Information Fusion 101 (2024) 101985

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

A knowledge graph-supported information fusion approach for multi-faceted conceptual modelling

Zheyuan Chen, Yuwei Wan, Ying Liu *, Agustin Valera-Medina

Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

ARTICLE INFO

Keywords:
Information integration
Multi-faceted modelling
Knowledge graph-supported data fusion

ABSTRACT

It has become progressively more evident that a single data source is unable to comprehensively capture the variability of a multi-faceted concept, such as product design, driving behaviour or human trust, 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. The consideration of intra-facets relationships is also indispensable. In this context, a knowledge graph (KG), which can aggregate the relationships of multiple aspects by semantic associations, was exploited to facilitate the multi-faceted conceptual modelling based on heterogeneous and semantic-rich data. Firstly, rules of fault mechanism are extracted from the existing domain knowledge repository, and node attributes are extracted from multi-sourced data. Through abstraction and tokenisation of existing knowledge repository and concept-centric data, rules of fault mechanism were symbolised and integrated with the node attributes, which served as the entities for the concept-centric knowledge graph (CKG). Subsequently, the transformation of process data to a stack of temporal graphs was conducted under the CKG backbone. Lastly, the graph convolutional network (GCN) model was applied to extract temporal and attribute correlation features from the graphs, and a temporal convolution network (TCN) was built for conceptual modelling using these features. The effectiveness of the proposed approach and the close synergy between the KG-supported approach and multi-faceted conceptual modelling is demonstrated and substantiated in a case study using real-world data.

1. Introduction

Multi-faceted modelling has received much attention in a wide range of research areas, such as product design [1] and social science [2], due to its ability to capture the complexity of real-world problems more accurately. A multi-faceted concept is a concept that can be described and analysed from multiple perspectives or facets, each capturing a different aspect of the concept. For instance, in the manufacturing domain, a machine's performance can be affected by numerous factors, such as physical properties, operating conditions, and external environmental factors. Considering these multiple facets can lead to a more comprehensive understanding of the concept and improved decision-making.

In essence, multi-faceted modelling provides additional information from which those concepts can be better and more comprehensively modelled. However, another major issue in multi-faceted modelling is the intra-aspects relationships amongst different facets, which is usually ignored under the assumption that features from different facets are independent of each other. In the meantime, for the modelling of a multi-faceted concept, no single data source can capture the complexity of all the factors relevant to a multi-faceted concept [3]. This introduces challenges in handling heterogeneous data from different sources and integrating them effectively to generate meaningful insights. For example, in the era of IIoT, data science is fuelling the rapid development of more intelligent manufacturing by enabling a shift towards data-driven decision-making. In the production workshop of the manufacturing sector, the number of IoT devices is constantly increasing. As a result, the amount of information being exchanged between devices is expanding rapidly. Accordingly, it is advantageous to combine information from different sensors to compensate for their limitations.

Furthermore, the necessity to incorporate information fusion can be summarised as follows: Firstly, multisource heterogeneous data is typically presented in large quantities, whereas fusion can function to

* Corresponding author.
E-mail address: LiuY81@Cardiff.ac.uk (Y. Liu).

https://doi.org/10.1016/j.inffus.2023.101985
Received 8 November 2022; Received in revised form 28 June 2023; Accepted 23 August 2023
Available online 25 August 2023

reduce the size and dimensions of the dataset by extracting information that is useful [4]. Secondly, it is possible to combine data from multiple sources on the same phenomenon to generate a collection of collective values, while it is also possible to utilise potentially unnoticed events to generate informative outcomes [5]. Thirdly, due to the multiplicity of facets associated with a concept of this nature, a single measuring method is usually insufficient when addressing problem spaces of this type. Alternatively, multiple data sources can facilitate the collection of more comprehensive information [6].

Amongst the heterogeneous data surrounding a multi-faceted concept, a significant amount of this data consists of rich information, such as technical documents and time-series data generated through manufacturing processes. It is worth noting that these data are usually both structured and non-structured, time-varying, multisource, and semantically rich [7]. Processing such data requires the use of semantic technologies, such as ontology and the semantic web [8]. Nevertheless, the documentation contains a substantial amount of information. Ontologies can be used to develop semantic descriptions of such concepts. However, the development of a unified, standardised, and fine-grained description is challenging [9]. As a matter of fact, there is a growing need for sophisticated approaches to assist machines in understanding the context of multi-sourced content resulting from the proliferation of industrial data. Meanwhile, there are numerous challenges for analysts, including heterogeneity in the data sources and formats, a lack of comprehensive and integrated knowledge repositories, and a discrepancy in terminology [10].

Since an ontology is based on rule representations, it has 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 [11]. In order to provide additional detail regarding the limitations and circumstances of our issue, creating models for multifaceted concepts within the engineering field requires intricate manoeuvring of numerous components. A diverse range of domains, including mechanical, electrical, chemical and even the human factor, contribute to these factors. A wide range is covered by the diverse domains. The involvement of domain experts with a comprehensive understanding of the nuances and interplay of these factors is necessary to interpret these intricacies. Effective interpretation requires a thorough comprehension of the subject matter. Given this context, knowledge graphs (KGs) can 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-dependant 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]. Many recent works elaborate on how KG can aid in organising and networking massive and heterogeneous manufacturing resources semantically. This feature of KG proves highly beneficial in our context, where the data originates from diverse and heterogeneous sources [9,14]. In this case, KGs can be used as a basis for developing multi-faceted modelling by extracting the semantics, which are retrieved from different vocabularies and semantic repositories, which are used to enrich the semantic description of resources using annotations. In contrast, although distrust is recognised to play an equivalently important role as trust, the investigation of utilising knowledge graphs in multi-faceted modelling is still in its infancy.

With the aim of addressing these challenges and facilitating multi-faceted modelling, this paper proposes a fusion architecture for the modelling of multi-faceted concepts, leveraging the power of knowledge graphs (KGs). Knowledge graphs can facilitate the construction of unified standard representations for data fusion by representing knowledge in the form of entities and relations. They can be used as a basis for developing multi-faceted modelling by extracting the semantics, which are retrieved from different vocabularies and semantic repositories and is used to enrich the semantic description of resources using annotations. In this study, we propose the use of knowledge graphs (KGs) as a novel approach to address the challenges of integrating heterogeneous data sources. KGs provide a unified representation of complex relationships amongst entities, attributes, and relationships, facilitating data integration and fusion. By leveraging semantic technologies, KGs enrich the understanding of the data and help bridge the gap between heterogeneous data sources with varying formats, terminologies, and structures. Additionally, KGs offer scalability and extensibility, allowing for easy updates and extensions with new data or domain knowledge. Finally, KGs support advanced querying and reasoning capabilities, enabling more sophisticated analyses and inferences to be drawn from the integrated data. 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 GCN model was applied to extract features from these temporal graphs, and these features were fed into a TCN model for fault concept modelling. As opposed to existing methods, this method is capable of fault diagnosis based on fault mechanism knowledge or data-driven models and combines them with knowledge-based fault diagnosis to form a fault alert diagnosis system for auxiliary decision-making.

