k \leftarrow \mathrm{h}_v^k / \|\mathrm{h}_v^k\|_2, \forall v \in \mathbf{V}$
end
$\mathbf{z}_v \leftarrow \mathrm{h}_v^u, \forall v \in \mathbf{V}$

Figure 5.4 gives the schematic diagram of forwarding propagation in the GraphSAGE algorithm. The first key step is sampling with the depth K, which samples each node randomly. It is noted that the number of sampling is constant for each node, as it reduces the computational complexity for the improvement of computational efficiency. Assuming that the number of sampling is k, the selected opportunity of each neighbourhood is equal within the scope of depth K. If the number of neighbourhoods is less than k, the replacement method has been implemented until k neighbourhoods are selected totally. Otherwise, k neighbourhoods of all neighbourhoods are sampled at a random frequency. The next step is to aggregate information from the selected neighbourhoods into the target node so as to update the representation of the target node. Similarly, the representations of other nodes are updated. Finally, the aggregated representations are used to predict labels. The edge representation between two entities is presented through these two nodes' representations. Meanwhile, the edges are labelled in the link prediction task.

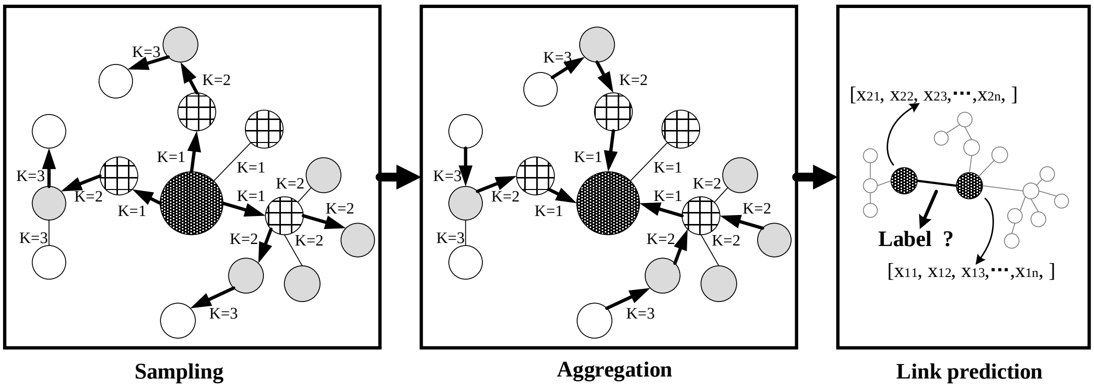

Figure 0.4 The schematic diagram of forwarding propagation in the GraphSAGE algorithm.

In this study, the sample of edges is divided into two types: positive samples and negative samples. The positive sample is defined as the existing edge between two nodes in the original graph, and the negative sample is regarded as the non-existed edges similarily. In this context, the relation completion task can be considered as a binary classification. For learning the weights of aggregators and embeddings, the cross-entropy function is usually utilised as a loss function in the binary classification task, as shown as follows.

$$\mathcal{L} = -\frac{1}{N}\sum_i \mathcal{L}_i = \frac{1}{N}\sum_i -[y_i \cdot \log(p_i) + (1-y_i) \cdot \log(1-p_i)] \quad (5.5)$$

where: $y_i \in \{0, 1\}$ is label, $y_i = 1$ for the positive sample, and $y_i = 0$ for the negative sample; p_i is the predicted probability, which sample i is a positive sample.

Figure 5.5 illustrates an overview of the link prediction model based on GraphSAGE. For starters, KGs are used to derive and compute the node-feature matrices and adjacency matrices. In this context, KGs are expressed in terms of the attributes of the entities (defined as node features) and the relationship features (regarded as adjacency matrices). Then, the two types of matrices are fed into a two-layer GraphSAGE model, which is a supervised training model. The features of each node have been updated by its neighbour within the share parameters. After propagating two GraphSAGE layers, the updated representations of each node have been obtained. Compared with the common GraphSAGE model, the sample labels are different from the node classification. Since the goal task is regarded as the link classification, the edges are divided into two different types: positive edges and negative edges. In terms of positive edges, the real-existing edges are defined as positive edges. Similarly, the non-existing edges are considered negative edges. In the next step, the embeddings of each edge have been represented by two linked entities in the proposed model. Accordingly, the proposed model employs the rule that minimises the classification error to achieve the final embeddings of nodes. In this context, the proposed model has been already trained to estimate the possible edges.

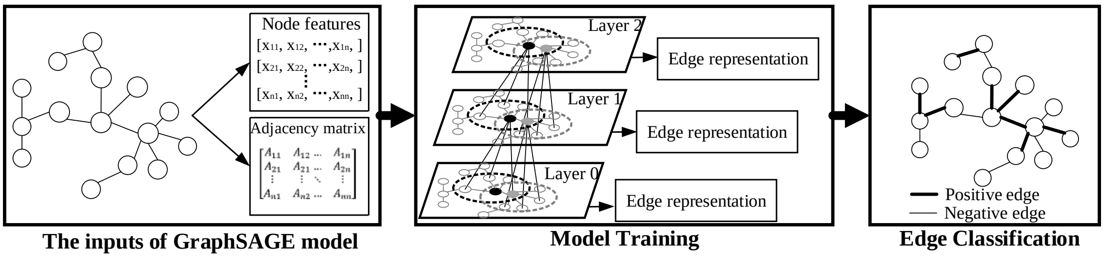

Figure 0.5 The architecture of the two-layer GraphSAGE model.

5.2.3 Domain-centric KG Construction and Visualisation

The key steps of KG construction and visualisation are present in this section. The potential triplets have been discovered and generated by the GraphSAGE-based relation completion model. After that, all of these triplets are integrated into a united triplet repository. Through using open-source platforms, the RDF language is used to describe each triplet for the domain-centric knowledge representation. In other words, the domain-centric KG is constructed and stored in these platforms. Meanwhile, graph visualisation tools are often embedded in these platforms. In this context, KG visualisation is usually accompanied by KG construction. In fact, there are also several independent open-source software to achieve the KG visualisation. At present, there are several open-source graph databases available at this time, including *Neo4j*, *Mongo*, *Gephi*, and *Grakn*. Due to the operational ease and highly performant visualisation, the *Gephi* has been employed to visualise the generated KG.

5.3 Case Study

Cold rolling is recognised as an important pillar of global economic growth in the production of electrical steel strips. In the manufacturing of metal sheets and strips, cold rolling is an important process because of its advantages with regard to accuracy, efficiency, and output rate. Presently, cold rolling contributes to the improvement of the properties of steel strips on changes both in the microstructure and thickness of the steel. Since the properties that have been improved include surface smoothness, tensile strength, yield strength and hardness, cold-rolled products usually have superior mechanical properties, small dimensional tolerances and high-quality surfaces (Wu et al., 2021). Furthermore, as science and technology continue to advance, the quality requirements for steel strip products from cold rolling processes are becoming more detailed and demanding. Therefore, it is imperative increasingly to analyse and monitor the quality of cold-rolled products.

This section set up a real-world experimental study on the cold rolling process to validate the proposed framework. In this case, a domain knowledge graph (KG) was constructed to address the issue of strip breakage in the cold rolling process of the steel industry. The KG was designed to capture the complex relationships between various factors that contribute to breakages, including material properties, processing conditions, and equipment parameters. The KG was constructed using a combination of ontology design, relation extraction, and relation completion techniques.

5.3.1 Ontology Design for Strip Breakage in Cold Rolling

In regard to the modern steel industry, steel strips are produced by cold rolling in a high-speed, high-precision, and continuous process. It is not uncommon for cold rolling to encounter certain defects in the same manner as the process of metal forming. Technical reports indicate that various types of defects exist in steel strips from cold rolling production. It is worth noting that the technical reports used in the knowledge extraction process contain detailed information on the various types of defects encountered in steel strips during cold rolling production. These reports provide insights into the causes and effects of edge cracking, burrs in the centre, surface defects, and buckling. The reports further explain how a high silicon concentration in electrical steel leads to brittleness, resulting in breaks during cold rolling. The technical reports serve as a valuable resource for understanding the complex factors that contribute to breakage and for developing effective strategies for predicting and preventing this failure. The most common defects encountered in the sheet metal rolling process include edge cracking, burrs in the centre, surface defects, and buckling. Especially electrical steel is an iron alloy containing high percentages of silicon. Alloys containing a high silicon content have a low magnetisation loss as a result of the high electrical resistivity. As a result of a high silicon concentration, the strip becomes brittle, resulting in breaks during cold rolling. As the most serious defect, strip breakage needs to be paid special attention to as it causes huge financial losses. Specifically, strip breakage has damaged rolls and mill accessories badly, not only the increase of production costs. Hence, the information integration of different resources around the cold rolling process of the steel industry contributes significantly to the prediction of this failure. The proposed model was validated by carrying out an experimental study using real-world data in this section. An electrical steel manufacturer equipped with a reversing mill for cold rolling provided the experimental data. Figure 5.6 illustrates the entire manufacturing workflow that contributes to cold-rolled coil in the steel industry. The relevant procedures contribute to the influence in different degrees of the quality of cold-rolled products, including the hot rolling process, properties of hot-rolled coils, annealing process, pickling process, cold rolling process, quality inspection, etc.

The hot-rolled coils have been fed into the annealing machine to increase their ductility, reduce their hardness, and make them more workable, as this processing alters the physical and chemical properties of the coil material. Then, a pickling treatment is performed on metal products to remove impurities, such as stains, inorganic contaminants, rust, and scale. In conjunction with emulsion, the steel coils that have been processed are transmitted to a cold rolling mill for flat deformation. The above steps are repeated to get the finished size. After straightening these coils, the quality of coil-rolled products was inspected by technicians who work on the site. The strip breakage defect has been selected and marked manually. Finally, the qualified cold-rolled coils are cut into the required length for packing and storing.

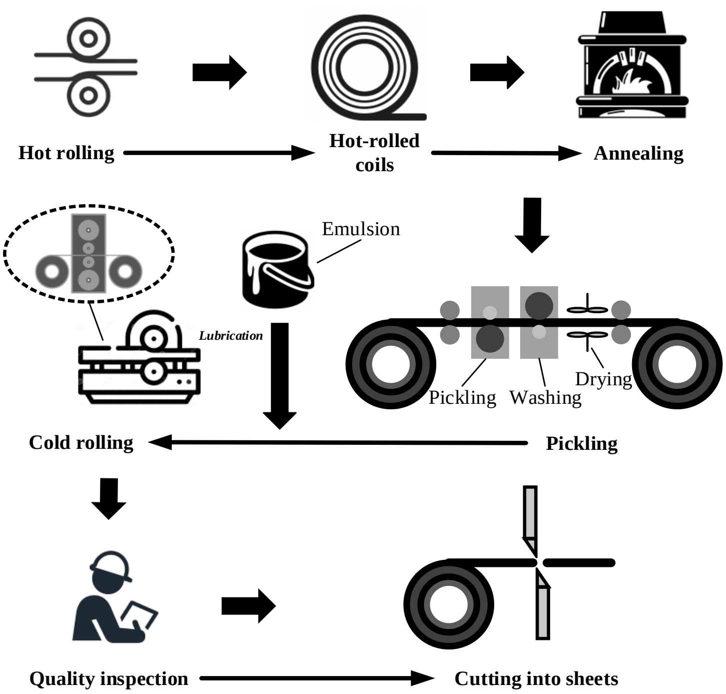

Figure 0.6 The manufacturing workflow of cold-rolled products in the steel industry.

Following the manufacturing workflow of cold-rolled products, the strip breakage knowledge should be derived from steel-coil material, chemical reagents, transportation, processing, treatment, etc. The class hierarchy considered the concepts of processes, facilities, products, operations, parameters, chemicals, fault diagnosis, etc. Meanwhile, these seven parts were regarded as the top classes, as shown in Figure 5.7. Based on the provided concepts, the subclasses of each top class were determined subsequently. Since the cold-rolling process plays the most crucial role in influencing the strip breakage phenomenon, the processes were divided into the cold-rolling process and other processes. Facilities were composed of the mill, rolls and other equipment. For operations, two different operations were recognised in the cold rolling, including cold-rolling operation and observation. In terms of products, coils were classified into cold-rolled coils and hot-rolled coils. Chemicals were regarded as the element contents. Process parameters and product parameters were two types of parameters. For fault diagnosis, the cold-rolled coils were labelled as *'Normal'* or *'Strip-breakage'*.

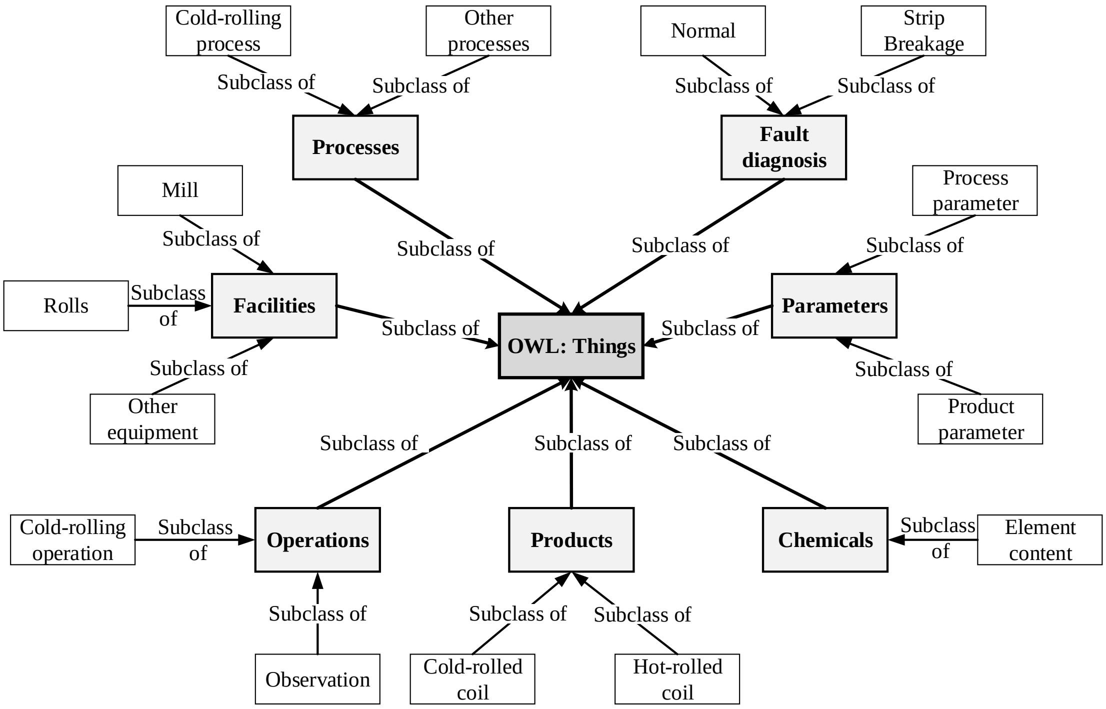

Figure 0.7 The hierarchical structure of top classes of the strip breakage-centric ontology.

Table 5.2 presents details of relevant concepts and characteristics extracted from multiple sources. In this section, five resources are regarded as the contributing factors to the strip-breakage phenomenon, including the hot-rolling process, annealing, pickling, emulsion, cold-rolling process and quality inspection. In this context, the specific strip-breakage knowledge has been extracted from the five resources in the next stages. The dataset is collected and stored from these resources, which covers a production period of six months. Specifically, this historical dataset contains 1254 samples, including 94 variables.

Table 0.2 Details of relevant concepts and characteristics extracted from multiple sources.

Sources	Concepts and variables
The hot-rolling process	Hot-rolled coil properties, such as chemical contents, crown, quench temperature etc.
Annealing & Pickling	Annealing temperature, Jetflow speed etc.
Emulsion	Dirt result, pH, conductivity, chloride index etc.
Cold-rolling process	The rolling operation, equipment, machine performance, tension, measured parameters etc.
Quality inspection	Cold-rolled coil properties, such as weight ingoing, width, weight outgoing etc.

In regard to the relation extraction, three correlation methods are selected to decide whether the relations between each pair of variables from the relevant sources are shown in Table 5.2, including Pearson correlation coefficient, Kendall Tau rank correlation coefficient, and Spearman rank relational coefficient. The results of these three correlation methods are shown in Appendix B.

After taking absolute values, the heat maps of the three types of correlation results are presented in Figure 5.8, Figure 5.9, and Figure 5.10 accordingly. In the light of the absolute values, the correlation results have been classified into five levels, which are very weak relation or no relation ($0 < |correlation\ coefficient| < 0.2$), weak relation ($0.2 \leq |correlation\ coefficient| < 0.4$), moderate relation ($0.4 \leq |correlation\ coefficient| < 0.6$), strong relation ($0.6 \leq |correlation\ coefficient| < 0.8$), and a very strong relation ($0.8 \leq |correlation\ coefficient| \leq 1$). When the correlation result fell within the range of very weak relation or no relation, the relation between two different variables was supposed to be the non-existing relation. Otherwise, the relation was considered as the existing relation. For example, since the absolute Pearson correlation value is 0.998 between *'Z2Temp min'* and *'Z2Temp min'*, it has been supposed that there is a relation between these two entities. Similarly, there is no relation between *'Crown min'* and

'Width' as the absolute Pearson correlation value is 0.009 between these two variables. The absolute Kendall Tau correlation between *'Gauge'* and *'Weight outgoing'* took 0.435, which falls within the range of moderate relation. It means that the relation between these two variables was considered as the existing one. On the contrary, there is no relation between *'Gauge'* and *'TOFF'*. Since the absolute Spearman correlation value is 0.402 between *'GasFlow min'* and *'Z6Temp std dev*, it has been supposed that there is a relation between these two entities. Similarly, there is no relation between *'GasFlow min'* and *'Heavy end'* as the absolute Pearson correlation value is 0.004.

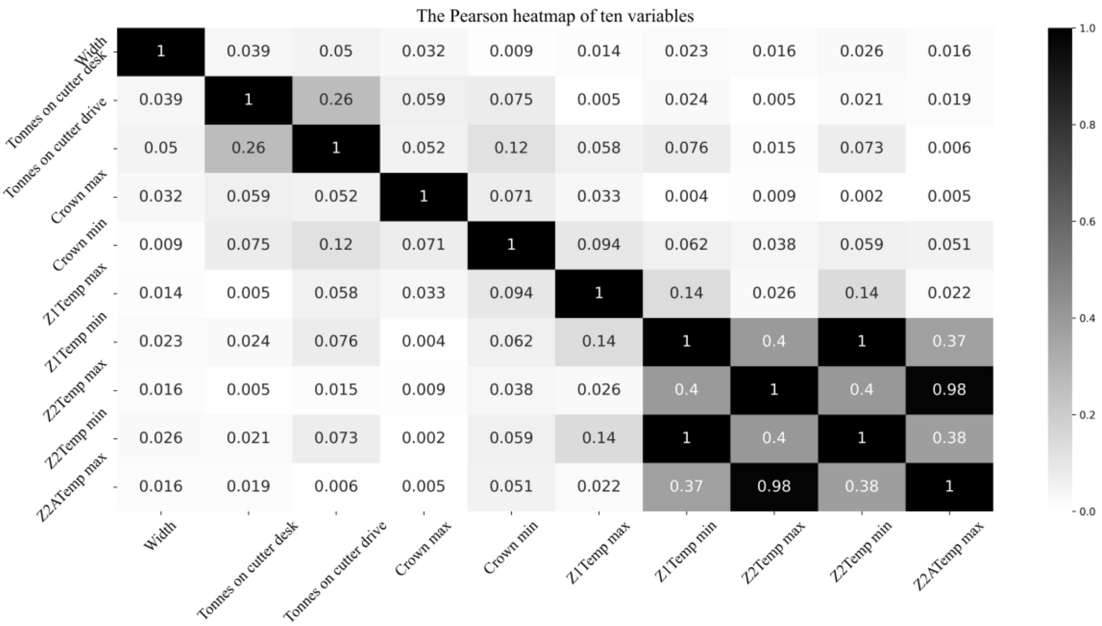

Figure 0.8 The heat map of Pearson correlation coefficients.