The main contributions of this paper are as follows: (1) A novel fusion architecture for multi-faceted modelling: a new approach to multi-faceted model concepts based on knowledge graphs and data fusion is proposed. This architecture addresses the limitations of current methods that either rely solely on data-driven models or focus exclusively on knowledge-based fault diagnosis. (2) A knowledge graph is built based on the existing knowledge repository and workshop empiricism from domain experts surrounding the multi-faceted concept. This work combines domain knowledge from experts with data-driven insights to create a comprehensive knowledge graph. This approach enables better modelling of multi-faceted concepts by capturing the complex relationships amongst different facets and providing a unified representation for data fusion. (3) Practical application to industrial fault diagnosis: We demonstrate the effectiveness of our proposed approach in the context of industrial fault diagnosis. This method outperforms existing techniques in terms of precision, recall, and F1 score, showcasing its potential for real-world applications.

The remainder of this paper is organised as follows. Related studies are reviewed in Section 2 regarding KG construction, KG-aided data fusion and information integration for multi-faceted modelling is provided. The flowchart of the proposed methodology is illustrated and described for multi-faceted conceptual modelling with the support of a knowledge graph in Section 3. Section 4 presents the conducted experiments and results. Section 5 concludes the paper.

2. Related work

As discussed in the preceding, one of the fundamental issues in modelling a multi-faceted concept is to integrate multi-sourced information with the connection of semantic gaps. In this context, this section reviews existing research work on multi-faceted modelling with a focus on knowledge graph-aided methods. To be specific, some recent research on KG construction for analysis under an industrial context was introduced in the first place. Secondly, recent studies on knowledge graph-aided data fusion were investigated. Lastly, the studies on multi-faceted modelling by information integration related to this paper were also discussed.

2.1. Knowledge graph construction under industrial context

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 [15]. 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 [12]. 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 [13]. 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.

As a method of representing and managing knowledge, ontologies utilise predefined classes and properties to provide machines and humans with an easy-to-understand model of knowledge [16]. In contrast, an ontology describes only the general entities or concepts that share the same properties rather than a particular individual within a given field. From recommendation systems [17] to product family design [18,19], ontologies have been employed in a variety of applications.

The use of KGs in industry and academia is becoming more widespread; however, their construction suffers from incompleteness, which negatively affects their utility as real-life tools [20]. Consequently, the research community has provided technical solutions to solve this problem, commonly referred to as KG augmentation or completion approaches. In the process of KG completion, new facts are added to the KG by applying new probability entities and/or new relationships. A link prediction algorithm can be applied to a number of uses, including predicting new friends in social networks and developing recommender systems for a variety of other applications. Recent attention has been directed toward a new cohort of models in this context. An entity-related KG model embeds the constituents (entities and relationships) in a semantically continuous space of low dimension [21].

2.2. Knowledge graph-aided data fusion

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 [22]. 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 [15]. 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 [23] proposed a scheme for calculating non-linear distributions of IoT data using deep learning. As a result of the fusion of multisource 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 [24] 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 [25], the topological and semantic similarities between multiple sources of knowledge using two knowledge graphs are analysed. With the aim of implementing semi-automatic linkage amongst 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 [26].

Data fusion tasks using graph neuron networks (GNN) and their variants are showing promise in a variety of applications due to the advancement of graph neuron networks (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 in [27] which combines 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 [28] 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.

Besides the GNN approaches for data fusion tasks, recent studies have shown that combining multiple sources of information can improve and regularise the feature representation, leading to better performance in conceptual modelling tasks. This idea is related to the concept of multiple knowledge representations, as proposed by [29]. In our proposed method, we leverage multiple features from different sources, such as time-series data and domain knowledge represented in the knowledge graph, to model the fault diagnosis problem from multiple perspectives. By aggregating these multiple features through the proposed graph-based fusion approach, we aim to capture a more comprehensive understanding of the system and achieve better performance in fault diagnosis. The benefits of leveraging multiple features have been observed in various domains, including natural language processing and computer vision. For instance, the Transformer architecture, as proposed by [29], has achieved state-of-the-art performance in various NLP tasks by combining multiple features through a self-attention mechanism. In our work, we aim to apply a similar concept to the fault diagnosis problem and demonstrate its effectiveness through experimental evaluations. In this case, leveraging multiple features is a promising direction for improving conceptual modelling tasks, and our proposed method is an example of how this concept can be applied to the fault diagnosis problem.

Lastly, in recent years, there have been increasing efforts to develop fault diagnosis methods using knowledge graphs. A fault diagnosis method based on a knowledge graph for rotating machinery. The method employs an unsupervised learning algorithm to cluster the equipment state data and utilises the knowledge graph to capture the semantic relationships between the data. Similarly, Yang et al. developed a knowledge graph-based approach for fault diagnosis of power transformers [30]. The method utilises a hybrid model that combines deep learning and graph convolutional networks to extract features from the knowledge graph and perform fault diagnosis. Guo et al. proposed a fault diagnosis framework based on knowledge graphs and deep learning for wind turbines [31]. The approach involves constructing a knowledge graph from the equipment data and using graph convolutional networks to learn the relationships between the equipment features and detect faults. In a similar vein, Wu et al. developed a knowledge graph-based approach for fault diagnosis of reciprocating compressors [32]. The method utilises a graph neural network to learn the relationships between the equipment data and perform fault diagnosis. Li et al. proposed a fault diagnosis method based on a knowledge graph and adversarial training for rolling bearings [33]. The approach involves constructing a knowledge graph from the equipment data and using an adversarial learning algorithm to improve the robustness of the fault diagnosis model. These recent works highlight the potential of knowledge graph-based methods for fault diagnosis and demonstrate the effectiveness of using graph-based models for multi-faceted feature modelling and data fusion.

2.3. Information integration for multi-faceted modelling

There are emerging approaches for multi-touch interaction on IoT devices [34], and a few open problems can be identified to stimulate further research and development on information integration for multi-faceted modelling. The process of machine learning and data mining involves applying advanced techniques that are capable of fusing knowledge from a variety of datasets organically together. Several data fusion methodologies can be categorised according to the stage-based, feature level-based, and semantic meaning-based data fusion methodologies described in [35]. Using recent advances in data fusion, such as the semantic web and big data technologies, [29] propose an effective agricultural ontology-based knowledge fusion method to enhance the identification and fusion of new and existing data sets to enhance big data analytics. By incorporating noisy and sparse timestamp information into data fusion strategies based on voting, PageRank, and Knowledge-Based Trust, this paper examines the extent to which the performance of these strategies can be improved. Through a machine-learning-based approach, a machine-learning-based approach [36] was proposed based on timestamp propagation between web tables in order to overcome sparsity with the consideration of different types of noisy timestamps in the data fusion process. An algorithm based on Flexible Manifold Embedding was also proposed as a method for multi-metric graph fusion [37]. Recently, Transformer models have gained popularity in time series analysis, particularly in tasks that require the modelling of temporal information. A notable example is the work [38] which introduced the Align and Told framework for boosting text-video retrieval.

A holistic approach for integrating, enhancing, and unifying knowledge graphs are reviewed in [39] with the aim of providing a broad, complete, and systematic overview of the definitions and challenges of knowledge graph fusion. In order to allow multimodal fusion of incomplete data within a heterogeneous graph structure, a Heterogeneous Graph-based Multimodal Fusion (HGMF) approach is proposed [40]. Similar to this, a graph-based deep neural network was developed to simulate brain structure and function simultaneously as a function of Mild Cognitive Impairment (MCI) [41] and iteratively updated to maximise the ability to differentiate MCI patients by integrating functional information. For the process simplification of the product family design, a method for building a multi-faceted ontology that is semantically annotated has been proposed [1] to suggest semantically related annotations automatically based on the design and manufacturing repository. An approach was proposed during an investigation of undermined ground for historical coal mining with the aim of establishing the development aspects of multi-faceted geophysical modelling systems [42].

Several recent works have explored the use of knowledge graphs for data fusion and information integration, addressing various challenges in different domains [5,41,43]. Some of these approaches have focused on constructing knowledge graphs from heterogeneous data sources [7, 30], while others have proposed methods for enriching existing knowledge graphs with new data or external knowledge [9,12]. A few studies have also explored the use of knowledge graphs for fault diagnosis in the context of industrial systems [30,31]. However, most of these works either concentrate on specific aspects of the problem or lack a comprehensive framework that combines knowledge graph-based data fusion with advanced machine learning techniques for multi-faceted modelling.

In contrast, our proposed approach addresses this gap by integrating knowledge graph construction, data fusion, and multi-faceted modelling into a unified framework, leveraging the strengths of both knowledge-based and data-driven methods for fault diagnosis.

Inspired by the limitations of existing methods, we propose a novel framework that combines the use of a knowledge graph and graph convolutional neural network (GCN) to model the multi-faceted aspects of equipment faults. Our approach builds on prior research in the areas of fault diagnosis, knowledge graphs, and graph neural networks. By leveraging these established techniques and addressing their shortcomings, our proposed framework represents a significant contribution to the field of equipment fault diagnosis. Specifically, our approach overcomes the issue of ignoring intra-aspects relationships, which is often present in other multi-faceted modelling techniques. Through experiments on a real-world dataset, we demonstrate that our proposed framework achieves improved accuracy in fault diagnosis compared to existing methods. Overall, our proposed framework represents an innovative and effective solution to the challenges of equipment fault diagnosis.

3. The proposed approach

As reviewed in the last section, it is necessary to generate a unified semantic knowledge-based approach as a foundation for the analysis of multi-faceted events, which can effectively associate intra-facets relations and take advantage of dynamic process data. In this case, we proposed the framework of this multi-faceted modelling approach shown in Fig. 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 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.

3.1. 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 Fig. 1.

Fault diagnosis data is typically characterised by multivariate time-series data that capture various aspects of equipment operation. The data can come from a variety of sources, including sensors, logs, and maintenance records. In this work, we focus on the task of fault diagnosis for a specific type of equipment, and we construct a knowledge graph to model the relationships between the different types of data.

To construct the knowledge graph, we follow a set of critical steps to ensure that we capture the relevant information and fuse it effectively. First, we pre-process the data to handle missing values and normalise the data to facilitate comparison across different sources. We then transform the multivariate time-series data into a temporal graph representation, where each node represents a concept, and each edge represents a relationship between concepts. The construction of the knowledge graph

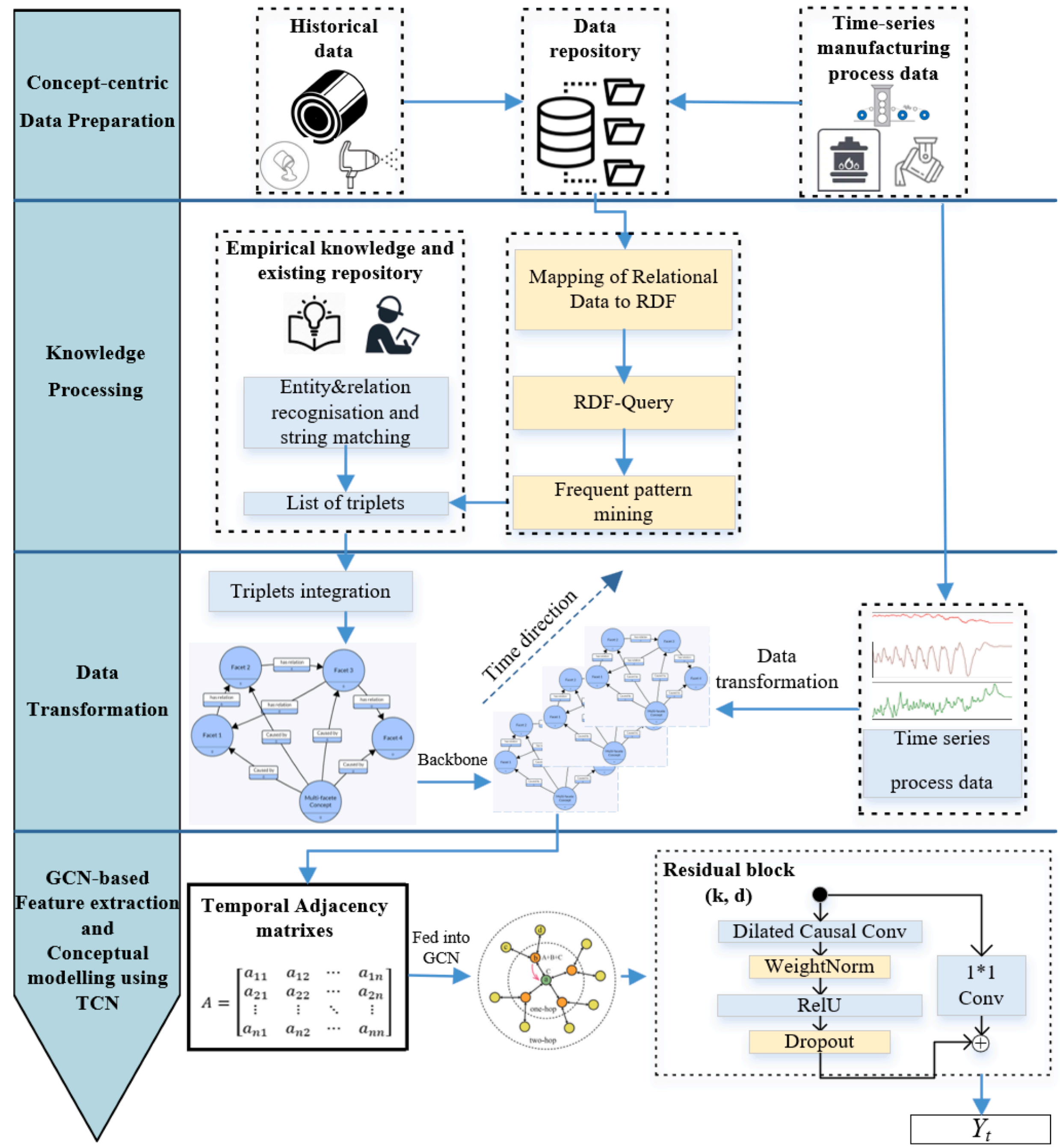

Fig. 1. The overall architecture of KG-supported multi-faceted conceptual modelling.

is based on data fusion issues, which requires the extraction of both structural and temporal features from the data.