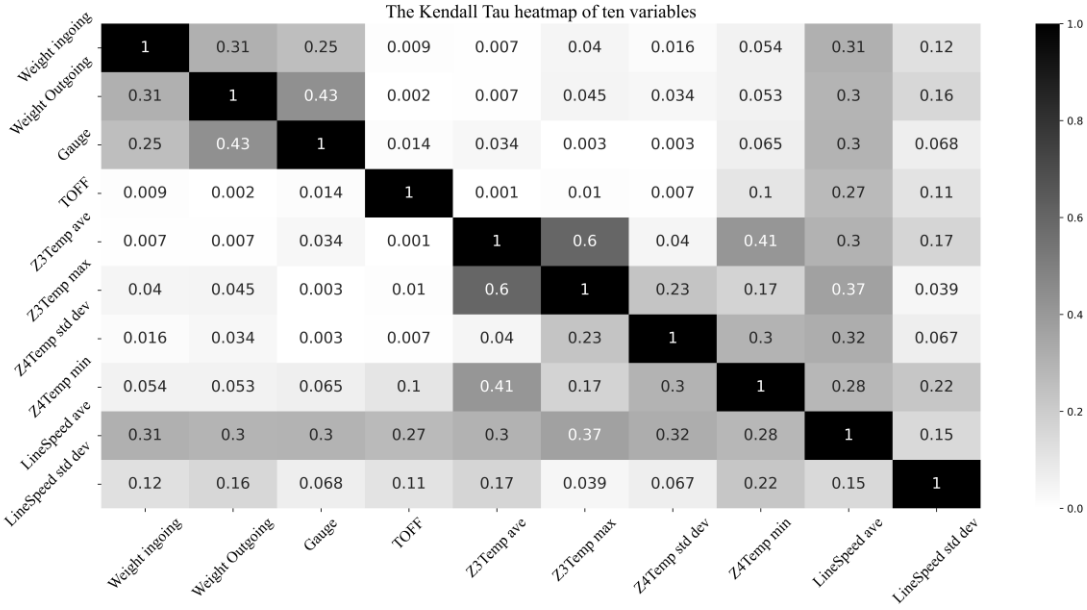

Figure 0.9 The heat map of Kendall Tau rank correlation coefficients.

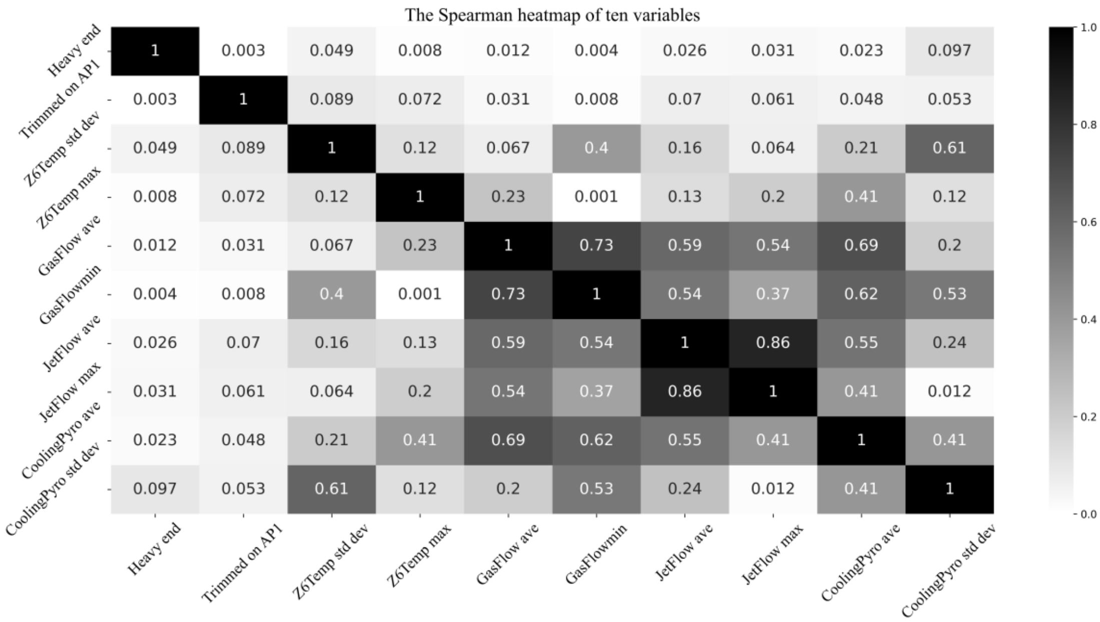

Figure 0.10 The heat map of Spearman rank correlation coefficients.

In this section, the strip breakage ontology was finished in open-source software (Protégé5.5), which supports OWL. The concepts, properties, and associated relationships of the strip-breakage ontology were defined in this software. Figure 11 shows an illustrative example of the classes, object properties and data properties of the strip-breakage ontology. There are seven main concepts in this ontology, including *'chemicals'*, *'cold-rolled coil patterns'*, *'facilities'*, *'manufacturing processes*, *'operations'*, *'parameters'*, and *'products'*. Totally, the second level of ontology comprises 20 subclasses, such as *'OES-SoIAI'*, *'Mill'*, *'Observation'*, *'accumulator'*, *'pickling process*, etc. Meanwhile, the object properties and the data properties were defined as well for a better understanding of the concepts in Figure 11.

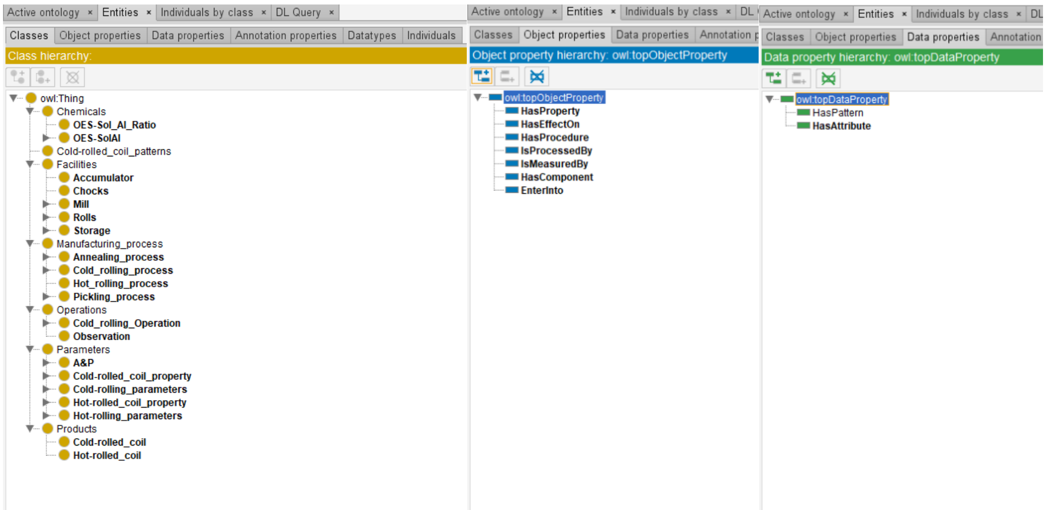

Figure 0.11 The classes, object properties and data properties of the strip-breakage ontology.

A performant Web-based platform, namely *'OOPS!'* (Cao et al., 2022), has been introduced to detect the pitfalls of ontologies in this section. After designing and constructing the strip-breakage ontology, the source codes were uploaded to the *'OOPS!'* website for quality evaluation of ontologies. As shown in Figure 12, the evaluation result means the designed ontology is error-free in logic consistency, reasoning and applicability.

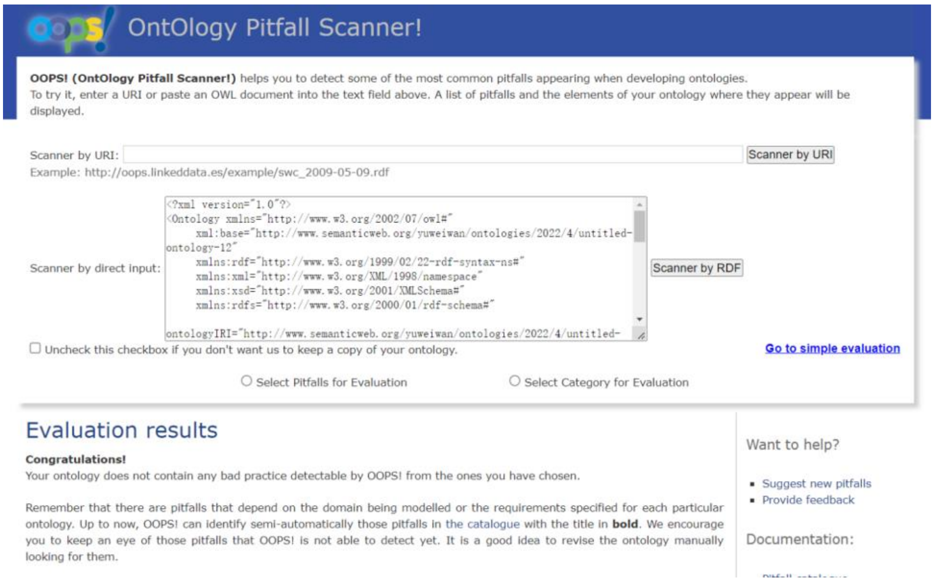

Figure 0.12 The evaluation result of the strip-breakage ontology.

5.3.2 Relation Completion for Strip Breakages

In this section, the experiments were conducted to predict the potential edges for relation completion of the strip-breaking-centric KG. A brief overview of the dataset is provided in the following. The dataset contains 13113 samples, which are divided into two types (existing edges and non-existing edges). Regarding training and testing data split, the 5-fold cross-validation tests were conducted to evaluate the performances, and the mean values of the five folds were outputted and labelled. Specifically, 80 per cent of the overall data were selected for training, and the remaining 20 per cent was reserved for testing.

As mentioned above, the experiments were conducted using our proposed two-layer GraphSAGE model. Meanwhile, three common machine learning algorithms were built to compare the performances, including back propagation neural network (BPNN), SVM, and random forest (RF). For the BPNN model, the hidden unit number and learning rate were set to 50 and 0.71, respectively, and the training epoch was set to 500. In terms of the SVM model, the kernel function was set as a radial basis function. Lastly, for the RF model, the number of estimators was set to 600. On the basis of that, the edges were fed respectively into three different models (BPNN, SVM, and RF) after calculating by the representations of entities. Meanwhile, as shown in

Figure 5, the representations of graphs (entity matrices and adjacency matrices) were fed into the two-layer GraphSAGE mode to predict the potential for relation completion.

Moreover, in the task of relation completion, the goal is to output the potential edges by predicting the edge types. Since the edges are classified into two types (existing edge and non-existing edge), the link prediction is considered a binary classification. In this context, five following metrics were introduced to evaluate the performances of the proposed model, including accuracy, precision, recall, F1-Score, and False Alarm Rate (FAR):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5.6}$$

$$Precision = \frac{TP}{TP+FP} \tag{5.7}$$

$$Recall = \frac{TP}{TP+FN} \tag{5.8}$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{5.9}$$

$$FAR = \frac{FP}{FP+TN} \tag{5.10}$$

where: TP, TN, FP, FN mean true positive, true negative, false positive, and false negative, respectively.

Table 5.6 gives the detailed binary classification result by five different metrics of BPNN, SVM, RF, and GraphSAGE models. It is obvious that the GraphSAGE model showed advantages over the other three machine learning models. Specifically, the GraphSAGE model surpassed the other machine learning baselines in terms of accuracy (82.02%), precision (76.6%), recall (79.26%), and F1-Score (77.9%). Moreover, the experimental results show that a smaller value on the FAR metric

(16.14%) was found using the GraphSAGE model than in the other three models. In this context, the GraphSAGE classifier performed well across all different performance metrics, showing that incorporating the neighbourhood information is better for the relation completion task in this study. Moreover, for the considered three benchmark models, it can be observed that the RF model achieved much better performances on all different metrics than SVM and BPNN models. Although the BPNN classifier performed better on recall than the SVM classifier, these two classifiers performed at about the same level during the comparison experiments.

Table 0.3 The binary classification result of four models.

Models	Accuracy	Precision	Recall	F1-Score	FAR
BPNN	76.99	70.45	73.17	71.78	20.46
SVM	77.34	71.13	72.96	72.04	19.74
RF	78.06	72.17	73.48	72.82	18.89
GraphSAGE	82.02	76.6	79.26	77.9	16.14

Figure 13 illustrates an example of relation completion through link prediction using the GraphSAGE model. After training the GraphSAGE classifier, the existing graphs were fed into the proposed model to learn the edge representations and discover the potential edges. As shown in Figure 13, the missing relations were found to complete the strip-breakage KG, including the relation between *'DSP width'* and *'Z6Temp std dev'* and the relation between *'Gauge average (microns)'* and *'Crown min (microns)'*.

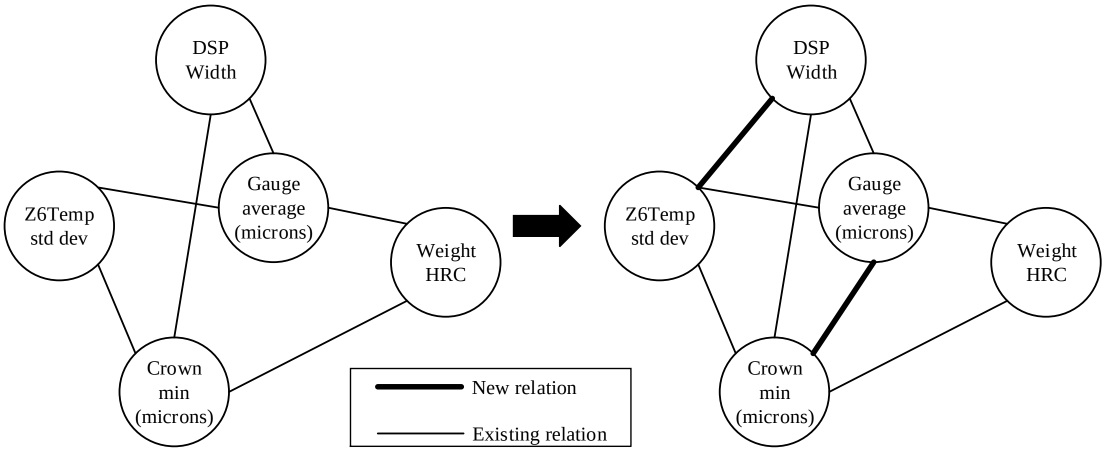

Figure 0.13 An illustrative example of relation completion.

5.3.3 Breakage-centric KG Visualisation

Based on the hierarchy structure refinement and relation completion, all triplets were integrated and linked to constructing the strip-breakage-centric KG of the cold rolling process. In this section, the free software *Gephi* was implemented to store and visualise the generated KG as it performs well on operation and function. Specifically, 2295 triplets are integrated and imported into *Gephi* to construct and visualise the strip-breakage KG in the cold rolling process of the steel industry. In Figure 5.14, an entire strip-breakage-centric KG composed of seven subclasses is present. The entire strip-breakage KG contained 230 entities and 2295 relations in total.

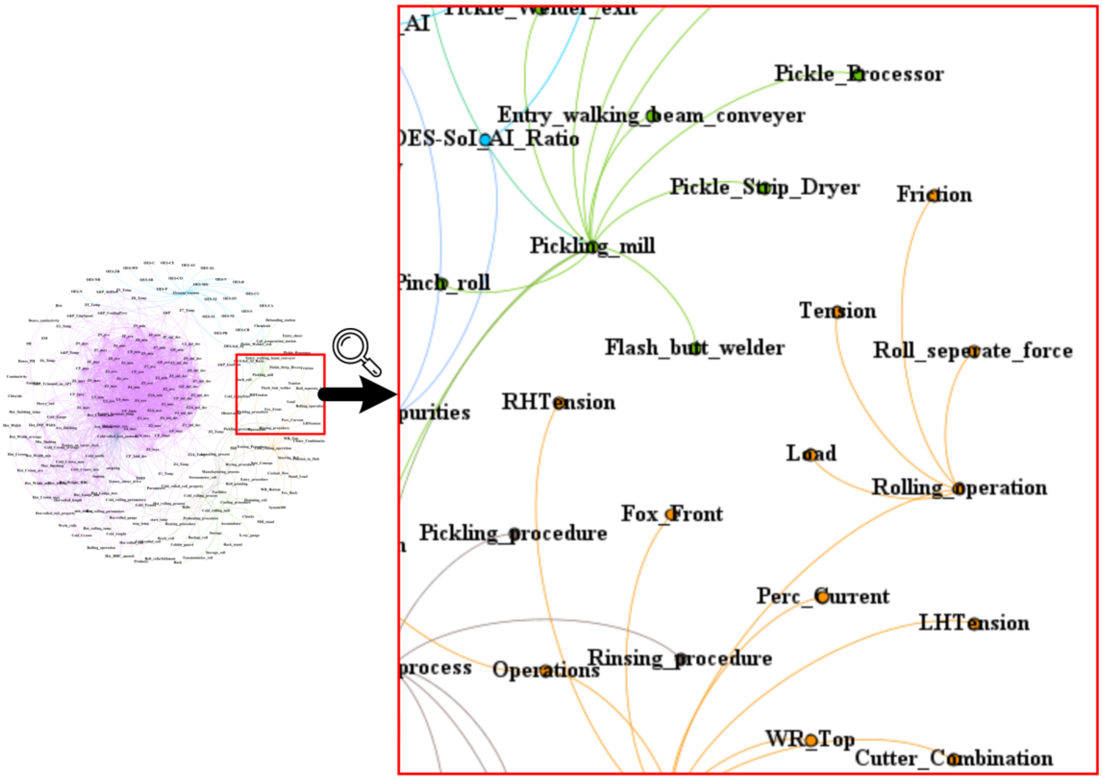

Figure 0.14 The entire strip-breakage KG of the cold-rolling process.

5.4 Discussion

In real-industrial scenarios, the relevant information from various modalities or sources is usually related to one intricate phenomenon, process, or system. The integration of data from multiple sources is essential for obtaining more comprehensive and expressive information than that obtained from any one source alone. In this context, this study presents a general framework of domain-centric KG construction as a semantic bridge which leverages multi-source data to provide a united and versatile view of understanding a phenomenon, process, or system. The main contributions of this study lie in hierarchy structure refinement developed to reduce the participation of knowledge alignment from domain experts and relation completion deployed to enhance the reliabilities of the KGs. The hierarchy structure refinement and relation completion modules have been validated using a real-world case study from the cold rolling process of the steed industry, and results from the different metrics confirm the accuracy and efficiency of the proposed framework.