To model the relationships between the concepts, we establish a set of generation rules based on the domain knowledge and the data itself. These rules are used to create the edges between the nodes in the knowledge graph, which captures the relationships between different concepts. For example, a rule may specify that if there is a certain combination of sensor readings, it is likely that the equipment is in a particular state.

The resulting knowledge graph is a concept-centric KG that is specifically designed to capture the relevant information for fault diagnosis. The KG is constructed by combining the multivariate time-series data with domain knowledge, which allows us to model the complex relationships between the different types of data. In the following sections, we will describe in more detail how we extract and process the structural and temporal features of the knowledge graph using GCN and TCN models.

3.1.1. Entities and relations extraction from concept-centric data and existing knowledge repository

Fault diagnosis data is typically characterised by multivariate time-series data that capture various aspects of equipment operation. The data can come from a variety of sources, including sensors, logs, and maintenance records. In this work, we focus on the task of fault diagnosis for a certain type of equipment, and we construct a knowledge graph to model the intra-relationships between the different facets.

The construction of the knowledge graph involved several critical steps to ensure the relevant information was captured and effectively fused. Initially, the data was pre-processed to handle missing values and normalise it to enable comparison across various sources. Subsequently, the multivariate time-series data was transformed into a temporal graph representation where each node represented a concept, and each edge indicated a relationship between concepts.

To construct the knowledge graph, structural and temporal features were extracted from the data. These features were obtained based on generation rules that were established using both domain knowledge and the data. The rules were utilised to create edges between the nodes, thus representing the relationships between different concepts. For example, a rule might have been employed to suggest that a particular state of equipment could be inferred from a particular combination of sensor readings.

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 [44] 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.

As shown in Fig. 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.

3.1.2. CKG generation by rules and node attributes integration

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…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 Fig. 2.

The generation of the fault KG involves several steps, including the definition of concepts and their attributes related to the fault diagnosis problem, the transformation of fault data into a structured format, the application of KG construction algorithms, and the enrichment of the KG with external knowledge sources. To better illustrate the generation process, we provide some examples of generation rules in this section. These rules are defined by domain experts and used to construct the fault KG. The outcome of the process was a concept-centric knowledge graph that was specifically designed to capture information relevant to fault diagnosis. The knowledge graph was constructed by combining multivariate time-series data with domain knowledge to enable the modelling of complex relationships between different types of data. In the subsequent sections, the extraction and processing of structural and temporal features of the knowledge graph using GCN and TCN models will be described in more detail.

3.2. Temporal graph-based data transformation

In a typical manufacturing process, each data instance carries a default time stamp, indicating that it belongs to a specific time period. To exploit the temporal characteristics of this data effectively, we aim to build a model that incorporates both the state characteristics of the time series and the internal relationships of the concept-centric features. 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

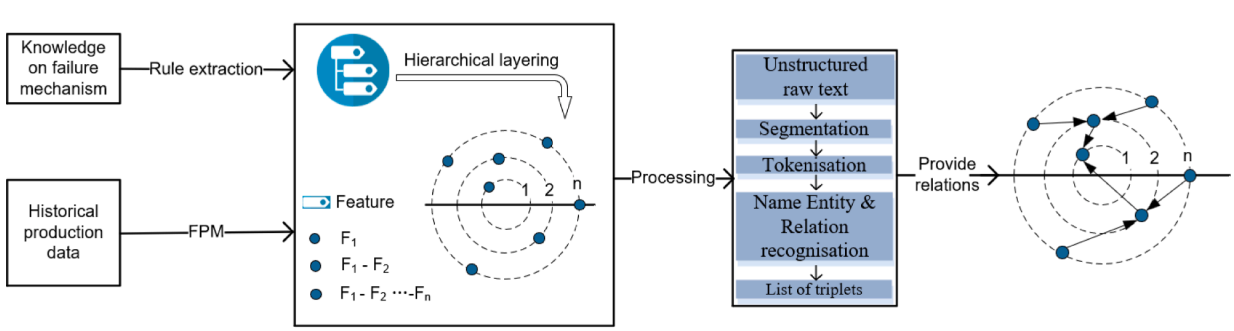

Fig. 2. Semantic illustration of CKG construction.

and their temporal dependency in the dynamic working system. In this context, the concept-centric knowledge graph can be expressed as follows:

$$G = (\mathrm{V},\ E,\ \mathrm{A}) \tag{1}$$

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 $X_\tau = (X_1,\ X_2,\ X_{t-\tau+1}\ldots,\ X_t) \in R^{t*N}$ 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 [45]. 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 $G_\tau = (G_1,\ G_2,\ G_{t-\tau+1}\ldots,\ G_t)$ at each time interval as shown in Fig. 3.

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

$$G^{\mathrm{t}} = (V^{\mathrm{t}},\ E^{\mathrm{t}}, \boldsymbol{A}^{\mathrm{t}}) \tag{2}$$

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:

$$E^{\mathrm{t}} = \left\{ e^t_{jk} \middle| \forall \mathrm{j},\ \mathrm{k} \in V^{\mathrm{t}} \right\} \in \mathbb{R}^{N\times N} \tag{3}$$

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

$$\boldsymbol{A}^{\mathrm{t}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{4}$$

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 amongst 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:

$$P_{X_1,\ X_2} = \frac{\sum_{i=1}^{n}(X_{1i} - \overline{X_1})(X_{2i} - \overline{X_2})}{\sqrt{\sum_{i=1}^{n}(X_{1i} - \overline{X_1})^2}\sqrt{\sum_{i=1}^{n}(X_{2i} - \overline{X_2})^2}} \tag{5}$$

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:

$$\mathbf{A}_{i,j} = e^{P_{X_1,\ X_2}} \tag{6}$$

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

3.3. CKG-supported multi-faceted conceptual modelling

This section outlines the process of multi-faceted concept modelling using the proposed knowledge graph-based data fusion framework. It describes the incorporation of the constructed knowledge graph into the modelling process and explains the steps involved in feature extraction and classification for fault diagnosis.

3.3.1. Temporal graph feature extraction using GCN

To extract meaningful features from the knowledge graph, we transform triples in the graph into their corresponding low-dimensional vector embeddings using a graph convolution model, in this case, the GCN. Firstly, an adjacency matrix $A \in \mathbb{R}^{N\times N}$ 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 multisource 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 multisource 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

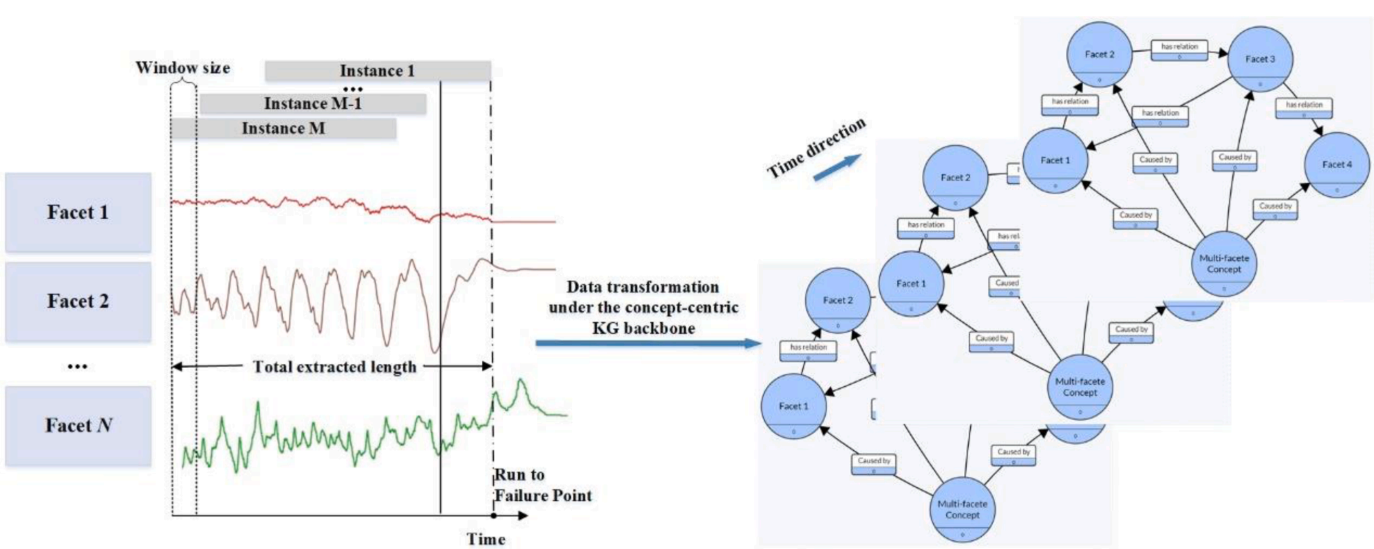

Fig. 3. Transformation and temporal graph representation of the run-to-failure process data.

Layer-wise propagation rule for a multilayer GCN:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{7}$$

Here, as there are no self-connections in the graph, $\widetilde{A} = A + I_N$ is defined as the adjacency matrix of the graph, amongst which I_N is the identity matrix. Moreover, a layer-specific trainable weight matrix is described as $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$ 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 $H^{(0)} = X$.

3.3.2. Conceptual modelling using TCN

After extracting the aggregated features from the temporal graphs using GCN, we further process these features through a Temporal Convolutional Network (TCN) for multi-faceted conceptual modelling. The TCN is optimised to model the temporal dependencies and further capture the evolving patterns in the data, ultimately aiding in more accurate and efficient fault diagnosis. 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, 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 CNNs perform 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 [46]. 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 [47]. Through autoregressive prediction and a long memory, the architecture is simple (e.g., no skip connections across layers, as shown in Fig. 4). In addition, it is capable of achieving both very deep networks because it uses dilated convolutions that allow the receptive field to be exponentially expanded.

As shown in Fig. 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 [47].

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. 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.

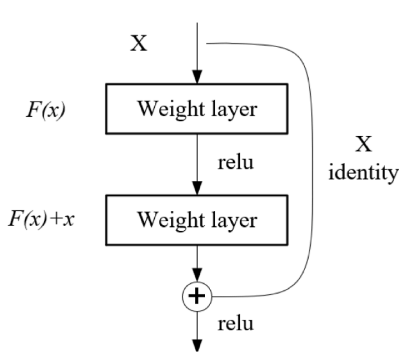

Fig. 4. Illustration of skip connections across layers.

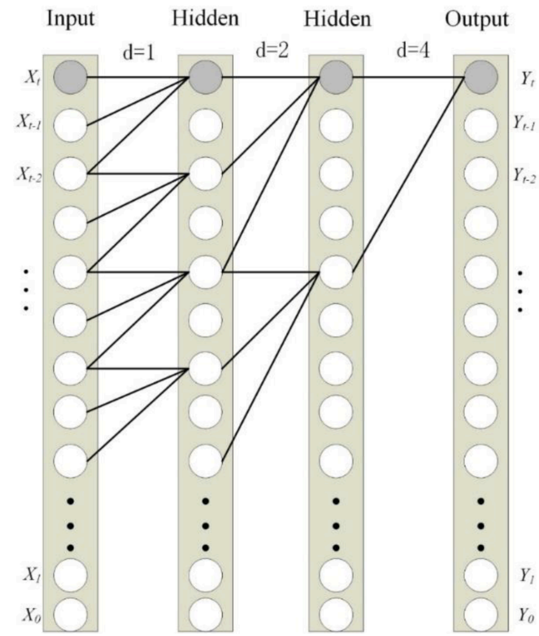

Fig. 5. Illustration of TCN structure.

Pseudocode for CKG-TCN multi-faceted modelling [48]
Input: - Multivariate time-series process data $X_\tau \in R^{t*N}$

- Correlations amongst t sequences $PCCs \leftarrow X_\tau$
- CKG backbone $G = (V, E, A)$

Output: CKG-TCN modelThe procedure of CKG-TCN Modelling:

1: Initialise the backbone Concept Knowledge Graph (CKG) $G = Initialise\_CKG()$
2: Convert time-series data into temporal graph based on CKG and compute adjacency matrix
3: ***For*** $t = 1$ to $Length(X_\tau)$ ***do***:
4: Compute the temporal graph $G[t] = Convert\_Data\_To\_Graph(X_\tau[t], G)$
5: ***For each*** measurement i,j in X_τ ***do***
6: Compute adjacency matrix $A[t] = Compute\_Adjacency\_Matrix(A[t])$
7: Calculate the weighted adjacency matrix $A[t]_{weighted} = e^{PCCs_{(x_i,x_j)}}$
8: ***End for***
9: Update the graph's edges $G[t] = Update\_Edges(G[t], A[t]_{weighted})$
10: ***End for***
11: Stack the $A[t]$ for all t in 1 to $Length(X_\tau)$ to form A_τ (3D tensor of stacked matrices).
11: Stack the matrices $V[t]$ representing vertices for all t in 1 to $Length(X_\tau)$ to form V_τ.
12: ***For each*** matrix M in $[A_\tau, V_\tau]$ ***do:***
13: Compute the normalisation of M.
14: ***End for***
15: Initialise the Graph Convolutional Network (GCN) model.
16: Apply the GCN to the normalised A_τ and V_τ to extract features F_τ
17: Initialise the Temporal Convolutional Network (TCN) model.
18: ***While*** the model's performance on a validation set has not yet converged ***do:***
18: Apply the TCN to F_τ for temporal feature extraction and prediction.
19: Update the model parameters.
20: Check if the model's performance on a validation set has converged.
21: ***End While***
22: ***Return*** the final trained CKG-TCN model.

Following the implementation of the multi-faceted modelling process, the procedure is encapsulated into a succinct pseudocode which presents a clear, step-by-step process on how the CKG-TCN model is built and trained. In summary, the CKG backbone is established with the identified concept-centric features. In a loop iterating through the length of the input data, every data instance at each time step is converted into a temporal graph based on the defined CKG schema. The adjacency matrix of each graph is computed, subsequently being transformed into a weighted adjacency matrix, which then updates the graph's edges. After completing the iteration through the input data, the matrices representing the vertices in each graph are stacked to form V_τ. The adjacency matrix A_τ and the vertices V_τ are then normalised, followed by the initialisation of the TCN model. The iterative process then commences, which involves the extraction of features from A_τ and V_τ using a GCN, followed by the application of the TCN to the features, thus producing F_τ that incorporates both intra-facet and temporal information.