Firstly, six sources of 5M1E, which contribute to one phenomenon, were considered to cover the information versatility of ontology in data-driven intensive scenarios comprehensively. The developed ontology was populated with instances after predefining the hierarchical structure of top classes in the steel industry. These top classes were defined around the strip-breakage phenomenon in the cold rolling process, including processes, fault diagnosis, facilities, parameters, operations, products, and chemicals. With the aim of refining and completing the proposed hierarchy structure, three correlation methods were developed to reduce the participation of knowledge alignment from domain experts, thereby avoiding time consumption and massive labour. As mentioned above, these three correlation methods (Pearson correlation coefficient, Kendall Tau rank correlation coefficient, and Spearman rank relational coefficient) were proposed as a quantifiable criterion to discover and decide the relations between each pair of variables for hierarchy structure refinement. Based on three experiments on the correlation coefficient, $|relational\ coefficient| \geq 0.2$ was supposed to be relations, or it was not. By extracting relation automatically, the experimental results illustrated that the proposed framework of hierarchy structure refinement avoids the extensive knowledge alignment by domain experts. Moreover, the Web-based platform was introduced to detect the pitfalls of ontologies. It was confirmed that quantifiable and fast evaluation provides an objective view of the quality and feasibility of ontologies. Combined with the quality evaluation manually, it was more effective and efficient than that of a single evaluation in constructing a reliable industrial ontology. The solid feasibility of strip-breakage ontology design shows that the general proposed framework can be broadened to other industrial ontologies.

In addition, the existing knowledge was exploited to discover the missing relations between each couple of entities for KG completion, thereby enhancing the reliabilities of the KGs. Specifically, the missing relations were inferred and added to the existing KG, which resulted in increasing the number of edges. Recently, as a set of advanced techniques for graph representation learning, GNN techniques have been full of vigour and tremendous potential. Based on that, a GraphSAGE-based link prediction model was built to complete the strip-breakage KG. Specifically, since the edges were classified into two types (existing edges and non-existing edges), the relation completion task was transferred into a binary classification task. Moreover, the three most common machine learning algorithms (BPNN, SVN, and RF) were developed to compare the performances. In this context, five metrics were deployed to evaluate the proposed model roundly, including accuracy, precision, recall, F1-Score, and FAR. The experimental results of relation completion indicated the power of the GNN-based approach for link prediction in constructing a domain-centric KG. Compared with the most common machine learning models (BPNN, SVN, and RF), the GraphSAGE classifier achieved the most performance on five different metrics consistently. The main reason is that this classifier captures the neighbourhoods' information on the inherent graph structure to discover the missing and potential relations. The stable performances show the ability of robustness to infer links for relation completion, which can be applied in other industrial scenarios.

One of the significant contributions of this study was the hierarchy structure refinement module, which was developed to reduce the participation of domain experts in knowledge alignment, thereby avoiding time consumption and massive labour. Three correlation methods were proposed as a quantifiable criterion to discover and decide the relations between each pair of variables for hierarchy structure refinement. The experimental results illustrated that the proposed framework of hierarchy structure refinement avoids extensive knowledge alignment by domain experts. Furthermore, the web-based platform was introduced to evaluate the quality and feasibility of ontologies objectively. The proposed methodology was effective and efficient in constructing a reliable industrial ontology.

Secondly, the proposed framework leveraged existing knowledge to discover missing relations between entities for KG completion, thereby enhancing the reliability of the KGs. The GNN-based approach for link prediction demonstrated the power of the GraphSAGE classifier for constructing a domain-centric KG. The experimental results showed that the GraphSAGE classifier consistently outperformed the most common machine learning models, namely BPNN, SVN, and RF, on five different metrics. The reason for the GraphSAGE classifier's superior performance was its ability to capture the neighbourhood's information on the inherent graph structure to discover the missing and potential relations. The stable performance of the proposed framework demonstrated its robustness in inferring links for relation completion, which can be applied in other industrial scenarios.

Overall, the proposed framework's effective validation in the real-world case study and the advantages it offers over traditional methods suggest its potential for broader applications beyond the steel industry. The contributions of the hierarchy structure refinement and relation completion modules provide a stepping stone towards developing reliable and robust domain-centric KGs. However, further research is needed to explore the proposed framework's generalization to other domains and to investigate the scalability of the proposed methodology for constructing large-scale KGs.

5.5 Summary

It is a typical characteristic that information from various sources contributes to one phenomenon in data-intensive industries. The use of KGs has been demonstrated to be an effective way to bridge the semantic gap for covering the versatility of a phenomenon in industries. Therefore, it has been proven that KG has been developed as a semantic organisation to tackle massive multi-sourced information embedded in manufacturing processes and products. With the aim of building a reliable domain-centric KG, this study proposed a framework based on hierarchy structure refinement and relation completion. An illustrative real-world case of the cold rolling process in the steel industry was conducted to validate the proposed framework.

The process of domain-centric industrial ontology construction was first described in this study. Specifically, the multi-source information was extracted and integrated to design a comprehensive strip-breakage ontology of the cold rolling process. Meanwhile, three correlation methods are applied to extract relations between each pair of variables, where $|relational\ coefficient| \geq 0.2$ was supposed to be the existing relation. Otherwise, it is considered a non-existing relationship. Subsequently, the missing and potential relations were detected and predicted through the existing knowledge in strip-breakage KG, achieving better performances on five different metrics by the GraphSAGE model than the other three baseline machine learning models. Lastly, all triplets were linked and imported into the *Gephi* software for strip-breakage-centric KG construction and visualisation.

Chapter 6 Knowledge Graph-Aided Information Fusion for Multi-faceted Fault Modelling in Steel-making

6.1 Introduction

As discussed in Chapter 4, it has become progressively more evident that a single data source is unable to comprehensively capture the variability of multi-faceted concepts in steel-making, which has diverse semantic orientations. Therefore, multi-faceted conceptual modelling is often conducted based on multi-sourced data covering indispensable aspects, and information fusion is frequently applied to cope with the high dimensionality and data heterogeneity. Meanwhile, in Chapter 5, we proposed a semantic approach by constructing a fault-centric KG for the processing and managing of semantic data and domain knowledge in steel-making. However, besides the ability for knowledge management and sharing, KG can aggregate the relationships of multiple aspects by semantic associations, which can be exploited to facilitate multi-faceted conceptual modelling based on heterogeneous and semantic-rich data.

Given this context, in this chapter, KG was applied to facilitate the construction of unified standard representations for data fusion by representing knowledge in the form of entities and relations. Although domain-independent (open-world) KGs such as Wiki are widely used, domain-dependent KGs offer a greater range of benefits and can provide a positive return on investment [12]. Meanwhile, it is common to capture domain knowledge in KGs, which are then used to enrich semantics with specific conceptual representations of entities [13]. In this case, KGs can be used as a basis for developing multi-faceted modelling by extracting the semantics, which is retrieved from different vocabularies and semantic repositories, which are used to enrich the semantic description of resources using annotations.

Based on the above analysis, with the support of a knowledge graph, this chapter proposes a fusion architecture for the modelling of multi-faceted concepts. The purpose of this architecture is to capture data related to the concept-centric domain derived from heterogeneous sources into a formal knowledge graph representation illustrating the concept-centric domain. To this end, a multi-faceted modelling method for fault diagnosis based on knowledge graphs and data fusion was developed. Based on equipment mechanisms derived from concept-centric data and empirical knowledge rules, a concept-centric knowledge map was derived with temporal characteristics surrounding the manufacturing failure. Subsequently, concept-centric knowledge graphs were constructed, and multivariate time-series data was transformed into a temporal graph representation of the data sequence. Finally, a Graph Convolution Network (GCN) model was applied to extract features from these temporal graphs, and these features are fed into a Temporal Convolution Network (TCN) model for fault concept modelling.

The remainder of this chapter is organised as follows. The flowchart of the proposed methodology is illustrated and described for multi-faceted conceptual modelling with the support of a knowledge graph in Section 6.2. Section 6.3 presents the conducted experiments, and the results are shown in Section 6.4. These results are discussed in Section 6.5. Section 6.6 summarise this chapter.

6.2 Methodology

In this section, we proposed the framework of this multi-faceted modelling approach shown in Figure 6.1. Firstly, an initial set of concept-relevant data is gathered from various sources, and the concept-centric features are identified and utilised for the development of the CKG backbone. Then, the empirical knowledge recorded from technical documents and the existing knowledge repository is processed by NLP (natural language processing) tool for triplets' generation. Meanwhile, the structured data is mapped to its classes using the data to RDF (resource description framework) technique, and the CKG backbone is generated with the integration of those triplets. Thirdly, the time-series process data is transformed into a stack of temporal graphs under the CKG backbone. Finally, through spatial-temporal graph convolution using a temporal GCN model, the aggregated feature representing the intra-faceted and temporal characteristics of the graphs is extracted. The aggregated feature is then fed into a TCN model for multi-faceted conceptual modelling.

The primary contribution of the proposed multi-faceted modelling approach shown in Figure 6.1 is the integration of heterogeneous data sources, including structured and unstructured data, to construct a CKG backbone. The backbone provides a foundation for the generation of triplets from technical documents and the existing knowledge repository. Additionally, the approach employs spatial-temporal graph convolution and temporal convolutional networks to extract the aggregated feature representing the intra-faceted and temporal characteristics of the graphs for multi-faceted conceptual modelling. This approach can effectively capture the interrelationships between different concepts and the evolution of the knowledge graph over time, which is crucial for real-world applications.

The literature review informed several aspects of the proposed approach. Specifically, the use of natural language processing techniques for triplet generation from technical documents and existing knowledge repositories is a widely adopted strategy. The mapping of structured data to its classes using data to RDF conversion is also a well-established technique in knowledge graph construction. Moreover, the use of graph convolutional neural networks for graph representation learning has gained considerable attention in recent years. The proposed approach builds on these established techniques by integrating them into a comprehensive framework that addresses the unique challenges of multi-faceted modelling.

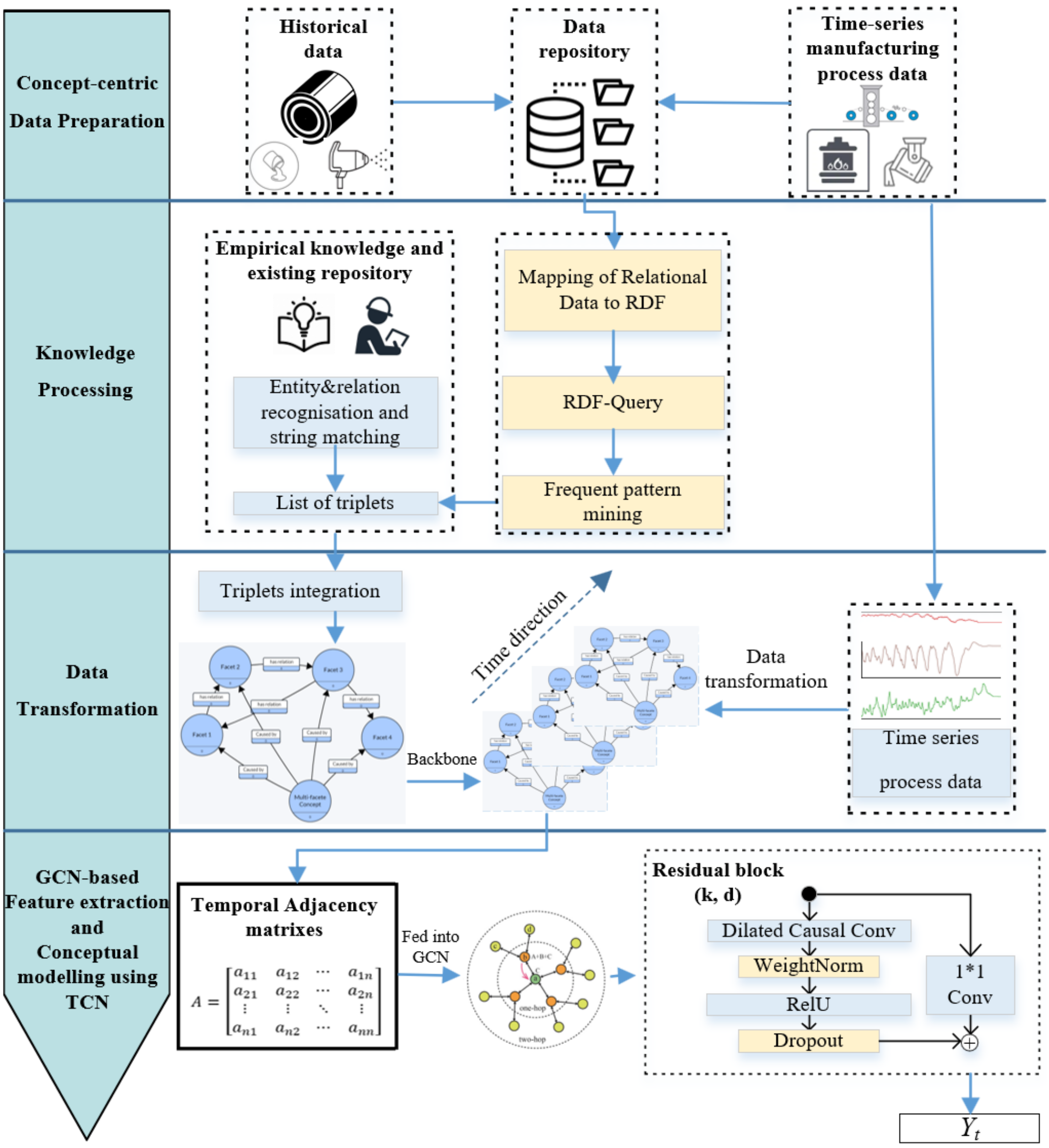

Figure 0.1 The overall architecture of KG-supported multi-faceted conceptual modelling.

6.2.1 Fault Concept-centric Knowledge Graph Construction

The internal relationships and the temporal characteristics of the concept-related features were captured using a knowledge graph in this section. By semantic mapping, we relate the semantics of temporal process data with knowledge graphs. This layer primarily integrates semantic relationships between the facets of our concerning concept, as shown in Figure 6.1. Generally, the knowledge graph is represented by triples, which include the subjects, predicates, and objects, or (h, r, t). In symbolic notation, *h* represents the subject, and *t* represents the object. These two nodes are referred to as the head node and tail node, respectively. An edge or relationship is the predicate, as expressed by the *r*. Facts are denoted by each instance (triple). For the structured data extracted from the enterprise information management platforms, the data to RDF (Corcho et al., 2020) approach is applied to structured data to its corresponding properties and classes.

In terms of the unstructured information stored in the existing knowledge repository, our intention is to transform such information into an available computational form. As this information is merged with the information on the process behaviour, the material, the task schedule, and other facets concerning the concept, the integration of such information can thereby aid intelligence decision-making.

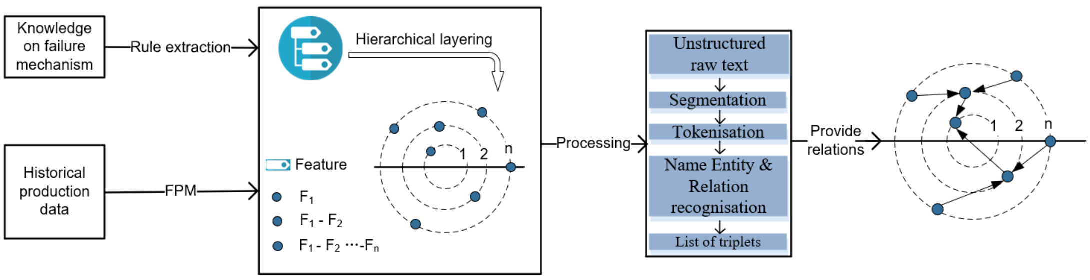

Figure 0.2 Semantic illustration of CKG construction.

As shown in Figure 6.2, the attribute is used as an edge. Attributes consist of the specific description of an entity, which represents several facets surrounding the concept. These facets can be obtained from the existing knowledge repositories, such as device description files and production diaries. As part of the construction of CKG, string matching is primarily used to obtain the names and attributions from unstructured data. In accordance with the extraction process, the data is organised into triples and placed in the knowledge graph. A knowledge graph will serve as a repository for aggregating and conveying real-world information. Entity nodes are represented by nodes, while edge nodes are represented by relationships between entities. Knowledge about the properties of the concept may be introduced through the construction of CKG. Therefore, as a result of this information, the model can better fit the concept-centred feature representation and gain an understanding of more complex correlations between these features.

With the generation of a list of triplets from both structured and unstructured data, the CKG will be generated by triplets mapping and integration. Firstly, the rules of mechanism knowledge are summarised, and extract indexes I_1, $I_2 \ldots I_n$ are extracted as the classification basis. Accordingly, a level-1 concept node F_1 of the corresponding level is generated, and the same to F_2 until F_n. At the same time, the features of operational data are extracted as node attributes. Secondly, the operation data of the equipment is marked according to the regular nodes, and the concept classification and prediction of the operating data are carried out to generate node relations at different levels. Thirdly, the rule nodes and node relations are saved in the form of the triplets of inferior concept nodes, relation, and superior concept nodes. The rule of the mechanism chain from level 1 to level N is constructed, which contains node information and the relationship between nodes. Finally, according to the relationship between nodes, multiple regular mechanism chains are fused into a complete rule map. In view of mechanism knowledge, the concept-centric features are used as a classification basis to summarise rules, the summarised indicators and attribute names are symbolised, and the symbols of indicators are combined as rule nodes. Nodes of the same grade are divided into the same level. The node level depends on the type and quantity of data indicators. Lower-level nodes and upper-level nodes are subordinate to each other, indicating that upper-level nodes are further classified into lower-level nodes, as shown in Figure 6.2.

6.2.2 Temporal Graph-based Data Transformation Relation Completion

It is necessary to note that default time stamps are attached to the original process data, which means that each time series data set is generated for a specific period of time. For a better understanding of the temporal characteristics of the data, the intention is to construct a model that incorporates both the state characteristics of the time series and the internal relationships of the concept-centric features. This work is characterised by the fact that sequence diagnosis is considered at each stage as a temporal event with corresponding time labels and attribute values associated with its occurrence. It is necessary to take into consideration that the collected data sequence is arranged into the Spatiotemporal graph that corresponds to the sequence at each time stamp as part of establishing possible interactions between attributes and their temporal dependency in the dynamic working system. In this context, the concept-centric knowledge graph can be expressed as follows:

$$\mathrm{G} = (\mathrm{V}, E, \mathrm{A}) \tag{6.1}$$

with N nodes $v_i \in \mathrm{V}$ is the vertices set, and edges E is the edges set.

According to our previous study, given a manufacturing process data $X_\tau = (X_1,\ X_2,\ X_{t-\tau+1} \dots, X_t) \in R^{t*N}$ of the t time intervals, τ is the time interval size which is the window size, and our target is to predict whether the failure will take place within this specific time window (Chen et al., 2020b). However, the potential interactions between attributes were not established in the previous study. For the inclusion of such important relational information, we design the transformation of X_τ into its corresponding stack of temporal graphs $G_\tau = (G_1,\ G_2,\ G_{t-\tau+1} \dots, G_t)$ at each time interval as shown in Figure 3.