4. Case study

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 [45], 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 [49]. When it comes to cold rolling, it is inevitable that failures such as edge cracks, central bursts, surface defects, and buckles will take place [50]. Strip breakages are amongst 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 [51]. A retrospective analysis of root causes has been conducted in previous studies on strip breakage [45], 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 [52]. 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.

4.1. Data and knowledge repository description

This section introduces the experimental setup used to evaluate the performance of our proposed method. We will describe the dataset, the implementation details, and the evaluation metrics employed to assess the effectiveness of our approach in the context of fault diagnosis.

4.1.1. Data description for breakage-centric multi-sourced data

To delve deeper into the complexity of strip breakage faults in the production process, we utilise a comprehensive real-world dataset obtained from a cold rolling mill to validate our proposed method. The cold rolling mill is a typical representative of a multi-stage process, where the strip's thickness is systematically reduced through several interconnected rolling stands. The dataset is multi-faceted, comprising multisource heterogeneous data from the cold-rolling process. This data includes time-series data reflecting a sequence of data points indexed in time order, event data pertaining to specific occurrences during the production process, and sequential data capturing the order of various operations. Specifically, the dataset contains time-series data collected from various sensors and equipment, including raw entry speed, shape AP front, total load feedback, tramp sap result, measured slip, and gauge average. The dataset utilised in this research is derived from multisource heterogeneous data that we have previously collected, analysed, and discussed in our prior work [45,53,54]. It encompasses various domains, thereby presenting a unique and rich array of information.

To be specific, the data was collected over a period of six months and included both normal and faulty operation data. The faulty operation data consists of four different types of faults: strip breakage, roll eccentricity, edge crack, and roll damage. Each fault type includes multiple instances, and the dataset is balanced with an equal number of normal and faulty instances. The 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. The dataset was chosen because it provides a rich source of information that allows us to demonstrate the effectiveness of our proposed method in handling multisource heterogeneous data and modelling complex relationships between attributes. Furthermore, the dataset has been previously used in other studies, which ensures the validity and reliability of the results obtained in our experiments.

It should be noted that each selected coil only broke once. Thus, the dataset contained 1256 coils, amongst which 354 are broken strip coils, covering three months of production. In order to pre-process the data, we have applied normalisation and scaling to ensure the values are comparable across different attributes. Additionally, we have identified the temporal dependencies between attributes and determined the appropriate window size for monitoring the most relevant data related to the system's condition. Furthermore, it is possible to calculate the breakage point in detail using full-resolution data, resulting in more accurate classification labels. Some of the collected features are shown in Table 1.

Table 1 presents the description of the six features used in this study for fault diagnosis. The first feature is Raw entry speed, which is the measurement of strip speed at the entry. It is an important parameter that affects the quality of the final product. The second feature, Shape AP Front, records the shape at the front end of the coil after annealing and pickling. The third feature is Total load feedback, which is the force equal to the pressure on a strip that pushes the load apart. This feature

Table 1
Example of typical features from the cold rolling process.

No.	Name	Description
1	Raw entry speed (m/min)	Measurement of strip speed at the entry
2	Shape AP Front	Anneal and pickling recorded the shape at the front end of the coil
3	Total load feedback (T)	A force equal to the pressure on a strip pushes the load apart
4	Tramp Sap Result	Oil Saponification value against incoming saponification value
5	Measured slip (%)	Displacement of a strip coil versus a working roll
6	Gauge average (mm)	The average measured gauge at digital signal processing

can provide information about the load-bearing capacity of the equipment. The fourth feature is Tramp Sap Result, which is the oil saponification value against the incoming saponification value. This feature can help detect the presence of contaminants in the system. The fifth feature, Measured slip, is the displacement of a strip coil versus a working roll. This feature can provide insights into the condition of the equipment and help detect the presence of defects. The final feature is the Gauge average, which is the average measured gauge at digital signal processing. This feature can help detect changes in the thickness of the strip and can provide valuable information for quality control.

To present a clearer picture of the dataset, we summarise the key characteristics of our dataset in Table 2. In general, the dataset for this study was collected from 1256 coils across the hot rolled coils (HRC), annealing and pickling (A&P), emulsion, and cold rolling process (CRP) phases. The HRC data records the physical and chemical properties of each incoming feedstock hot rolled coil across 47 variables. A&P data comprises 18 variables, capturing the real-time annealing and pickling process on each incoming hot rolled coil at a frequency of 50 Hz. The emulsion data, which is recorded daily, has eight variables. Overall, these features represent important parameters that can provide insights into the condition of the equipment and help detect faults. By incorporating these features into the knowledge graph-based framework proposed in this study, it is possible to achieve accurate fault diagnosis and improve the overall performance of the system.

4.1.2. Knowledge repositories surrounding strip breakage

In most cases, it is not easy to construct the backbone of KG under a domain environment without the collaboration of domain experts [55]. 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 Fig. 6.

Building upon this, we construct a comprehensive concept-centric Knowledge Graph (KG) for the cold-rolling mill. This KG encapsulates the inherent complexity of the multi-faceted features associated with strip breakage. More specifically, it captures a broad range of potential correlations and interactions amongst these features. The constructed KG includes different types of entities, such as time-series data, event data, and sequential data, and it illustrates their interconnections. For instance, it might portray how a change in a specific time-series variable (like speed) could potentially trigger a particular event (such as a system warning), ultimately leading to a sequence of operations indicative of a strip breakage fault. By mapping out these intricate relationships, the KG provides an accurate and holistic representation of the complex correlations between different types of data, facilitating more comprehensive and accurate fault diagnosis.

The primary contribution of the proposed multi-faceted modelling approach shown in Fig. 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 coloured nodes in the graph represent different facets that categorise the entities in the CKG backbone, and the edges between nodes indicate the relationships amongst them. One key observation from the graph is that there is a high degree of interconnectedness amongst 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.

Table 2
Example of typical features from the cold rolling process.

Data source	HRC	A&P	Emulsion	CRP
Representative features	Chemical content	Annealing temperature	Dirt result, pH	speed, tension
Number of Features	47	18	9	17
Type of Data	Batch/time series	Batch/time series	Batch	time series
Sampling frequency	Per coil	Per 0.01s	Per day	Per 0.01s

4.2. Evaluation metrics

In this section, we provide a detailed explanation of the evaluation metrics used in our experiments to assess the performance of the proposed fault diagnosis model. The following evaluation metrics were used: accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy measures the proportion of correct predictions amongst all predictions. It is calculated as the ratio of the number of correctly classified samples to the total number of samples in the test set. A higher accuracy indicates better performance. Precision measures the proportion of true positives amongst all positive predictions. It is calculated as the ratio of the number of true positives to the sum of true positives and false positives. Precision indicates how often the model correctly identifies positive samples. A higher precision indicates that the model produces fewer false positives. Recall measures the proportion of true positives amongst all actual positives. It is calculated as the ratio of the number of true positives to the sum of true positives and false negatives. Recall indicates how often the model correctly identifies actual positive samples. A higher recall indicates that the model produces fewer false negatives. F1-score is the harmonic mean of precision and recall. It is a balanced measure that takes both precision and recall into account. It is calculated as the ratio of the product of precision and recall to their sum. F1-score ranges from 0 to 1, with a higher value indicating better performance. AUC-ROC (Area Under the Receiver Operating Characteristic Curve) measures the ability of the model to distinguish between positive and negative samples. It is calculated as the area under the curve of the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate at different classification thresholds.

To evaluate the modelling performance, the metrics shown in Eqs. (8)–(11) are used in this experiment.

$$Accuracy = \frac{|f\_set \ \cap \ K|}{|\ K|} \tag{8}$$

$$Precision = \frac{|f\_set \ \cap \ K|}{|\ K|} \tag{9}$$

$$Recall = \frac{|f\_set \ \cap \ K|}{|f\_set|} \tag{10}$$

$$F1 \ = \frac{2 \cdot Precision \cdot Recall}{Precision \ + \ Recall} \tag{11}$$

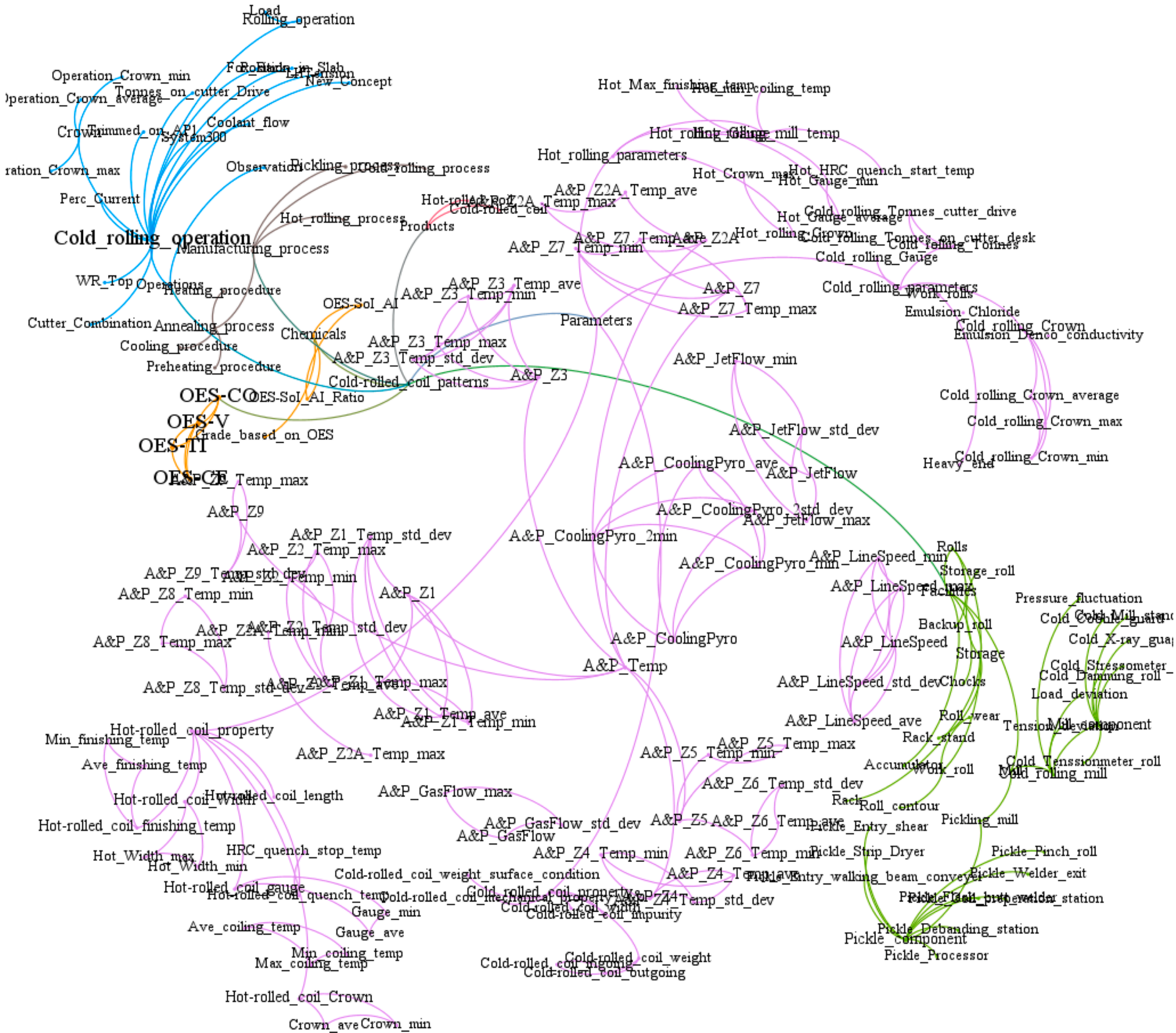

Fig. 6. The overall breakage-centric KG.

To better clarify, '*f_set*' refers to the set of all true positive predictions made by our model, which means the faults that were correctly identified by the model. '*K*', on the other hand, refers to the set of all positive predictions made by our model, including both the true positives (correct predictions) and false positives (incorrect predictions). 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.

4.3. Experimental setup on multi-faceted modelling

To elucidate our experimental setup in greater detail, we examined two different scenarios. The very first step in our process was understanding the nature of our data and the requirements of our research problem. It is sequential, time-dependant, and highly dimensional, with possible complex interactions. This recognition led to our decision to adopt models that can account for these characteristics, steering us towards exploring RNNs, and, later, the CKG-TCN model. Therefore, in the first scenario, our primary focus was to evaluate the efficacy of the CKG-TCN approach by benchmarking its performance against six other prevalent machine learning algorithms. These algorithms were Random Forest (RF), Support Vector Machine (SVM), k-Means, Long Short-term Memory (LSTM) network, Gated Recurrent Unit (GRU), and ConvLSTM. In detail, the number of estimators in the RF was fixed at 10, and the SVM algorithm utilised Gaussian kernels as the kernel function. For k-Means, we elected a cluster number of 3. The LSTM and GRU were configured with hidden state ht and window sizes of 5 and 16, respectively. The ConvLSTM, by default, was set with a kernel size of 20.

Additionally, transformer architecture [38] has gained significant attention in recent years due to its ability to model long-range dependencies and its parallelisation capabilities. In light of these developments, we include a Transformer baseline in our experiments to provide a comprehensive comparison with our proposed CKG-TCN method. A transformer baseline is also implemented, which has shown success in similar tasks. The Transformer model is adapted to our problem by configuring its input and output layers to match the dimensions of our dataset and by fine-tuning its hyperparameters to achieve optimal performance. We implemented the Transformer model and trained it on the same dataset to obtain the performance metrics for comparison. Besides the transformer, Temporal Graph Convolutional Networks (T-GCN) [56] and Graph Attention Networks (GAT) [57] have been implemented as the baseline as well. The T-GCN model combines both Graph Convolutional Networks (GCNs) and Gated Recurrent Units (GRUs) to model the spatial and temporal dependencies of graph-structured data. T-GCN is suitable for time-series prediction tasks on graph-structured data, making it a relevant baseline for the study. In this case, regarding our comparison of the CKG-TCN with the Transformer, it was a decision driven by an understanding that both these models leverage temporal information differently. Transformers excel at handling long-term dependencies but may overlook local interactions due to their global attention mechanism. CKG-TCN, on the other hand, is adept at capturing both local and distant interactions, hence offering a more comprehensive view of the data. The GAT model introduces an attention mechanism to the graph convolution operation, allowing nodes to weigh the contributions of their neighbours adaptively. GAT has been shown to achieve competitive performance in various graph-based tasks, making it a suitable SOTA baseline. 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.