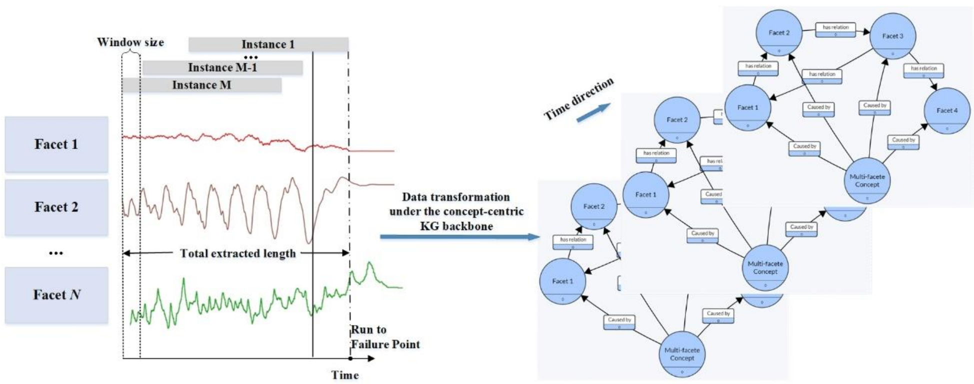

Figure 0.3 Transformation and temporal graph representation of the run-to-failure process data.

In this case, using the backbone structure of the CKG derived from the knowledge repository,

$$G^{\mathrm{t}} = (V^{\mathrm{t}}, E^{\mathrm{t}}, \boldsymbol{A}^{\mathrm{t}}) \tag{6.2}$$

with N nodes $v_i \in V^{\mathrm{t}}$ are vertices set following the time stamps and edges E^{t} are the edges set expressed as:

$$E^{\mathrm{t}} = \left\{ e_{jk}^{t} \middle| \forall \mathrm{j}, \mathrm{k} \in V^{\mathrm{t}} \right\} \in \mathbb{R}^{N \times N} \tag{6.3}$$

where $e_{jk}^{t} = 1$ if j, k are connected, when $e_{jk}^{t} = 0$, then j, k are disconnected. $\boldsymbol{A}^{\mathrm{t}} \in \mathbb{R}^{N \times N}$ is the adjacency matrix derived from the nodes:

$$\boldsymbol{A}^{\mathrm{t}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{6.4}$$

It is important to understand that the relationships between different concept-centric facets in failure diagnostic tasks are not explicitly provided in comparison with the task in traffic forecasting, where the adjacency matrix can be computed by using Euclidean distance among stations in a traffic network. The Euclidean distance is not the most appropriate metric to choose for modelling relationships between two measurements when they are in the Euclidean domain since proximity does not necessarily imply close relationships. In this way, we construct a weighted adjacency matrix between the different measurements based on the PCCs between them. The formulation of PCCs for sequence X_1 nd X_2 is:

$$P_{X_1, X_2} = \frac{\sum_{i=1}^{n} ({X_1}_i - \overline{X_1})({X_2}_i - \overline{X_2})}{\sqrt{\sum_{i=1}^{n} ({X_1}_i - \overline{X_1})^2} \sqrt{\sum_{i=1}^{n} ({X_2}_i - \overline{X_2})^2}} \tag{6.5}$$

In a nutshell, P_{X_1, X_2} is a number between -1 and 1 that indicates the extent to which two variables are related. In order to calculate the weighted adjacency matrix, the following formula is employed:

$$\mathbf{A}_{\boldsymbol{i},\boldsymbol{j}} = e^{P_{X_1, X_2}} \tag{6.6}$$

In this case, $\boldsymbol{A}_{\boldsymbol{i},\boldsymbol{j}}$ can be used to calculate the relationship between $\boldsymbol{A}^{\mathrm{t}}$ at different time intervals. The following GCN model takes $\boldsymbol{A}^{\mathrm{t}}$ and V^{t} as the model input.

6.2.3 KG-supported Multi-faceted Modelling

With the aim of mining implicit knowledge from a graph, triples in the knowledge graph are transformed into their corresponding low-dimensional vector embedding using either a knowledge representation learning model or a graph convolution model.

Firstly, an adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a square matrix that represents a finite graph. During the construction of the adjacency matrix, elements represent whether pairs of vertices in the graph are adjacencies or not. The degree matrix is $D_{ii} = \sum_i \boldsymbol{A}_{ij}$. When the multi-source data has been imported into the knowledge graph, information (represented by V and E) has been contained. By populating multi-sourced data into knowledge graphs, an embedding approach is necessary in order to transform the data from these graphs into information that can be used for multi-source conceptual modelling. In this study, as a convenient way to accomplish the embedding process, GCN is used to extract the connected features in an end-to-end manner. In other words, GCN updates each node respectively to their neighbourhoods.

Specifically, for the purpose of performing the temporal graph convolution, the GCN model takes $\boldsymbol{A}^{\mathrm{t}}$ and V^{t} as the input with the output feature $\overline{V_{\tau}} \in R^{\tau * \mathrm{N} * \mathrm{N}}$. The core theory of GCN is demonstrated as follows: Given a specific graph-based neural network model $f(X, A)$, here is a Layer-wise propagation rule for a multilayer GCN:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{6.7}$$

Here, as there are no self-connections in the graph, $\widetilde{\boldsymbol{A}} = \boldsymbol{A} + I_N$ is defined as the adjacency matrix of the graph, among which I_N is the identity matrix. Moreover, a layer-specific trainable weight matrix is described as $\widetilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ and the activation function is denoted as $\sigma(\cdot)$. The activations matrix is defined as $H^{(l)} \in \mathbb{R}^{N\times D}$ in the l^{th} layer. X is the original node attribute matrix where $H^{(0)} = X$.

A CKG enables the merging and organisation of time series data knowledge in order to predict faults in subsequent devices. Multiple sensors generate and collect data during the manufacturing process. With the transformation of multivariate time-series data into a stack of temporal graphs, the output feature $\overline{V}_{\tau}$can be fed into the TCN model for further concept modelling. The Temporal Convolutional Network (TCN), a method of processing time-series data, utilises dilated causal convolution and residual connections in order to address the problems discussed above. Dilated causal convolutions are used only for elements that precede the current element, while CNN performs convolution on elements adjacent to the current element. A hierarchy of temporal convolutional filters was first developed for the purpose of examining long-range patterns using the TCN approach (Lea et al., 2017). In TCNs, there are two main characteristics: (1) convolutions are causal, and (2) the network can map a sequence of any length to an output sequence of the same length, similar to RNNs. A generic convolutional architecture for sequential data is the basis of the proposed architecture (Bai et al., 2018). Through autoregressive prediction and a long memory, the architecture is simple (e.g., no skip connections across layers, as shown in Figure 4). In addition, it is capable of achieving very deep networks because it uses dilated convolutions that allow the receptive field to be exponentially expanded.

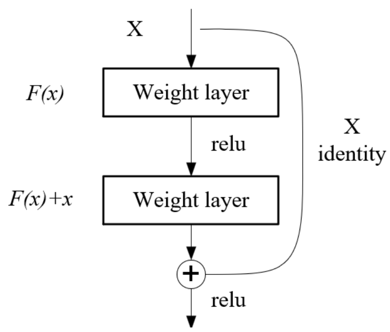

Figure 0.4 Illustration of skip connections across layers

As shown in Figure 6.5, there are three configurations of dilation factors d: 1, 2 and 4. Each subsequent filter tap is separated by a fixed dilation. As the dilations and filter size *k* increase, the receptive field is effectively expanded. As a result, each input will be filtered in some way (Bai et al., 2018).

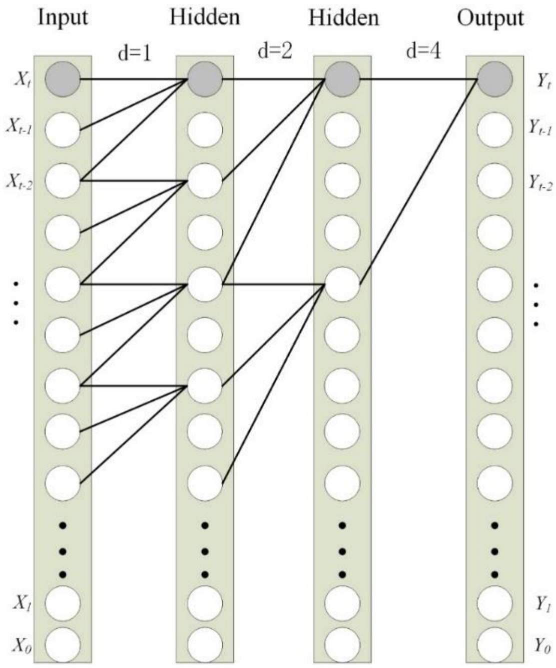

Figure 0.5 Illustration of TCN structure

Pre-activation residual connection scheme is applied to the dilated causal convolution subnetwork, which means that the BN and activation function is placed prior to the dilated causal convolution operations.

Table 0.1 Pseudo-code of the proposed approach

Pseudo-code of the proposed procedure (Jiang et al., 2022)
Input: $X_\tau \in R^{t*N}$ is the manufacturing process data, $PCCs \leftarrow X_\tau$ are the correlations among t sequences, $\mathrm{G} = (\mathrm{V}, E, \mathrm{A})$ is the CKG backbone **Output:** CKG-TCN model 1: $X_\tau = (X_1, X_2, X_{t-\tau+1} \dots, X_t) \in R^{t*N}$ 2: $\mathrm{G} = (\mathrm{V}, E, \mathrm{A})$ 3: **while** $e < E$ **do** 4: **for each** X_i in X_τ , $i \in t$, Transform X_i to Temporal graph G_i under G add G_i to a stack of graph G_τ 5: Compute adjacent matrixes $\boldsymbol{A}^\tau$ from G_τ 6: Calculate attribute interactions $PCCs \leftarrow X_\tau$ 7: Calculate the aggregated feature V_τ 8: Concept modelling using the TCN 9: Update the model parameters 10: **if** model convergence, **then** 11: Return the CKG-TCN 12: **end if** 13: **end while**

TCNs take advantage of the advantages of CNNs and RNNs. Since the RNN-based structure has several defects, the dilated causal convolution has been selected [25] to extract temporal dynamics. Long-term dependency can be learned by dilated causal convolution in a non-recursive manner, which results in a greater receptive field without significantly increasing computational cost.

6.3 Experimental Setup

To verify the effectiveness of the proposed approach, a multi-faceted concept with both temporal characteristics and internal correlations across the surrounding facets is an ideal modelling target. In this regard, the modelling task of strip breakage, a miscellaneous production failure in cold rolling, is taken as an illustrative study. First of all, there is a long history of research conducted regarding this failure, which means there are sufficient and reliable knowledge repositories on the side of this failure concept. Secondly, it has been verified the triggers of strip breakage are multi-faceted and various (Chen et al., 2021b), which drives the urgency for integrating multi-sourced data accordingly. Furthermore, since steelmaking is a sequential process, the steelmaking production line is typically compact and strongly correlated, which indicates the necessity of considering semantic relationship complexity across multiple sources.

It is a fact that cold rolling is one of the most important techniques used in the metal processing industry in order to produce sheets and strips due to its high efficiency and production rate (Hou et al., 2007). When it comes to cold rolling, it is inevitable that failures such as edge cracks, central bursts, surface defects, and buckles will take place (Mashayekhi et al., 2011). Strip breakages are among these failures which require special attention, as they result in significant increases in production costs and cycle times, as well as significant damage to mill accessories (Johnson and Mamalis, 1977). A retrospective analysis of root causes has been conducted in previous studies on strip breakage (Chen et al., 2021b), which has discussed the causes of strip breakage and classified it into four categories which are material, equipment malfunction, rolling operation, and work roll features.

It has been stated that strip breakages can take place due to a variety of factors. Therefore, it is imperative to examine the problem of strip breakage from multiple perspectives, including the analysis of feedstock properties, the examination of equipment malfunctions, the analysis of improper rolling process operation, and other factors. In this context, no single data source can capture the variety of breakage-centric factors that contribute to this production failure. Hence, it is necessary to merge data from multiple sources for the generation of collective information on strip breakage modelling using a data-driven approach. Also, owing to the wealth of domain knowledge regarding strip breakage and its causes, it is advantageous to integrate data from various sources with the utilisation of such knowledge.

Data for this study was provided by a steel manufacturer that manufactured electrical steel and used a reversing mill for cold rolling. In this material, this element increases its electrical resistivity, reducing magnetic losses. During cold rolling, the strip becomes brittle due to a higher concentration of silicon, resulting in more breakages (Takami et al., 2011a). The experiments are conducted on a 64-bit Windows server with 32 GB RAM and one Core i7-9700 K CPU as well as an NVIDIA GeForce 2080ti GPU for training time decrease. Python was the utilised platform for experiment implementation, and Pytorch was applied to build deep learning models (Paszke et al., 2017).

6.3.1 Data and Knowledge Repository Description

A production data acquisition (PDA) system was installed on the production site for the purpose of collecting raw data regarding the cold rolling process in this study. With the aid of this automated system, equipped with accurate measurement devices, variables related to cold rolling can be measured, including speed, tension, eccentricity, and roll gap position. Continuous monitoring and recording of data are carried out in real-time at a frequency of 100 *Hz* in order to document the continuous condition of the mill. In comparison with raw data, a lower sampling rate leads to distortion as a result of the high correlation between neighbouring data points. Due to this, our analysis was conducted using full-resolution data collected at a sampling rate of 100 *Hz* to obtain the most information possible from the PDA-recorded data. Furthermore, it is possible to calculate the breakage point in detail using full-resolution data, resulting in more accurate classification labels. It should also be noted that each selected coil only broke once. Thus, the dataset contained 1256 coils, among which 354 are broken strip coils, covering three months of production.

In most cases, it is not easy to construct the backbone of KG under a domain environment without the collaboration of domain experts (Jin, 2018). In the domain-centric knowledge representation, each triplet is described using the RDF language utilising open-source platforms. Therefore, these platforms are responsible for building and storing domain-centric knowledge bases. The strip-breakage-centric KG was constructed based on the refinement of the hierarchy structure and the completion of the relationships. A summary of related studies on strip breakage and the cause analysis can be categorised into four different facets, namely material-related issues, equipment malfunctions, rolling operations, and the rolls pushing the strips. Following the approaches stated in Section 3.1, CKG was constructed according, and the free software Gephi was used to visualise the CKG, as shown in Figure 6.6. Specifically, 2295 triplets are integrated and imported into *Gephi* to construct and visualise the strip-breakage KG in the cold rolling process of the steel industry. In Figure 6.6, an entire strip-breakage-centric KG composed of seven subclasses is present. The entire strip-breakage KG contained 230 entities and 2295 relations in total.

The coloured nodes in the graph represent different facets that categorize the entities in the CKG backbone, and the edges between nodes indicate the relationships among them. One key observation from the graph is that there is a high degree of interconnectedness among entities across different facets, suggesting that the proposed approach is able to capture the interdependence and complexity of real-world knowledge domains. Moreover, the different colours in the legend correspond to different facets such as process, material, equipment, and so on. The fact that entities are distributed across multiple facets and are interconnected within and between facets indicates that the proposed approach can effectively capture the multi-faceted nature of industrial processes.

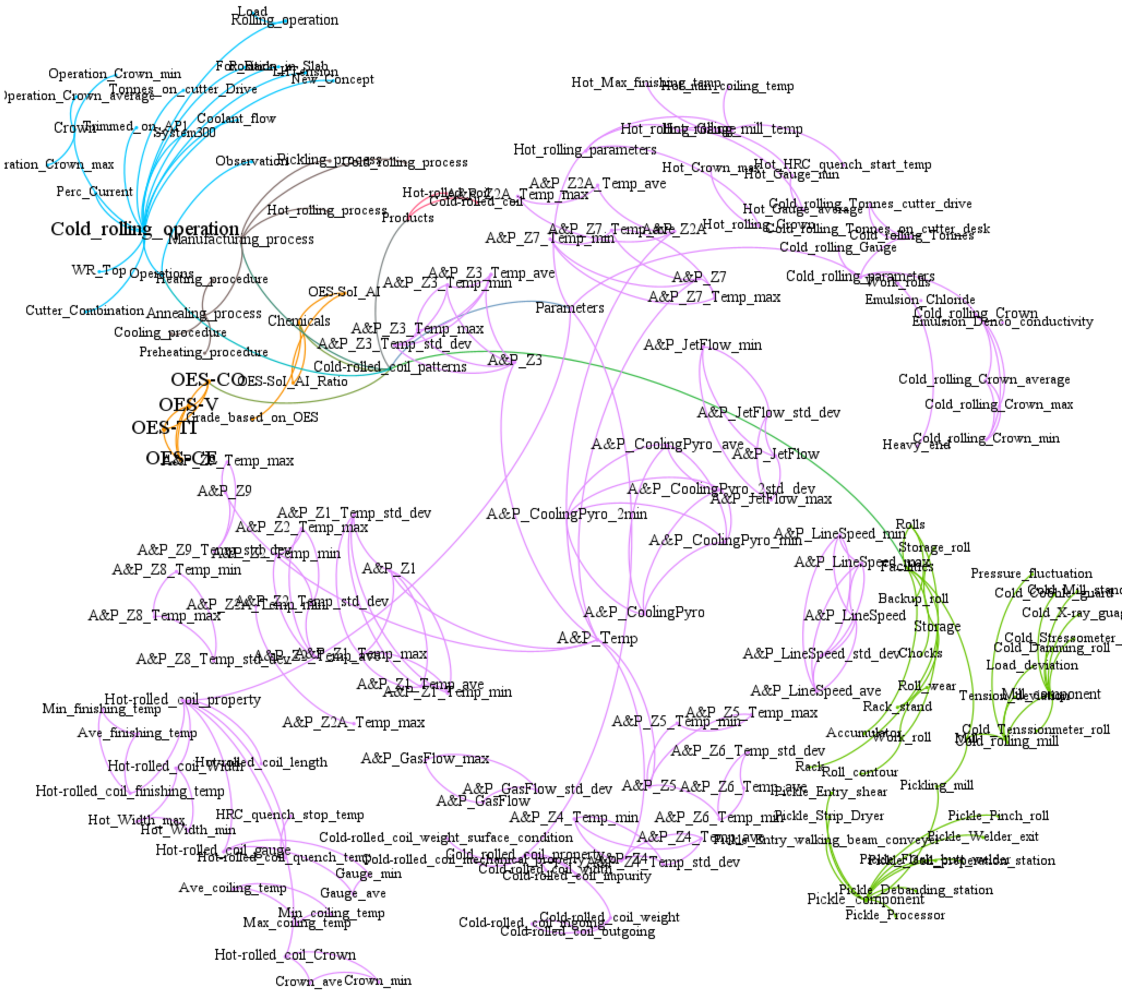

Figure 0.6 The overall breakage-centric KG

6.3.2 Model Setup

The experiments were set up in two scenarios: firstly, to verify the merits of the prediction performance, the modelling results of the CKG-TCN approach are compared with other prevailing machine learning algorithms. Under this scenario, six machine learning algorithms, which are Random Forest (RF), Support Vector Machine (SVM), *k*-Means, Long Short-term Memory (LSTM) network, Gated Recurrent Unit (GRU) and ConvLSTM, are compared with the proposed CKG-TCN.

To be specific, in this scenario, we have set the number of estimators in the RFs to 10. In SVMs, Gaussian kernels are employed as the kernel function. For *k*-Means, the cluster number is set to 3. LSTM and GRU have hidden state *ht* and window sizes of 5 and 16, respectively. By default, the kernel size of the ConvLSTM is set to 20. In addition, within the graph-aided approaches, we set two different baselines: we compare our method CKG-TCN with its non-GCN version CKG-TCN-noGCN.

Secondly, to verify the merits of this graph-supported modelling approach, two different fusion strategies were conducted: one is the transformation of multi-sourced sequential data into graph format using temporal graph convolution, and the other strategy does not conduct the transformation (inputs are the numerical values). Since the output of temporal graph convolution is a 3D tensor involving both attribute interactions and temporal dependency, conventional ML algorithms such as RFs are not suitable for this scenario. In this case, LSTM, GRU, ConvLSTM and TCN are selected for the experiments in two different fusion strategies. The parameters are the same as in scenario one.

6.3.3 Evaluation Metrics

To evaluate the modelling performance, the metrics below are used in this experiment.

$$Accuracy = \frac{|f\_set \cap K|}{|K|} \tag{6.8}$$

$$Precision = \frac{|f\_set \cap K|}{|K|} \tag{6.9}$$

$$Recall = \frac{|f\_set \cap K|}{|f\_set|} \tag{6.10}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{6.11}$$

A precision rate can be defined as the proportion of positive samples predicted from the predicted results. Recall, which is the percentage of faults correctly predicted in a test set, is a measure of how many faults appear in the test set. F1 is the analysis of both precision and recall.

6.4 Experimental Results

6.4.1 Exploration Experiments on Window Size

It is important to consider the window size when modelling time-series data because it has a substantial influence on performance. Therefore, in order to evaluate the effect of this parameter, we assess the performance of the TCN (without converting the time-series data into temporal CKG) modelling with the intention of examining its temporal trends. As the frequency is 100 Hz, the default window size is set to 0.01 seconds.

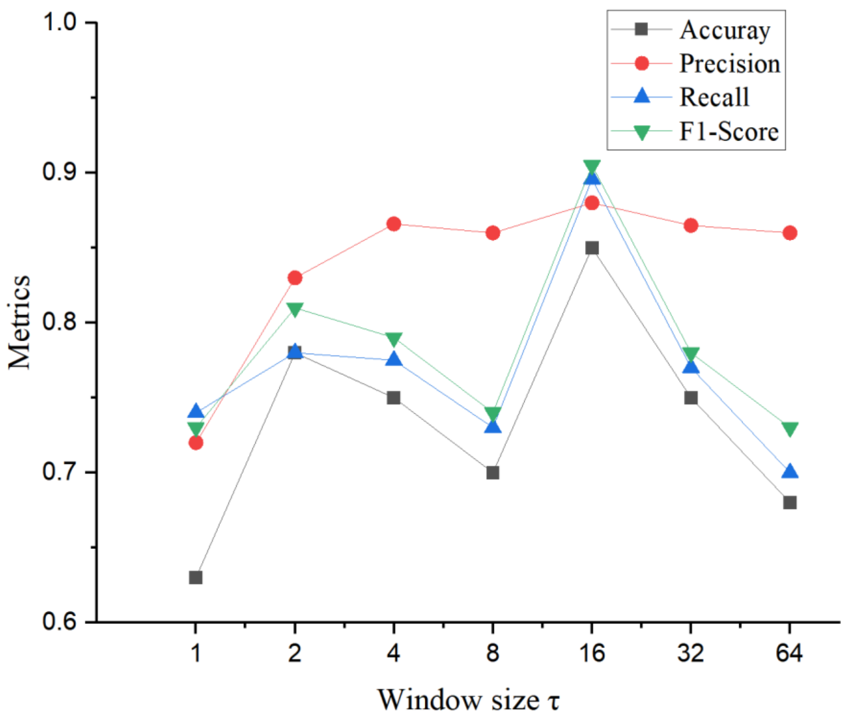

Figure 0.7 Evaluating the performance of the TCN at different window sizes τ

Figure 7 illustrates a similar trend in all the metrics. It is obvious to see the best results are achieved when τ = 16 in terms of all different metrics. The precision reaches the highest when τ = 16. Accordingly, τ = 16 may represent the most relevant granularity of process data for fault modelling, and this window size is chosen as the most suitable parameter for the following experiments.

6.4.2 Performance Comparison with Prevailing ML algorithms

As stated in Section 6.3.2, six algorithms which are K-Means, RFs, SVM, GRU, LSTM, and ConvLSTM, are compared with the CKG-TCN. By performing 10-fold cross-validation, accuracy, precision, recall, and F1-score are obtained.

Table 0.2 Comparison with other prevailing machine learning algorithms.

Competitors		Accuracy %	Precision %	Recall %	F1
Conventional	*k*-Means	52.25	72.32	43.32	0.42
	SVM	56.59	74.15	64.17	0.67
	RF	71.67	79.15	80.92	0.77
RNN/CNN-based	LSTM	75.33	86.29	81.37	0.82
	GRU	74.52	86.23	80.11	0.83
	ConvLSTM	76.30	83.46	82.35	0.82
Graph-aied	CKG-TCN-noGCN	76.21	84.81	84.32	0.80
	CKG-TCN	**81.20**	**86.33**	**87.46**	**0.88**

In Table 6.2, 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.

Since RNNs take temporal information into account, they perform better than conventional SVMs, K-Means, and RFs. More specifically, RFs perform better than SVMs and K-Means because they aggregate weak classifiers into stronger classifiers. CKG-TCN, on the other hand, has the best performance. In terms of accuracy, precision, and F1-score, it performs better than SVM and K-Means. With respect to Accuracy and F1-Score, it is superior to RFs, GRU, TCN, and ConvLSTM.It is because CKG-TCN considers not only temporal dependency but also attribute interactions, so it performs better across all metrics because we consider attribute interactions as well as temporal dependency.

Meanwhile, within the graph-aided approaches, there is a great difference regarding the performance with respect to Accuracy, Precision, and F1-Score. We can see non-GCN method performs significantly worse on all different classification metrics. This illustrates how GCN can be used to learn the embedding vectors of various causes of breakage within a graph. Possibly, this is because the GCN layer fits the feature representations derived by the model. As GCN is included, the node is able to aggregate more information. To obtain a feature representation with rich information, the CKG-TCN combines the features learned by each GCN layer and the features of the node itself. As a result, the feature representations obtained by CKG-TCN may be more accurate and contain less noise than those obtained by CKG-TCN without GCN.

6.4.3 Performance Comparison with Prevailing ML algorithms

In this section, the comparison results between graph-aided fusion methods and the conventional fusion methods without the support of KG are introduced. Through this ablation experiment, the effectiveness of the KG-supported fusion strategy is verified. As stated in Section 4.4, LSTM, GRU, ConvLSTM and TCN are compared under two different fusion strategies. Strategy one: the sequential numerical data are transformed into CKG format, and the input of deep learning models are 3D-tensor derived from temporal graph convolution, which is $\overline{V_\tau} \in R^{\tau * \mathrm{N} * \mathrm{N}}$. Strategy two: the raw sequential data are numerically concatenated and then fed into the proceeding deep learning models. For the comparison of modelling performance, considering the imbalance of the dataset, Accuracy and Recall are used as the metrics under 10-fold cross-validation. The performance impact of KG-supported and nonKG-supported when conducting the fusion task is demonstrated in Figure 6.8 and Figure 6.9.

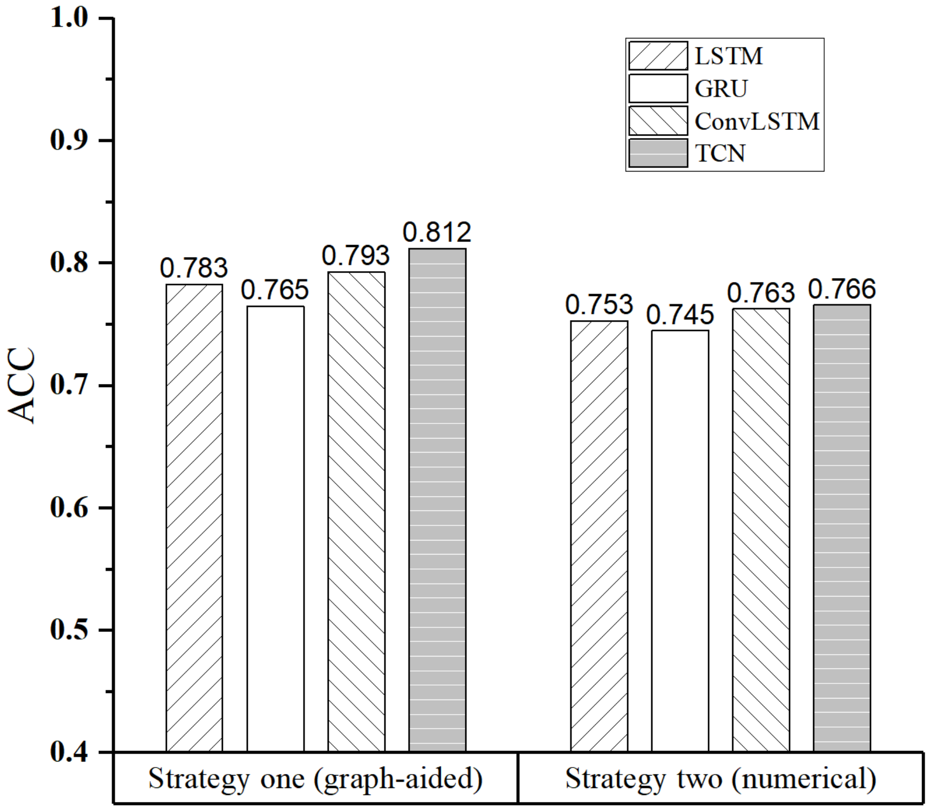

Figure 6.8 Modelling performance in terms of ACCs in different strategies

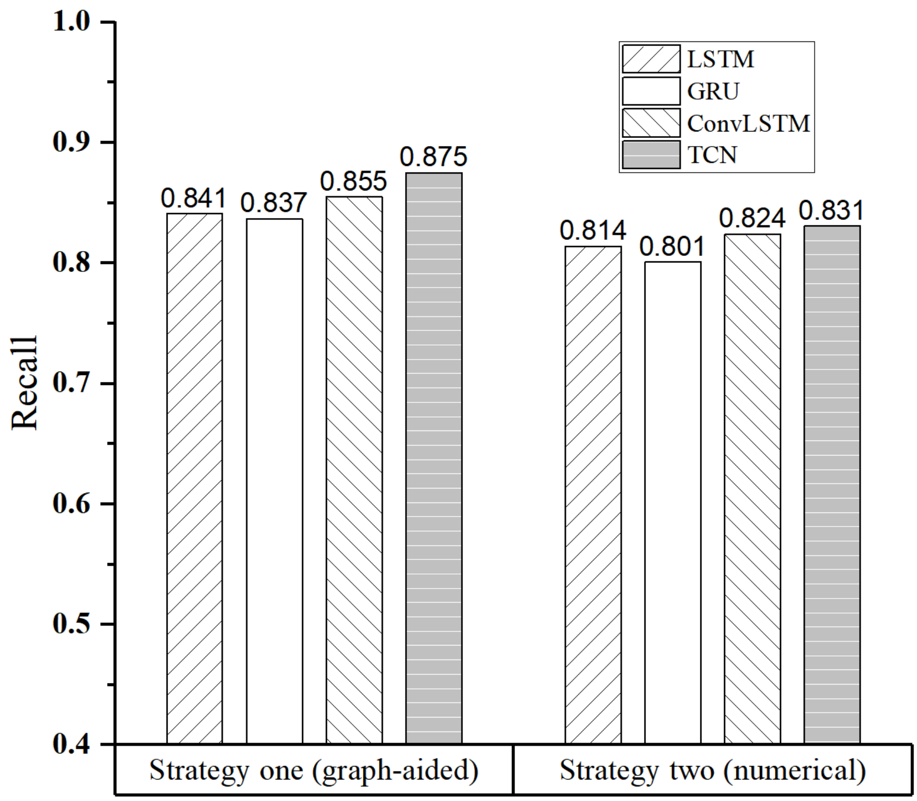

Figure 6.9 Modelling performance in terms of Recalls in different strategies

In Figure 6.8, we present the modelling results for various fusion strategies. As a general rule, graph-aided models in strategy one perform better than graph-aided models in strategy two in terms of both ACC and recall. As a result, the model we designed was able to provide a better fitting of the features as well as a more accurate representation of the potentially complex correlations between concept-centric features. As strip breaks occur instantly, the model performance may differ due to the fact that only a detailed representation is able to capture the momentary pattern before the strip breaks. Moreover, as a miscellaneous production failure, interactions and correlations among various attributes cannot be ignored. In contrast to conventional numerical fusion systems, the multi-source numerical fusion approach does not provide accurate associated knowledge regarding the complex semantic relationships between the data sources.

To be specific, within both the graph-aided and numerical strategy, the RNN-CNN-based algorithms (ConvLSTM and TCN) outperform the original RNN-based (GRU and LSTM) algorithms with respect to both accuracy and recall. It shows that the RNN-CNN-based model can not only capture the temporal characteristics of the process data but also extract rich information from such data through the convolution layer. Despite this, graph-aided strategies require more computational resources than conventional approaches due to the complexity of the models. Additionally, it is usually necessary to select hyperparameters for RNN-based models to achieve the desired performance.

6.5 Discussion

These experiments provide some key insights into the modelling results. The first important consideration is to determine the best window size for monitoring the most relevant data related to the working system's condition. A reliable model for failure prediction can be developed by utilising the best temporal dependency. To achieve higher performance levels, it is also crucial to establish attribute dependencies to ensure that the system works correctly.

Secondly, one significant disadvantage of SVMs, K-Means, and *k*-Means is their inability to depend on the working system. The status of the present is dependent on the status of the past. This renders a single health record insufficient to evaluate the current state of health. While RNNs are designed to extract temporal characteristics, they do not consider the internal relations among features. We can create a reliable model for concept modelling by fully utilising the temporal dependency to gain a better understanding. Meanwhile, compared with RNN-based models, the RNN-CNN-based model can not only capture the temporal characteristics of the process data but also extract rich information from such data through the convolution layer.

Thirdly, it can be concluded from the comparison experiments of graph-aided fusion and conventional data fusion that there are potential transfer relationships between the multi-sourced attributes. It is difficult to obtain accurate associated knowledge regarding the relationships between multiple data sources using numerical fusion approaches without semantic information mining across the concept-centric attributes. In this case, to achieve better performances, it is crucial to establish an approach with the inclusion of attribute dependencies across the concept-centric features. It may be that the graph-aided approach can obtain both temporal feature embedding and attributes relationship embedding at the same time.

Finally, since the GCN is used to fit the temporal and attributes relationship feature representations, the inclusion of graph features extraction by GCN is essential. A node can aggregate more information by using GCN extraction and aggregation without experiencing an excessive amount of noise. CKG-TCN provides a feature representation with rich information and less noise by combining the features learned by each layer of the GCN and the characteristics of the node. Thus, CKG-TCN-noGCN may result in more accurate feature representations containing less noise than CKG-TCN-noGCN.

6.6 Summary

The development of a knowledge graph-aided multi-faceted modelling method was proposed as a means of overcoming the limitations of conventional equipment fault diagnosis. With the construction of concept-centric KG, the multivariate time-series data was transformed into a temporal graph representation of the data sequence, and GRL techniques were applied to extract features from these temporal graphs, and these features were fed into a ML model for fault concept modelling. The experimental results show: (1) the KG-aided fusion strategy shows merits against the numerical fusion strategy since it considers intra-feature relationships; (2) The graph feature extraction using GCN provides a feature representation with rich information and less noise, which results in more accurate results. Methodologically, this approach improves the accuracy and convergence speed of fault diagnosis, enables constructing a domain map of equipment fault diagnosis, and combines mechanism knowledge and data-driven methods of multi-faceted conceptual modelling.

Chapter 7 Achievements and Conclusions

7.1 Achievements

In this study, the principal objective was to model the multi-faceted and instantaneous production failures of steel-making in order to optimize the quality control at modern steel-making facilities. The motivation for this study was discussed at the beginning of this thesis, and it was underpinned by a discussion of the background of data mining in the steel industry. While data mining has been given considerable attention in the steel industry, little research has focused on fault diagnosis utilizing multi-sourced data and fault-related knowledge. For the purpose of steel-making fault diagnosis, this study analysed multi-source data and fault-related knowledge. Historically, only numerical steel-making process data have been used in related research, and they are usually derived from only one source. As a result of industry 4.0, semantic data is becoming available, as well as techniques to handle it. Meanwhile, the emerging KG and GRL techniques enable knowledge processing and ML analysis on graphs. The first chapter of this thesis contains four research questions. Based on the work achieved, the answers to the research questions are obtained.

Subsequently, the literature review was provided with the relevant technologies and relevant research to determine the state-of-the-art research. Firstly, the steel-making process was examined through three main aspects, including the steel-making process, cold rolling process and strip breakage as two specific examples. Then, data mining and its applications in the steel-making industry were reviewed. Concisely, as the core part of the steel-making industry, the specific tasks and the applied techniques were involved in this section. Also, the relevant studies of KG were introduced concerning

KG construction and graph representation learning. Lastly, the studies on the strategies and KG-aided techniques under an information fusion context were investigated.

The first research question is: what is an appropriate data mining framework for fault diagnosis in steel-making with the full exploitation of these resources? Following the understanding of the state-of-the-art, a framework was designed for fault diagnosis based on the multi-sourced data and existing domain knowledge in steel-making. In this framework, the stage of data and knowledge collection, data and knowledge processing, and data transformation are specified. After the processing, GRL techniques are applied to extract graph features. Finally, these features are fed into ML pipelines for the modelling results, which can be used to leverage the decision support for quality control in steel-making.

With the aim of achieving fault modelling in a predictive manner with the justification of facets surrounding the production failure, a multi-faceted modelling approach was proposed. In this study, a typical product failure named strip breakage was selected. With the aim of minimising the occurrence and impact of strip breakage, a micro-level prediction of strip breakage based on historical process data was achieved. The 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 was conducted. The key findings in this study are: (1) Among the models built on the proposed LSTM network, the best performance was achieved when features from all three facets were analysed together; (2) In comparison with those prevailing machine learning algorithms, the LSTM network can achieve better algorithm performance under higher computational loads.

Following the determination of the fault modelling approach, an issue in the cold rolling process is the multitudinous factors contributing to this fault. Meanwhile, among the heterogeneous data surrounding multi-faceted concepts in steel-making, a significant amount of data consists of rich semantic information. With the consideration of better utilisation of these resources, the third research question was how to manage and exploit these resources in steel-making using KG considering their heterogeneity. In this context, a framework of domain-centric KG construction has been presented to avoid the massive participation of domain experts, as well as to refine KGs and discover the missing relations in this study. This proposed KG construction framework serves as a model that aims to design a reliable ontology and complete relations in constructing the domain-centric KG for information management and knowledge sharing in steel-making. The experimental results indicate: (1) the proposed framework of hierarchy structure refinement avoids the extensive knowledge alignment by domain experts; (2) compared with the most common machine learning models, the GraphSAGE classifier achieved the most performance regarding relation completion.

With the proposed semantic approach of constructing a fault-centric KG in steel-making, the final research question further explores the ability of KG to facilitate multi-faceted conceptual modelling based on heterogeneous and semantic-rich data. With the construction of concept-centric KG, the multivariate time-series data was transformed into a temporal graph representation of the data sequence, and GRL techniques were applied to extract features from these temporal graphs, and these features were fed into a ML model for fault concept modelling. The experimental results show: (1) the KG-aided fusion strategy shows merits against the numerical fusion strategy since it considers intra-feature relationships ; (2) The graph feature extraction using GCN provides a feature representation with rich information and less noise, which results in more accurate results.

7.2 Future Works

The aim of this thesis is to research data mining in the steel-making industry with a focus on fault diagnosis. For a severe production failure which occurs instantaneously, 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 production failure, we achieved a micro-level prediction of strip breakage based on historical process data. The 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 production failure in the near future can also be vital for post-analysis, such as in determining what countermeasures should be used. For further work of this work, 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 the material defect, which is another critical issue for strip breakage, were 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. For real cold rolling practice, even if we considered all the causes of strip breakage beforehand, the 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, in 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.

For future work regarding the construction of steel-making domain KG: firstly, 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. Meanwhile, 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, were 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. Future research work can focus on improving the performance of relation completion. In addition, it remains a challenge to develop and refine KGs automatically. Combined with graph theories, the representation learning of KG is also interesting. For future work on KG-aided modelling, as this study is conducted using a graph-based approach, it is more suitable for cases in which there are enough attributes and more interactions between the attributes. This may limit the adaptivity of the proposed approach. Furthermore, it would be beneficial to explore a more precise adjacency matrix and a more effective spatial-temporal structure.

7.3 Conclusions

In conclusion, this study aimed to model multi-faceted and instantaneous production failures of steel-making in order to optimize quality control in modern steel-making facilities. The thesis proposed several frameworks, including a fault diagnosis framework based on multi-sourced data and existing domain knowledge in steel-making, a multi-faceted modelling approach to predict a typical steel production failure, a domain-centric knowledge graph construction framework, and a KG-based approach to facilitate information fusion for multi-faceted steel-making concept modelling.

The proposed frameworks offer several capabilities, such as utilizing semantic data and knowledge processing techniques to enhance fault diagnosis and modelling, enabling the integration of heterogeneous data to construct a domain-centric knowledge graph, and facilitating the fusion of information for multi-faceted concept modelling. However, the frameworks have certain limitations and constraints. For instance, the proposed fault diagnosis framework may not be able to address certain production failures that do not have a divinable manifestation in rolling process variables. Moreover, the construction of the domain-centric knowledge graph may require domain expert assistance to refine and discover missing relations. Finally, the

KG-based approach for multi-faceted concept modelling may have limitations in cases with fewer attributes and interactions between the attributes.

Furthermore, this study has demonstrated the application of the proposed frameworks in the industrial case study of steel-making. The results show the potential of these frameworks to improve quality control and contingency mitigation in the steel-making industry. The multi-faceted modelling approach has been applied to predict the occurrence of strip breakage, a common production failure in steel-making. The proposed approach has achieved a micro-level prediction of strip breakage based on historical process data, which can assist mill operators in developing contingency mitigation strategies and post-analysis. Similarly, the proposed domain-centric KG construction framework has been applied to construct a strip-breakage-centric KG of the cold rolling process. The results show the potential of this framework to refine KGs and discover missing relations. Finally, the proposed KG-aided modelling approach has been applied to facilitate multi-faceted conceptual modelling based on heterogeneous and semantic-rich data in the industrial case study of steel-making. The results show the potential of this approach to provide a feature representation with rich information and less noise, resulting in more accurate results. To sum up, this study has contributed to the field of data-driven fault diagnosis and quality control in the steel-making industry by proposing and demonstrating the potential of multi-faceted modelling, domain-centric KG construction, and KG-aided modelling approaches. While the proposed frameworks have limitations and constraints, they provide promising avenues for future research and practical applications in the steel-making industry.

References

ABU-SALIH, B. 2021. Domain-specific knowledge graphs: A survey. Journal of Network and Computer Applications, 185, 103076.

ABU-SALIH, B., AL-TAWIL, M., ALJARAH, I., FARIS, H., WONGTHONGTHAM, P., CHAN, K. Y. & BEHESHTI, A. 2021. Relational learning analysis of social politics using knowledge graph embedding. Data Mining and Knowledge Discovery, 35, 1497-1536.

ALTıNKAYA, H., ORAK, I. M. & ESEN, I. 2014. Artificial neural network application for modeling the rail rolling process. Expert Systems with Applications, 41, 7135-7146.

BAI, S., KOLTER, J. Z. & KOLTUN, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

BATAL, I., FRADKIN, D., HARRISON, J., MOERCHEN, F. & HAUSKRECHT, M. Mining recent temporal patterns for event detection in multivariate time series data. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012. ACM, 280-288.

BAUERNHANSL, T., TEN HOMPEL, M. & VOGEL-HEUSER, B. 2014. Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung· Technologien· Migration, Springer.

BELKIN, M. & NIYOGI, P. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation, 15, 1373-1396.

BENGIO, Y., COURVILLE, A. & VINCENT, P. 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35, 1798-1828.

BHATTACHARYA, D., MISHRA, A., PODDAR, G. P. & MISRA, S. 2016. 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, 5, 15-22.

BREIMAN, L. 2001. Random forests. Machine learning, 45, 5-32.

BRONSTEIN, M. M., BRUNA, J., LECUN, Y., SZLAM, A. & VANDERGHEYNST, P. 2017. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine, 34, 18-42.

CAI-YAN, L. & YOU-FA, S. Application of data mining in production quality management. 2009 Third International Symposium on Intelligent Information Technology Application, 2009. IEEE, 284-287.

CAO, Q., BEDEN, S. & BECKMANN, A. 2022. A core reference ontology for steelmaking process knowledge modelling and information management. Computers in Industry, 135, 103574.

CAO, S., LU, W. & XU, Q. Grarep: Learning graph representations with global structural information. Proceedings of the 24th ACM international on conference on information and knowledge management, 2015. 891-900.

CARDOSO, S. D., DA SILVEIRA, M. & PRUSKI, C. 2020. Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies. Knowledge-Based Systems, 194, 105508.

CEMERNEK, D., CEMERNEK, S., GURSCH, H., PANDESHWAR, A., LEITNER, T., BERGER, M., KLöSCH, G. & KERN, R. 2022. Machine learning in continuous casting of steel: A state-of-the-art survey. Journal of Intelligent Manufacturing, 33, 1561-1579.

CEMERNEK, D., GURSCH, H. & KERN, R. Big data as a promoter of industry 4.0: Lessons of the semiconductor industry. 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), 2017. IEEE, 239-244.

CHAMI, I., ABU-EL-HAIJA, S., PEROZZI, B., Ré, C. & MURPHY, K. 2020. Machine learning on graphs: A model and comprehensive taxonomy. arXiv preprint arXiv:2005.03675.

CHEN, C., JAFARI, R. & KEHTARNAVAZ, N. 2017. A survey of depth and inertial sensor fusion for human action recognition. Multimedia Tools and Applications, 76, 4405-4425.

CHEN, F., WANG, Y.-C., WANG, B. & KUO, C.-C. J. 2020a. Graph representation learning: a survey. APSIPA Transactions on Signal and Information Processing, 9.

CHEN, H., CHEN, L., KUANG, X., XU, A. & YANG, Y. A Safe and Intelligent Knowledge Graph Construction Model Suitable for Smart Gridaper Title. 2021 6th International Conference on Image, Vision and Computing (ICIVC), 2021a. IEEE, 253-258.

CHEN, Z., LIU, Y., VALERA-MEDINA, A. & ROBINSON, F. 2019. Characterizing Strip Snap in Cold Rolling Process Using Advanced Data Analytics. Procedia CIRP, 81, 453-458.

CHEN, Z., LIU, Y., VALERA-MEDINA, A., ROBINSON, F. & PACKIANATHER, M. 2020b. Multi-faceted modelling for strip breakage in cold rolling using machine learning. International Journal of Production Research, 1-14.

CHEN, Z., LIU, Y., VALERA-MEDINA, A., ROBINSON, F. & PACKIANATHER, M. 2021b. Multi-faceted modelling for strip breakage in cold rolling using machine learning. International Journal of Production Research, 59, 6347-6360.

CHEN, Z., LIU, Y., VALERA MEDINA, A. & ROBINSON, F. 2020c. A multi-source feature-level fusion approach for predicting strip breakage in cold rolling.

COCCHI, M. 2019a. Data Fusion Methodology and Applications, Elsevier.

COCCHI, M. 2019b. Introduction: Ways and means to deal with data from multiple sources. Data Handling in Science and Technology. Elsevier.

CORCHO, O., PRIYATNA, F. & CHAVES-FRAGA, D. 2020. Towards a new generation of ontology based data access. Semantic Web, 11, 153-160.

CORTES, C. & VAPNIK, V. 1995. Support-vector networks. Machine learning, 20, 273-297.

CSER, L., KORHONEN, A., GULYAS, J., MANTYLA, P., SIMULA, O., REISS, G. & RUHA, P. Data mining and state monitoring in hot rolling. Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials. IPMM'99 (Cat. No. 99EX296), 1999. IEEE, 529-536.

CUI, X. & ZHAO, A. 2013a. Analysis the Causes of Deviation Breaking of Cold Rolling Strip. Xinjiang Steel, 1, 44-47.

CUI, X. & ZHAO, A. 2013b. Analysis the Causes of Deviation Breaking of Cold Rolling Strip. Xinjiang Steel, 44-47.

DA CUNHA, C., AGARD, B. & KUSIAK, A. 2006. Data mining for improvement of product quality. International journal of production research, 44, 4027-4041.

DOGAN, A. & BIRANT, D. 2021. Machine learning and data mining in manufacturing. Expert Systems with Applications, 166, 114060.

DOU, J., QIN, J., JIN, Z. & LI, Z. 2018. Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage. Journal of Visual Languages & Computing, 48, 19-28.

FALEIRO, R. M. R., VELLOSO, C. M., DE CASTRO, L. F. A. & SAMPAIO, R. S. 2013. Statistical modeling of charcoal consumption of blast furnaces based on historical data. Journal of Materials Research and Technology, 2, 303-307.

FAN, Z. & FRIEDMANN, S. J. 2021. Low-carbon production of iron and steel: Technology options, economic assessment, and policy. Joule, 5, 829-862.

FANG, W., MA, L., LOVE, P. E., LUO, H., DING, L. & ZHOU, A. 2020. Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. Automation in Construction, 119, 103310.

FAWCETT, T. 2006. An introduction to ROC analysis. Pattern recognition letters, 27, 861-874.

FOURNIER-VIGER, P., LIN, J. C.-W., GOMARIZ, A., GUENICHE, T., SOLTANI, A., DENG, Z. & LAM, H. T. The SPMF open-source data mining library version 2. Joint European conference on machine learning and knowledge discovery in databases, 2016. Springer, 36-40.

GAMBOA, J. C. B. 2017. Deep learning for time-series analysis. arXiv preprint arXiv:1701.01887.

GANESAN, B., PARKALA, S., SINGH, N. R., BHATIA, S., MISHRA, G., PASHA, M. A., PATEL, H. & NAGANNA, S. 2020. Link prediction using graph neural networks for master data management. arXiv preprint arXiv:2003.04732.

GéRARD, M., HASTINGS, M. & VERVOORT, P. 2007. Vibration monitoring of tandem mills at the Arcelor Mittal Tilleur cold rolling mill. Steel Times International, 31, 24.

GETOOR, L. & MACHANAVAJJHALA, A. 2012. Entity resolution: theory, practice & open challenges. Proceedings of the VLDB Endowment, 5, 2018-2019.

GRAVES, A. 2012. Supervised sequence labelling. Supervised sequence labelling with recurrent neural networks. Springer.

GRAVINA, R., ALINIA, P., GHASEMZADEH, H. & FORTINO, G. 2017. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. Information Fusion, 35, 68-80.

GROVER, A. & LESKOVEC, J. node2vec: Scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016. 855-864.

GUTHRIE, D., ALLISON, B., LIU, W., GUTHRIE, L. & WILKS, Y. A closer look at skip-gram modelling. Proceedings of the fifth international conference on language resources and evaluation (LREC’06), 2006.

HA, J., CHO, J., LEE, B. & HA, M. 2001. Numerical analysis of secondary cooling and bulging in the continuous casting of slabs. Journal of Materials Processing Technology, 113, 257-261.

HALL, D. L. & LLINAS, J. 1997. An introduction to multisensor data fusion. Proceedings of the IEEE, 85, 6-23.

HAMILTON, W., YING, Z. & LESKOVEC, J. 2017a. Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

HAMILTON, W. L., YING, R. & LESKOVEC, J. 2017b. Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584.

HAN, J., PEI, J. & TONG, H. 2022. Data mining: concepts and techniques, Morgan kaufmann.

HAN, Y., JUN, G. & TENG, L. 2018. Causes and Control Measures for Strip Break of Cold Rolling Mill in Baotou Steel. Science and Technology of Baotou Steel, 44, 25-28.

HAN, Z., ZHAO, J., WANG, W. & LIU, Y. 2016. A two-stage method for predicting and scheduling energy in an oxygen/nitrogen system of the steel industry. Control Engineering Practice, 52, 35-45.

HARANDI, M. T., SANDERSON, C., SHIRAZI, S. & LOVELL, B. C. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. CVPR 2011, 2011. IEEE, 2705-2712.

HOCHREITER, S. & SCHMIDHUBER, J. 1997. Long short-term memory. Neural computation, 9, 1735-1780.

HONGZHAO, H., YUE, L., JINZHOU, M. & JIANGHUA, S. 2010. Analysis and research on common problems and faults in cold rolling of strips. Heavy Machinery, 2.

HOU, F., ZHANG, J., CAO, J. & SHI, X. 2007. Review of chatter studies in cold rolling. Gangtie Yanjiu Xuebao(Journal of Iron and Steel Research), 19, 6-10.

HU, C., XIE, S., XIE, Y. & CHEN, X. Development of Domain Knowledge Graph: A Case Study on Flotation Process. 2021 6th International Conference on Robotics and Automation Engineering (ICRAE), 2021a. IEEE, 370-375.

HU, P.-H., ZHAO, H. & EHMANN, K. 2006. 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, 220, 1293-1303.

HU, Z. & HE, D. 2022. Operation scheduling optimization of gas–steam–power conversion systems in iron and steel enterprises. Applied Thermal Engineering, 206, 118121.

HU, Z., WEI, Z., SUN, H., YANG, J. & WEI, L. 2021b. Optimization of metal rolling control using soft computing approaches: a review. Archives of Computational Methods in Engineering, 28, 405-421.

ILIN, K. & BARANOV, G. Research on the influence of the strip tension on contact stress and cold rolling force with a strain-hardening. IOP Conference Series: Materials Science and Engineering, 2020. IOP Publishing, 012026.

ITO, K., SHIOYA, M., AGO, M., KUROKAWA, T., KOBAYASHI, H. & MORI, J. 2019. Production planning and scheduling technology for steel manufacturing process. Nippon Steel & Sumitomo Metal Technical report, 121, 31-47.

IWADOH, S. & MORI, T. 1992. Effect of work roll materials and progress of manufacturing technology on cold rolling and future developments in Japan. ISIJ international, 32, 1131-1140.

IWASAKI, M. & MATSUO, M. 2011. Change and development of steel-making technology. nippon steel technical report, 391, 88-93.

JABBAR, S., MALIK, K. R., AHMAD, M., ALDABBAS, O., ASIF, M., KHALID, S., HAN, K. & AHMED, S. H. 2018. A methodology of real-time data fusion for localized big data analytics. IEEE Access, 6, 24510-24520.

JI, S., PAN, S., CAMBRIA, E., MARTTINEN, P. & PHILIP, S. Y. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

JIANG, Y., DAI, P., FANG, P., ZHONG, R. Y. & CAO, X. 2022. Electrical-STGCN: An Electrical Spatio-Temporal Graph Convolutional Network for Intelligent Predictive Maintenance. IEEE Transactions on Industrial Informatics.

JIN, F., ZHAO, J., LIU, Y., WANG, W. & XU, Q. 2021. A scheduling approach with uncertainties in generation and consumption for converter gas system in steel industry. Information Sciences, 546, 312-328.

JIN, Z. 2018. Environment modeling-based requirements engineering for software intensive systems, Morgan Kaufmann.

JOHNSON, W. & MAMALIS, A. A survey of some physical defects arising in metal working processes. Proceedings of the Seventeenth International Machine Tool Design and Research Conference, 1977. Springer, 607-621.

JOLLIFFE, I. T. & CADIMA, J. 2016. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374, 20150202.

JUN, S., LEE, S. & CHUN, H. 2019. Learning dispatching rules using random forest in flexible job shop scheduling problems. International Journal of Production Research, 57, 3290-3310.

KANO, M. & NAKAGAWA, Y. 2008. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. Computers & Chemical Engineering, 32, 12-24.

KARIM, M. R., COCHEZ, M., JARES, J. B., UDDIN, M., BEYAN, O. & DECKER, S. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network.

Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019. 113-123.

KATONA, A., SPICK, R., HODGE, V., DEMEDIUK, S., BLOCK, F., DRACHEN, A. & WALKER, J. A. 2019. Time to Die: Death Prediction in Dota 2 using Deep Learning. arXiv preprint arXiv:1906.03939.

KEJRIWAL, M. 2019. Domain-specific knowledge graph construction, Springer.

KINGMA, D. P. & BA, J. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

KIPF, T. N. & WELLING, M. 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

KITCHIN, R. 2014. Big Data, new epistemologies and paradigm shifts. Big data & society, 1, 2053951714528481.

KONG, L., PENG, X., CHEN, Y., WANG, P. & XU, M. 2020. Multi-sensor measurement and data fusion technology for manufacturing process monitoring: a literature review. International Journal of Extreme Manufacturing, 2, 022001.

KUMAR, F., SHARMA, A. & ORAON, M. 2019. Future Research Potentials of Hot Rolling Process: A Review. Innovation in Materials Science and Engineering, 27-33.

KUMAR, S., MALLIK, A., KHETARPAL, A. & PANDA, B. 2022. Influence maximization in social networks using graph embedding and graph neural network. Information Sciences, 607, 1617-1636.

KUO, Y.-H. & KUSIAK, A. 2019. From data to big data in production research: the past and future trends. International Journal of Production Research, 57, 4828-4853.

LEA, C., FLYNN, M. D., VIDAL, R., REITER, A. & HAGER, G. D. Temporal convolutional networks for action segmentation and detection. proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 156-165.

LEE, S. Y., TAMA, B. A., CHOI, C., HWANG, J.-Y., BANG, J. & LEE, S. 2020. Spatial and sequential deep learning approach for predicting temperature distribution in a steel-making continuous casting process. IEEE Access, 8, 21953-21965.

LI, F., WU, J., DONG, F., LIN, J., SUN, G., CHEN, H. & SHEN, J. Ensemble machine learning systems for the estimation of steel quality control. 2018 IEEE International Conference on Big Data (Big Data), 2018a. IEEE, 2245-2252.

LI, H., ZHAO, Z., DONG, D., HAN, G., ZHANG, J., LIU, H. & YOU, X. 2018b. Edge-drop control behavior for silicon strip cold rolling with a sendzimir mill. Metals, 8, 783.

LI, Q., CAO, Z., TANVEER, M., PANDEY, H. M. & WANG, C. 2019. A semantic collaboration method based on uniform knowledge graph. IEEE Internet of Things Journal, 7, 4473-4484.

LI, X., CHEN, C.-H., ZHENG, P., WANG, Z., JIANG, Z. & JIANG, Z. 2020a. A knowledge graph-aided concept–knowledge approach for evolutionary smart product–service system development. Journal of Mechanical Design, 142, 101403.

LI, X., LYU, M., WANG, Z., CHEN, C.-H. & ZHENG, P. 2021. Exploiting knowledge graphs in industrial products and services: A survey of key aspects, challenges, and future perspectives. Computers in Industry, 129, 103449.

LI, Y., ZAKHOZHYI, V., ZHU, D. & SALAZAR, L. J. Domain Specific Knowledge Graphs as a Service to the Public: Powering Social-Impact Funding in the US. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020b. 2793-2801.

LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. R news, 2, 18-22.

LIN, Y.-J., SUH, C., LANGARI, R. & NOAH, S. 2003. On the characteristics and mechanism of rolling instability and chatter. J. Manuf. Sci. Eng., 125, 778-786.

LIU, B. 2015a. Study and application of decoupling control of roll tilt and rolling force difference to inhibit strip break during cold rolling. Metallurgical Industry Automation, 39, 40-44.

LIU, B. 2015b. Study and application of decoupling control of roll tilt and rolling force difference to inhibit strip break during cold rolling. Metallurgical Industry Automation, 40-44.

LIU, B., TANG, C. L. & GU, T. Q. A Method to Avoid Strip Breakage for Thin Strip Steel in Cold Rolling. Advanced Materials Research, 2014. Trans Tech Publ, 1211-1215.

LIU, Z. & ZHOU, J. 2020. Introduction to Graph Neural Networks. Synthesis Lectures on Artificial Intelligence and Machine Learning, 14, 1-127.

LOPES, M. R., COSTIGLIOLA, A., PINTO, R., VIEIRA, S. & SOUSA, J. M. 2019. Pharmaceutical quality control laboratory digital twin–A novel governance model for resource planning and scheduling. International Journal of Production Research, 1-15.

LOUHENKILPI, S. 2014. Continuous casting of steel. Treatise on process metallurgy. Elsevier.

LU, Y. & XU, X. 2017. A semantic web-based framework for service composition in a cloud manufacturing environment. Journal of manufacturing systems, 42, 69-81.

LU, Y., XU, X. & WANG, L. 2020. Smart manufacturing process and system automation–a critical review of the standards and envisioned scenarios. Journal of Manufacturing Systems, 56, 312-325.

MARTENS, H. 2015. Quantitative big data: where chemometrics can contribute.

MASHAYEKHI, M., TORABIAN, N. & POURSINA, M. 2011. Continuum damage mechanics analysis of strip tearing in a tandem cold rolling process. Simulation Modelling Practice and Theory, 19, 612-625.

MAZUR, V. L. & NOGOVITSYN, O. 2018. Theory and Technology of Sheet Rolling: Numerical Analysis and Applications, CRC Press.

MCCALLUM, A. 2005. Information extraction: Distilling structured data from unstructured text. Queue, 3, 48-57.

MEHRPOUYA, M., DEHGHANGHADIKOLAEI, A., FOTOVVATI, B., VOSOOGHNIA, A., EMAMIAN, S. S. & GISARIO, A. 2019. The potential of additive manufacturing in the smart factory industrial 4.0: A review. Applied Sciences, 9, 3865.

MEJOVA, Y. & SRINIVASAN, P. Exploring feature definition and selection for sentiment classifiers. Fifth international AAAI conference on weblogs and social media, 2011.

MISSBAUER, H., HAUBER, W. & STADLER, W. 2009. A scheduling system for the steelmaking-continuous casting process. A case study from the steel-making industry. International Journal of Production Research, 47, 4147-4172.

MORDIA, R. & VERMA, A. K. 2022. Visual techniques for defects detection in steel products: A comparative study. Engineering Failure Analysis, 106047.

MYKLEBUST, E. B., JIMENEZ-RUIZ, E., CHEN, J., WOLF, R. & TOLLEFSEN, K. E. Knowledge graph embedding for ecotoxicological effect prediction. International Semantic Web Conference, 2019. Springer, 490-506.

NG, K. K., CHEN, C.-H., LEE, C. K., JIAO, J. R. & YANG, Z.-X. 2021. A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives. Advanced Engineering Informatics, 47, 101246.

NGUYEN, H. L., VU, D. T. & JUNG, J. J. 2020. Knowledge graph fusion for smart systems: A survey. Information Fusion, 61, 56-70.

NICKEL, M., MURPHY, K., TRESP, V. & GABRILOVICH, E. 2015. A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104, 11-33.

ODUGUWA, V. & ROY, R. 2006. A review of rolling system design optimisation. International Journal of Machine Tools and Manufacture, 46, 912-928.

OTERO-CERDEIRA, L., RODRíGUEZ-MARTíNEZ, F. J. & GóMEZ-RODRíGUEZ, A. 2015. Ontology matching: A literature review. Expert Systems with Applications, 42, 949-971.

OU, M., CUI, P., PEI, J., ZHANG, Z. & ZHU, W. Asymmetric transitivity preserving graph embedding. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016. 1105-1114.

PANJKOVIĆ, V., GLOSS, R., STEWARD, J., DILKS, S., STEWARD, R. & FRASER, G. 2012. Causes of chatter in a hot strip mill: observations, qualitative analyses and mathematical modelling. Journal of Materials Processing Technology, 212, 954-961.

PARK, C. Y., KIM, J. W., KIM, B. & LEE, J. 2020. Prediction for manufacturing factors in a steel plate rolling smart factory using data clustering-based machine learning. IEEE Access, 8, 60890-60905.

PASZKE, A., GROSS, S., CHINTALA, S., CHANAN, G., YANG, E., DEVITO, Z., LIN, Z., DESMAISON, A., ANTIGA, L. & LERER, A. 2017. Automatic differentiation in pytorch.

PAULHEIM, H. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web, 8, 489-508.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. 2011. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12, 2825-2830.

PEI, J., HAN, J., MORTAZAVI-ASL, B., PINTO, H., CHEN, Q., DAYAL, U. & HSU, M.-C. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. Proceedings 17th international conference on data engineering, 2001. IEEE, 215-224.

PEROZZI, B., AL-RFOU, R. & SKIENA, S. Deepwalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014. 701-710.

PIETRANIK, M. & NGUYEN, N. T. 2014. A multi-attribute based framework for ontology aligning. Neurocomputing, 146, 276-290.

PUJARA, J., MIAO, H., GETOOR, L. & COHEN, W. Knowledge graph identification. International semantic web conference, 2013. Springer, 542-557.

PUNIA, S., NIKOLOPOULOS, K., SINGH, S. P., MADAAN, J. K. & LITSIOU, K. 2020. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. International Journal of Production Research, 1-16.

QIN, J., LIU, Y. & GROSVENOR, R. 2016. A categorical framework of manufacturing for industry 4.0 and beyond. Procedia Cirp, 52, 173-178.

QU, J., LI, S. & CHEN, J. 2017. Online Monitoring of Manufacturing Process Based on autoCEP. International Journal of Online Engineering, 13.

REGULSKI, K., ROJEK, G. & KUSIAK, J. Process of Ontology Construction of Rolling Metal Sheets Industrial Process. Key Engineering Materials, 2014. Trans Tech Publ, 978-985.

REN, H., DAI, H., DAI, B., CHEN, X., ZHOU, D., LESKOVEC, J. & SCHUURMANS, D. Smore: Knowledge graph completion and multi-hop reasoning in massive knowledge graphs. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022. 1472-1482.

ROS-YáñEZ, T., RUIZ, D., BARROS, J. & HOUBAERT, Y. 2004. Advances in the production of high-silicon electrical steel by thermomechanical processing and by immersion and diffusion annealing. Journal of alloys and compounds, 369, 125-130.

ROWEIS, S. T. & SAUL, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. science, 290, 2323-2326.

RUNCHANG, L., HUAN, L. & DA, Q. 2017. Cause Analysis of Strip Breakage during Cold Rolling Caused by Raw Material Defects. HEBEI M ETALLURGY, 22-25.

RUSNáK, J., MALEGA, P., SVETLíK, J., RUDY, V. & ŠMAJDA, N. 2020. The Research of the Rolling Speed Influence on the Mechanism of Strip Breaks in the Steel Rolling Process. Materials, 13, 3509.

SAMKO, O., MARSHALL, A. D. & ROSIN, P. L. 2006. Selection of the optimal parameter value for the Isomap algorithm. Pattern Recognition Letters, 27, 968-979.

SARDA, K., ACERNESE, A., NOLè, V., MANFREDI, L., GRECO, L., GLIELMO, L. & DEL VECCHIO, C. 2021. A multi-step anomaly detection strategy based on robust distances for the steel industry. IEEE Access, 9, 53827-53837.

SAUL, L., WEINBERGER, K., HAM, J., SHA, F. & LEE, D. D. 2006. Spectral methods for dimensionality reduction. Semisupervised Learning. Semisupervised Learning, 293-308.

SHAHAR, Y. 1997. A framework for knowledge-based temporal abstraction. Artificial intelligence, 90, 79-133.

SINGH, A. P. & GORDON, G. J. Relational learning via collective matrix factorization. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008. 650-658.

SINGHAL, A. 2012. Introducing the knowledge graph: things, not strings. Official google blog, 5, 16.

SOBASZEK, Ł., GOLA, A. & ŚWIĆ, A. Predictive scheduling as a part of intelligent job scheduling system. International Conference on Intelligent Systems in Production Engineering and Maintenance, 2017. Springer, 358-367.

SONG, G., WANG, X. & YANG, Q. 2018. Study on mathematical model of work roll wear in skin-pass rolling of hot steel strip. The International Journal of Advanced Manufacturing Technology, 97, 2675-2686.

SONG, G. W., TAMA, B. A., PARK, J., HWANG, J. Y., BANG, J., PARK, S. J. & LEE, S. 2019. Temperature Control Optimization in a Steel‐Making Continuous Casting Process Using a Multimodal Deep Learning Approach. steel research international, 90, 1900321.

STOREY, J. D. 2003. The positive false discovery rate: a Bayesian interpretation and the q-value. The Annals of Statistics, 31, 2013-2035.

SUYKENS, J. A. & VANDEWALLE, J. 1999. Least squares support vector machine classifiers. Neural processing letters, 9, 293-300.

SYAFRUDIN, M., ALFIAN, G., FITRIYANI, N. L. & RHEE, J. 2018. Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. Sensors, 18, 2946.

TAKAMI, K., MAHMOUDI, J., DAHLQUIST, E. & LINDENMO, M. 2011a. Multivariable data analysis of a cold rolling control system to minimise defects. The International Journal of Advanced Manufacturing Technology, 54, 553-565.

TAKAMI, K. M., MAHMOUDI, J., DAHLQUIST, E. & LINDENMO, M. 2011b. Multivariable data analysis of a cold rolling control system to minimise defects. The International Journal of Advanced Manufacturing Technology, 54, 553-565.

TAN, X., YAN, X.-T., JUSTER, N. P., RAGHUNATHAN, S. & WANG, J. 2008. Dynamic friction model and its application in flat rolling. Journal of materials processing technology, 207, 222-234.

TANG, L., LIU, J., RONG, A. & YANG, Z. 2001. A review of planning and scheduling systems and methods for integrated steel production. European Journal of operational research, 133, 1-20.

UMESHINI, S. & PSUMATHI, C. 2017. ASurvey ON DATA MINING IN STEEL INDUSTRIES. International Journal of Computer Science & Engineering Survey (IJCSES), 8.

VILLACAMPA, O. 2015. Feature selection and classification methods for decision making: a comparative analysis. Nova Southeastern University.

VILLALBA-DIEZ, J., ORDIERES-MERé, J., GONZáLEZ-MARCOS, A. & LARZABAL, A. S. 2022. Quantum Deep Learning for Steel Industry Computer Vision Quality Control. IFAC-PapersOnLine, 55, 337-342.

WANG, C. 2014a. Research on Fault Diagnosis of Belt Tearing in the Cold Rolling Process. Instrumentation Technology, 9, 16-20.

WANG, C. 2014b. Research on Fault Diagnosis of Belt Tearing in the Cold Rolling Process. Instrumentation Technology, 16-20.

WANG, D.-C., LIU, H.-M. & LIU, J. 2017. Research and development trend of shape control for cold rolling strip. Chinese Journal of Mechanical Engineering, 30, 1248-1261.

WANG, H., LI, G., LEI, Y., ZHAO, Y., DAI, Q. & WANG, J. 2005. Mathematical heat transfer model research for the improvement of continuous casting slab temperature. ISIJ international, 45, 1291-1296.

WANG, M., LI, Y., ZHANG, Y. & JIA, L. 2021a. Spatio-temporal graph convolutional neural network for remaining useful life estimation of aircraft engines. Aerospace Systems, 4, 29-36.

WANG, M., QIU, L. & WANG, X. 2021b. A survey on knowledge graph embeddings for link prediction. Symmetry, 13, 485.

WANG, S., WEI, X., NOGUEIRA DOS SANTOS, C. N., WANG, Z., NALLAPATI, R., ARNOLD, A., XIANG, B., YU, P. S. & CRUZ, I. F. Mixed-curvature multi-relational graph neural network for knowledge graph completion. Proceedings of the Web Conference 2021, 2021c. 1761-1771.

WANG, W. & ZHANG, M. 2018. Tensor deep learning model for heterogeneous data fusion in Internet of Things. IEEE Transactions on Emerging Topics in Computational Intelligence, 4, 32-41.

WANG, Z., MA, H., CHEN, H., YAN, B. & CHU, X. 2019. Performance degradation assessment of rolling bearing based on convolutional neural network and deep long-short term memory network. International Journal of Production Research, 1-13.

WITTEN, I. H. & FRANK, E. 2002. Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31, 76-77.

WORLDSTEELASSOCIATION. 2022. Steel Statistical Yearbook 2021 [Online]. https://worldsteel.org/steel-topics/statistics/steel-statistical-yearbook/. [Accessed].

WU, C., ZHANG, L., QU, P., LI, S. & JIANG, Z. 2021. Multiscale Interface Stress Characterisation in Cold Rolling. Metals and Materials International, 27, 1997-2013.

WU, Z., PAN, S., CHEN, F., LONG, G., ZHANG, C. & PHILIP, S. Y. 2020. A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems.

XIA, L., ZHENG, P., LI, X., GAO, R. X. & WANG, L. 2022. Toward cognitive predictive maintenance: A survey of graph-based approaches. Journal of Manufacturing Systems, 64, 107-120.

XIAO, Z. 2013. The analysis and improvement of FGC strategy on five tandem cold mill.

XIE, Y., YAO, C., GONG, M., CHEN, C. & QIN, A. K. 2020. Graph convolutional networks with multi-level coarsening for graph classification. Knowledge-Based Systems, 194, 105578.

XU, J. 2015. Cause Analysis for Strip-breaking of Non-oriented Silicon Steel. JIANGXI METALLURGY, 26-28.

YALCIN, A. S., KILIC, H. S. & DELEN, D. 2022. The use of multi-criteria decision-making methods in business analytics: A comprehensive literature review. Technological Forecasting and Social Change, 174, 121193.

YAN, K. & LI, H. 2006a. Causes and Countermeasures of Cold Strip Break in Masteel. Anhui Metallurgy, 36-37.

YAN, K. & LI, H. 2006b. Causes and Countermeasures of Cold Strip Break in Masteel. Anhui Metallurgy, 1, 36-37.

YAN, Y., SUN, Q., CHEN, J. & PAN, H. 2015. Effect of processing parameters on edge cracking in cold rolling. Materials and Manufacturing Processes, 30, 1174-1178.

YANG, C., LIU, Z., ZHAO, D., SUN, M. & CHANG, E. Network representation learning with rich text information. Twenty-fourth international joint conference on artificial intelligence, 2015.

YANG, H., YAN, X., DAI, X., CHEN, Y. & CHENG, J. 2020. Self-enhanced gnn: Improving graph neural networks using model outputs. arXiv preprint arXiv:2002.07518.

YANG, S., REBMANN, A., TANG, M., MORAVEC, R., BEHRMANN, D., BAIRD, M. & BEQUETTE, B. W. 2021. Process monitoring using causal graphical models, with application to clogging detection in steel continuous casting. Journal of Process Control, 105, 259-266.

YE, J., JANARDAN, R. & LI, Q. 2004. Two-dimensional linear discriminant analysis. Advances in neural information processing systems, 17.

YE, X., WU, X. & GUO, Y. Real-time quality prediction of casting billet based on random forest algorithm. 2018 IEEE International Conference on Progress in Informatics and Computing (PIC), 2018. IEEE, 140-143.

YIN, J., TANG, M., CAO, J., YOU, M., WANG, H. & ALAZAB, M. 2022. Knowledge-Driven Cybersecurity intelligence: Software Vulnerability Co-exploitation Behaviour Discovery. IEEE Transactions on Industrial Informatics.

YU, S., HE, T. & GLASS, J. R. 2020. Constructing a Knowledge Graph from Unstructured Documents without External Alignment.

YUN, I., WILSON, W. & EHMANN, K. 1998. Review of chatter studies in cold rolling. International Journal of Machine Tools and Manufacture, 38, 1499-1530.

ZAREMBA, W., SUTSKEVER, I. & VINYALS, O. 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

ZHANG, X., KANO, M., TANI, M., MORI, J. & HARADA, K. 2018. Defect data modeling and analysis for improving product quality and productivity in steel industry. Computer aided chemical engineering. Elsevier.

ZHANG, Y., JIN, R. & ZHOU, Z.-H. 2010. Understanding bag-of-words model: a statistical framework. International journal of machine learning and cybernetics, 1, 43-52.

ZHANG, Y., LI, Y., WEI, X. & JIA, L. Adaptive spatio-temporal graph convolutional neural network for remaining useful life estimation. 2020 International joint conference on neural networks (IJCNN), 2020a. IEEE, 1-7.

ZHANG, Y., ZHENG, X., LIU, L. & MA, H. Multivariate and multi-frequency LSTM based fine-grained productivity forecasting for industrial IoT. 2020 16th International Conference on Mobility, Sensing and Networking (MSN), 2020b. IEEE, 307-314.

ZHAO, J., LIU, Q., PEDRYCZ, W. & LI, D. 2012. Effective noise estimation-based online prediction for byproduct gas system in steel industry. IEEE Transactions on Industrial Informatics, 8, 953-963.

ZHAO, M., WANG, H., GUO, J., LIU, D., XIE, C., LIU, Q. & CHENG, Z. 2019. Construction of an industrial knowledge graph for unstructured chinese text learning. Applied Sciences, 9, 2720.

ZHENG, X., WANG, B., ZHAO, Y., MAO, S. & TANG, Y. 2021. A knowledge graph method for hazardous chemical management: Ontology design and entity identification. Neurocomputing, 430, 104-111.

ZHOU, B., HUA, B., GU, X., LU, Y., PENG, T., ZHENG, Y., SHEN, X. & BAO, J. 2021. An end-to-end tabular information-oriented causality event evolutionary knowledge graph for manufacturing documents. Advanced Engineering Informatics, 50, 101441.

ZUEHLKE, D. 2010. SmartFactory—Towards a factory-of-things. Annual reviews in control, 34, 129-138.

Appendix A. Advanced Data Analytics Technologies

A1 Data mining

The continuous development of complex systems and instrumental techniques has led to an immense amount of data being created, collected, and stored. Due to the large amount of data that can be utilized, data mining has received considerable attention from the information industry over the past few decades. As a highly application-driven discipline, data mining has seen great success in many applications, such as business intelligence (Yalcin et al., 2022), web search engines (Yin et al., 2022), and social networks (Kumar et al., 2022) etc. Essentially, data mining is the process of discovering interesting patterns, models, and other kinds of knowledge in large data sets, which is often used to refer to the entire knowledge discovery process (Han et al., 2022). In other words, it is through the process of data mining that knowledge can be converted from a large amount of data.

As a general technology, data mining can be applied to any type of data, provided the data are meaningful for a target application. Based on the data format, the datasets are classified into structured data, semi-structured data and unstructured data. In the structured type of data, the structure is uniform, record- or table-like, defined by their data dictionaries, such as data cubes, data matrices, and many data warehouses. Each attribute has a fixed value range and semantic meaning. A more sophisticated type of data refers to semi-structured data, such as a graph or network data. Beyond such structured or semi-structured data, there are also large amounts of unstructured data, such as text data and multimedia. Real-world data can often be a mixture of structured data, semi-structured data, and unstructured data. Based on application domains, data mining methods are often developed for mining some particular type of data, and their results can be integrated and coordinated to serve the overall goal.

As a knowledge discovery process, data mining typically involves data cleaning, data integration, data selection, data transformation, pattern and model discovery, pattern or model evaluation, and knowledge presentation. In connection with strip-breakage prediction in cold rolling, the entire process of data mining is described below:

- **Target Understanding:** Understanding the background knowledge of cold rolling in the steel industry after identifying the meaningful target, such as the entire process of cold rolling.

- **Data preparation:** Collecting and preparing data relevant to the strip-breakage phenomena from the steel industry.

  a. Data cleaning: As the data is collected from the real world, there is a possibility that it contains some impurities. It is necessary to remove noise and inconsistent data.

  b. Data integration: The relevant information obtained from multiple sources usually contributes to one intricate phenomenon in the industrial processes. Multiple data sources are combined, which usually leads to more expressive and informative information than that of each single data source.

  c. Data transformation: A summary or aggregation operation is performed on data to transform and consolidate the data into a form that is suitable for mining.

  d. Data selection: From the database, data relevant to the analysis task is retrieved.

- **Data mining:** The strip-breakage predicting models are constructed based on intelligent machine learning methods, which aim at extracting potential patterns.

- **Model evaluation:** With the aim of identifying the truly effective models, the models are evaluated using different metrics.

- **Operating:** After constructing and evaluating the data-driven models, they can be applied in the actual cold-rolled production management.

- **Knowledge presentation & Storage:** After obtaining the information for analysis, the previous stages are adjusted by extracting knowledge. Meanwhile, visualisation and knowledge representation techniques are deployed to store and present mined knowledge.

A2 Machine learning

Machine learning has rapid growth with the development of many new methodologies and applications in recent years, from conventional machine learning algorithms to graphical models and deep learning algorithms (Han et al., 2022). As mentioned previously, although machine learning develops rapidly, it is challenging to implement enormous volumes of time-series data. In recent years, prevailing machine learning algorithms have proven useful for manipulating multivariate time-series data. Detailed descriptions have not been provided for four benchmarking algorithms, which involve SVM, RF, and RNN. The details of these algorithms are presented below:

- **SVM** (Cortes and Vapnik, 1995) has become a widely used tool for supervised learning, with high versatility that extends across classification and regression tasks. In this thesis, SVM classifier was adopted to predict the strip breakage in the cold rolling process. A SVM decision function is actually an optimal "hyperplane" which is deployed to divide observations into two classes through patterns of information from known observations. By using this hyperplane, an unobserved data set can be labelled with the most likely label. Assuming the classification with bias less than ε is deemed as correct classification, the incorrect classification results in a high penalty to the algorithm. Kernel functions are the core components of SVM classifiers. x_i and x_j are two vectors. Four kernel functions have formed the most prevailing implementation:

$$\text{Linear kernel: } K\left(x_i, x_j\right) = x_i^T x_j \tag{A1}$$

$$\text{Polynomial kernel: } K\left(x_i, x_j\right) = (\gamma x_i^T x_j + b)^d \tag{A2}$$

$$\text{RBF kernel: } K\left(x_i, x_j\right) = exp(-\gamma \left\|x_i - x_j\right\|^2) \tag{A3}$$

$$\text{Sigmoid Kernel: } K\left(x_i, x_j\right) = tanh(\gamma x_i^T x_j + b) \tag{A4}$$

where γ, b, and d are parameters that need to be set.

- **RF** (Breiman, 2001) is a classifier consisting of a collection of tree-structured classifiers $\{h(X, \Theta_k)|\mathrm{k} = 1, 2, \ldots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\boldsymbol{X}$. Given an ensemble of classifiers $h_1(\boldsymbol{X}), h_2(\boldsymbol{X}), \ldots, h_k(\boldsymbol{X})$, and with the training set drawn at random from the distribution of random vector Y, **X**. The margin function (represented by mg is defined as follow:

$$mg(\mathbf{X}, Y) = av_k \boldsymbol{I}(h_k(\mathbf{X}) = Y) - \max_{j \neq Y} av_k \boldsymbol{I}(h_k(\mathbf{X}) = j) \tag{A5}$$

where $\boldsymbol{I}(\cdot)$ is the indicator function. The margin measures the extent to which the average number of votes at $\mathbf{X}, Y$ for the right class exceeds the average vote for any other class. The larger the margin, the more confidence in the classification. The generalization error is given below:

$$PE^* = P_{\mathbf{X},Y}(mg(\mathbf{X}, Y) < 0) \tag{A6}$$

where the subscript $\mathbf{X}, Y$ demonstrates that the probability is over the $\mathbf{X}, Y$ space. Eqs. (A7) indicates that $\{h(X, \Theta_k)|\mathrm{k} = 1, 2, \ldots, N\}$ follow the rule of large numbers as the value of N is large enough for the model, and the classifier has enough trees. Meanwhile, it has been proved that the upper limit of generalization error is convergent as the almost everywhere convergence of random vectors $\theta, \ldots, PE^*$. It is given as follows:

$$P_{\mathbf{X},Y}(P_\Theta(h(\mathbf{X},\Theta) = Y) - \max_{j \neq Y} P_\Theta(h(\mathbf{X},\Theta) = j) < 0) \tag{A7}$$

$$\bar{\varepsilon} \leq \frac{\bar{\rho}(1-s^2)}{s^2} \tag{A8}$$

where $\bar{\varepsilon}$ indicates the upper limit of the generalization error, and $\bar{\rho}$ means the average correlation coefficient between trees, and s represents the average classification performance of the decision trees. Eqs. (A8) illustrates that the larger the average correlation coefficient is, the larger the upper limit of generalisation error will be. Likewise, the larger the average classification is, the larger the upper limit of generalisation error will be. Essentially, the classification performance is affected by two factors, one is the overall performance of trees, and the other is the diversity between trees.

- **RNN** (Zaremba et al., 2014) is a type of deep learning algorithm which is well-known for processing sequential data. Unlike CNNs, RNNs provide feedback from the previous state to the current state of the hidden units. RNN-based models were built to predict the strip breakage in the cold rolling process in this thesis.

  In this structure, the input vectors are fed into the RNN instead of using a fixed number of input vectors, as done in the conventional network structures. Besides, this architecture can take advantage of all the available input information up to the current time. In addition, the depth of the RNN can be defined according to real conditions. It can be seen that the final output not only depends on the current input but also depends on the output of previously hidden layers. The mathematic process is defined below:

$$t_i = \boldsymbol{W}_{hx}x_i + \boldsymbol{W}_{hh}x_{i-1} + \boldsymbol{b}_h \tag{A9}$$

$$h_i = \sigma(t_i) \tag{A10}$$

$$s_i = \boldsymbol{W}_{oh} h_i + \boldsymbol{b}_y \tag{A11}$$

$$\hat{o} = g(s_i) \tag{A12}$$

where x_i indicates the input variables, $\boldsymbol{W}_{hx}$, $\boldsymbol{W}_{hh}$ and $\boldsymbol{W}_{ox}$ are weight matrices, $\boldsymbol{b}_h$ and $\boldsymbol{b}_y$ are bias vectors, σ and g are sigmoid functions, t_i, h_i and s_i are the temporary variables, and $\hat{o}$ is the expected output. The cost function is defined as follows:

$$f = \sum_i \left(\frac{\|\hat{o}_i - o_i\|}{2}\right) \tag{A13}$$

where o_i is the actual output. As such, the output at $t + 1$ is the joint function of the input at $t + 1$ and the historical data. The RNN simulates the correlation in sequential data, and the depth of the network is the time span.

A3 Graph representation learning

As data becomes increasingly interconnected and systems increasingly sophisticated, it is essential to make use of the rich and evolving relationships within our data. Generally, although relevant data obtained from multiple sources usually contributes to one intricate phenomenon, data are diverse, heterogeneous, and fragmented in real complex industrial scenarios. Meanwhile, multiple data sources usually lead to more expressive and informative information than that of each single data source. However, the conventional approaches concatenate feature vectors to integrate different facets, not considering the semantic gaps between them. In this context, data challenges revolve around relationships rather than just tabulating discrete data. In this case, graph representation has been energetically developed as it provides a powerful tool for connected data, such as KG. As a collection of nodes and links, graph data displays a powerful expressive ability and a high degree of modelling flexibility, making it a promising semantic network.

Specifically, a graph is represented by $G = (\mathbf{V}, \mathbf{E})$, where **V** denotes a set of nodes, and **E** is a set of edges between each pair of nodes. Specifically, the i_{th} node is indicated by $v_i \in \mathbf{V}$, and the features of all nodes are defined as $\mathbf{X}_v, \forall v \in \mathbf{V}$. Meanwhile, the adjacency matrix $\mathbf{A} \in \mathbf{R}^{|n|\times|n|}, A_{\boldsymbol{ij}} \in \{0, 1\}$ is usually used to describe **E**, which is a $|n| \times |n|$ square matrix. If an edge exists between node v_i and node v_j, then $A_{\boldsymbol{ij}} = 1$, otherwise $A_{\boldsymbol{ij}} = 0$.

By representing interacting entities in a graph, relational knowledge can be efficiently stored and accessed. Analysis of graph data can provide significant insights into community detection, behaviour analysis, and other useful applications, including node classification, link prediction, and clustering. However, when it comes to the inputs of conventional machine learning models, they usually take feature vectors representing objects in terms of tabular attributes (such as numeric attributes and categorical attributes). Thus, conventional machine-learning approaches cannot be directly applied to graph-structured data. In this context, it has been developed various graph representation learning techniques that convert raw graph data into a high-dimensional vector while preserving the intrinsic properties of the graph. There is also a term for this process called graph representation learning, which is becoming a hot topic and a challenge in recent years. As a result of a learned graph representation, machine-learning tools can be effectively utilized to perform downstream tasks.

Depending on the different theories, four types of graph representation learning methods were defined and have formed the most prevailing taxonomy, which are dimension-reduction-based methods, random-walk-based methods, matrix-factorization-based methods, and neural-network-based methods:

- **Dimension-reduction-based methods:** Methods based on dimension reduction have been developed to reduce high-dimensional graph data into a lower-dimensional representation while retaining the desired properties of the original data. These methods can be classified into linear and nonlinear two types.

- **Random-walk-based methods:** Random-walk-based methods sample a graph with a large number of paths by starting walks from random initial nodes, such as DeepWalk. These paths indicate the context of connected vertices. The randomness of walks gives the ability to explore the graph and capture both the global and the local structural information by walking through neighbouring vertices. After the paths are built, probability models such as skip-gram and bag-of-words can be performed on these randomly sampled paths to learn the node representation.

- **Matrix-factorization-based methods:** Matrix-factorization-based methods, also called graph factorization, was the first one to achieve graph embedding in time for node embedding tasks. With the aim of conducting the embedding, graph factorization factorizes the adjacency matrix of a graph. It corresponds to a structure-preserving dimensionality reduction process.

- **Neural-network-based methods:** Inspired by the success of CNNs, researchers attempt to generalize and apply them to graphs. The input can be sampled from a graph or the whole graph itself through paths. Researchers reformat the input graph to fit the original CNN model designed for the Euclidean domain. It is also possible to generalize the deep neural model to non-Euclidean graphs. Several neural-network-based graph representation learning methods are proposed, such as GCN, GraphSAGE, GAE etc.

A wide range of conventional methods is employed in graph representation learning in order to reduce dimensionality. Despite being mathematically transparent, the majority of them cannot be represented in graphs as high-order proximity. Instead of embedding the entire graph, DeepWalk-based methods sample the neighbourhood information of each node statistically. The main advantage of these approaches is that they capture long-distance relationships between nodes, whereas graphs may not fully preserve global information. By using statistics associated with pairwise similarities, matrix factorization methods are able to extract the possible patterns of the graph representation. In comparison to deep-learning algorithms that utilize only a local context window, these algorithms are capable of outperforming methods that rely on random walks. As a result, matrix factorization is inefficient and cannot be scaled to large graphs. As a result, proximity matrices and decompositions require more computing power and storage space. The advantage of factorization-based methods, such as LLE, Laplacian Eigenmaps, and graph factorization, is that they usually preserve first-order proximity in contrast to DeepWalk-based methods, which preserve second-order proximity.

Appendix B. Results of Correlation Tests

As stated in Section 5.3.1, Pearson correlation coefficient, Kendall Tau rank correlation coefficient, and Spearman rank relational coefficient were applied and results of these three correlation methods is shown in the Appendix B. B1 gives the Pearson correlation results of ten variables, such as *'Width'*, *' Tonnes on cutter Desk '*, *' Tonnes on Cutter Drive '*, *' Crown max '* etc. B2 presents the Kendall Tau rank correlation results of ten variables, which are '*Weight ingoing'*, *'Weight outgoing'*, *'Gauge'*, *'TOFF'* etc. B3 gives the Spearman rank correlation results of ten variables, such as *'Heavy end'*, *' Trimmed on AP1'*, *'Z6Temp std dev'* etc. These three correlation coefficient ranges from -1 to 1.

B1 Results of Pearson correlation coefficients of ten variables.

Variables	Width	Tonnes on cutter desk	Tonnes on cutter drive	Crown max	Crown min	Z1Temp max	Z1Temp min	Z2Temp max	Z2Temp min	Z2ATemp max
Width	1	0.039	0.05	0.032	-0.009	-0.014	0.023	0.016	0.026	0.016
Tonnes on cutter desk	0.039	1	0.262	-0.059	-0.075	-0.005	-0.024	0.005	-0.021	0.019
Tonnes on cutter drive	0.05	0.262	1	-0.052	-0.121	0.058	-0.076	-0.015	-0.073	-0.006
Crown max	0.032	-0.059	-0.052	1	-0.071	-0.033	0.004	-0.009	0.002	-0.005
Crown min	-0.009	-0.075	-0.121	-0.071	1	-0.094	0.062	0.038	0.059	0.051
Z1Temp max	-0.014	-0.005	0.058	-0.033	-0.094	1	-0.14	0.026	-0.144	0.022
Z1Temp min	0.023	-0.024	-0.076	0.004	0.062	-0.14	1	0.397	0.998	0.374
Z2Temp max	0.016	0.005	-0.015	-0.009	0.038	0.026	0.397	1	0.403	0.984
Z2Temp min	0.026	-0.021	-0.073	0.002	0.059	-0.144	0.998	0.403	1	0.382
Z2ATemp max	0.016	0.019	-0.006	-0.005	0.051	0.022	0.374	0.984	0.382	1

B2 Results of Kendall Tau rank correlation of ten variables.

Variables	Weight ingoing	Weight outgoing	Gauge	TOFF	Z3Temp ave	Z3Temp max	Z4Temp std dev	Z4Temp min	LineSpeed ave	LineSpeed std dev
Weight ingoing	1	0.31	-0.254	-0.009	0.007	-0.04	-0.016	-0.054	-0.311	-0.122
Weight outgoing	0.31	1	-0.435	0.002	-0.007	-0.045	-0.034	-0.053	-0.296	-0.158
Gauge	-0.254	-0.435	1	-0.014	0.034	0.003	0.003	-0.065	-0.3	-0.068
TOFF	-0.009	0.002	-0.014	1	-0.001	-0.01	-0.007	-0.103	-0.271	-0.113
Z3Temp ave	0.007	-0.007	0.034	-0.001	1	0.6	-0.04	0.412	-0.301	-0.166
Z3Temp max	-0.04	-0.045	0.003	-0.01	0.6	1	0.231	0.174	-0.369	-0.039
Z4Temp std dev	-0.016	-0.034	0.003	-0.007	-0.04	0.231	1	-0.3	-0.323	0.067
Z4Temp min	-0.054	-0.053	-0.065	-0.103	0.412	0.174	-0.3	1	-0.282	-0.223
LineSpeed ave	-0.311	-0.296	-0.3	-0.271	-0.301	-0.369	-0.323	-0.282	1	-0.152
LineSpeed std dev	-0.122	-0.158	-0.068	-0.113	-0.166	-0.039	0.067	-0.223	-0.152	1

B3 Results of Spearman rank correlations of ten variables.

Variables	Heavy end	Trimmed on AP1	Z6Temp std dev	Z6Temp max	GasFlow ave	GasFlow min	JetFlow ave	JetFlow max	CoolingPyro ave	CoolingPyro std dev
Heavy end	1	-0.003	0.049	-0.008	0.012	-0.004	-0.026	-0.031	-0.023	0.097
Trimmed on AP1	-0.003	1	-0.089	0.072	-0.031	0.008	-0.07	-0.061	0.048	-0.053
Z6Temp std dev	0.049	-0.089	1	0.124	-0.067	0.402	-0.16	0.064	-0.207	0.606
Z6Temp max	-0.008	0.072	0.124	1	0.231	0.001	0.134	0.197	0.408	0.125
GasFlow ave	0.012	-0.031	-0.067	0.231	1	0.734	0.592	0.54	0.69	-0.198
GasFlow min	-0.004	0.008	-0.402	0.001	0.734	1	0.539	0.367	0.621	-0.531
JetFlow ave	-0.026	-0.07	-0.16	0.134	0.592	0.539	1	0.857	0.55	-0.245
JetFlow max	-0.031	-0.061	0.064	0.197	0.54	0.367	0.857	1	0.41	-0.012
CoolingPyro ave	-0.023	0.048	-0.207	0.408	0.69	0.621	0.55	0.41	1	-0.413
CoolingPyro std dev	0.097	-0.053	0.606	0.125	-0.198	-0.531	-0.245	-0.012	-0.413	1