In the second scenario, we attempted to confirm the advantages of graph-supported modelling. 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). Likewise, the decision to analyse the impact of KG-supported and non-KG-supported strategies came after understanding the potential of graph-aided models to better fit features and more accurately represent complex correlations. Extensive experiments were performed to verify our hypothesis, and the results helped reinforce our decision. 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.

4.4. Performance comparison

In this section, we provide an analysis of the results obtained from our experiments. We will discuss the performance of our proposed method compared to other state-of-the-art approaches, highlighting the advantages and limitations of our method in the context of fault diagnosis.

4.4.1. 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 sampling frequency is 100 Hz, a default window size is set to 0.01 s.

Fig. 7 illustrates a similar trend in all the metrics. It is obvious to see the best results are achieved when $\tau = 16$ in terms of all different metrics. The precision reaches the highest when $\tau = 16$. Accordingly, $\tau = 16$ 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.

4.4.2. Performance comparison with other prevailing machine learning algorithms

As stated in Section 4.4, 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.

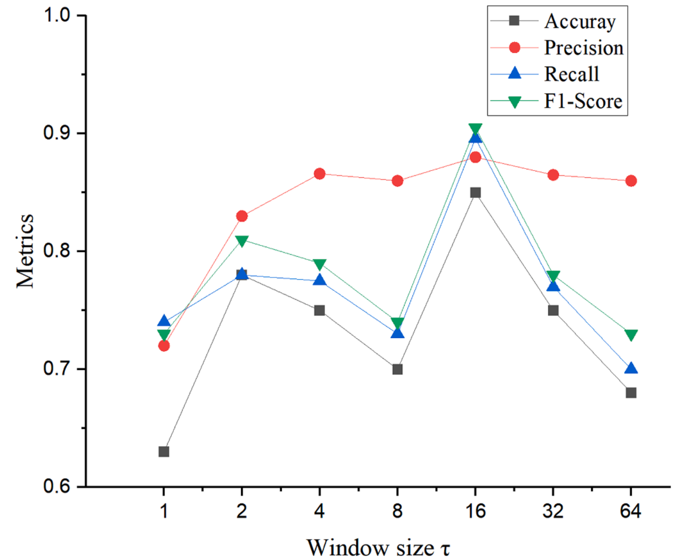

Fig. 7. Evaluating the performance of the TCN at different window sizes τ.

Table 3 exhibits the results of different models, combining RNN-based and traditional methods with varied feature sets, on their performance. These distinct methods' effectiveness is well illustrated by the insights the table offers. The significant enhancement of deep learning models based on RNNs in comparison to conventional techniques is mainly attributed to the default hyperparameter selection and distinct data representation methods utilised. Nonetheless, the degree of enhancement might fluctuate based on the particular assignment. Relying on their capability to take into account temporal information, RNN-based deep learning models were chosen as the preferred approach. Our dataset depends on this factor significantly. Our dataset's sequential and time-dependant nature is the reason for this. Therefore, the capacity of RNNs to maintain the recollection of preceding inputs proves to be highly advantageous. RFs were selected as they combine feeble classifiers to create more robust ones. Hence, they have higher chances of making precise predictions. To attain the desired performance, selecting hyperparameters becomes necessary due to the complexity of the model.

The CKG-TCN model, in particular, showed the highest performance amongst the tested models, as it leverages both temporal dependencies and attribute interactions, two key aspects in our dataset. The selection of this model was strategic as it offers several advantages, such as the integration of graph-aided data fusion and GCN-based feature

Table 3
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
Transformer-based	Transformer	**82.30**	86.20	86.50	0.86
Graph-based	T-GCN	78.12	84.87	85.96	0.85
	GAT	79.50	85.11	84.52	0.84
Graph-aided	CKG-TCN-noGCN	76.21	84.81	84.32	0.80
	CKG-TCN	81.20	**86.33**	**87.46**	**0.88**

extraction. This feature is crucial in our study as it helps provide a rich feature representation with less noise, something which our data critically required. 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. While our proposed CKG-TCN method demonstrates promising results, the Transformer baseline outperforms our approach in the conducted experiments. This indicates that the Transformer architecture is particularly effective in fusing features and modelling temporal information for this specific problem. However, our CKG-TCN method still offers several advantages, such as the integration of graph-aided data fusion and GCN-based feature extraction, which provide a rich feature representation with less noise.

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.

4.4.3. Comparison with different fusion strategies

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 $\overline{V_\tau} \in R^{\tau*N*N}$. 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 Figs. 8 and 9.

In Figs. 8 and 9, 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 amongst various attributes cannot be ignored. In contrast to conventional numerical fusion systems, the multisource 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.

4.5. Discussion

Based on validation, it was demonstrated that the proposed framework was capable of processing multisource heterogeneous data, and the proposed KG-aided fusion approach proved valuable for multi-faceted modelling tasks. Specifically, these experiments provide insight into the results of KG-aided data fusion for multi-faceted modelling. It is essential to determine the appropriate window size for monitoring the most relevant data related to the working system's condition as the first step. By employing the best temporal dependency model, it is possible to develop a reliable failure prediction model. The establishment of attribute dependencies is also crucial in order to achieve higher performance levels.

Moreover, their reliance on the operating system is a significant drawback of SVMs, K-Means, and k-Means. The current state and the past state are directly correlated. Hence, a sole medical record is

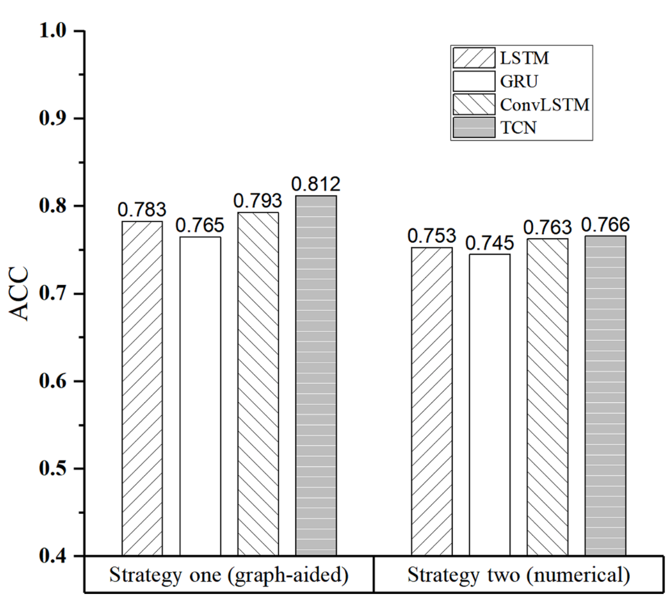

Fig. 8. Modelling performance in terms of ACCs in different strategies.

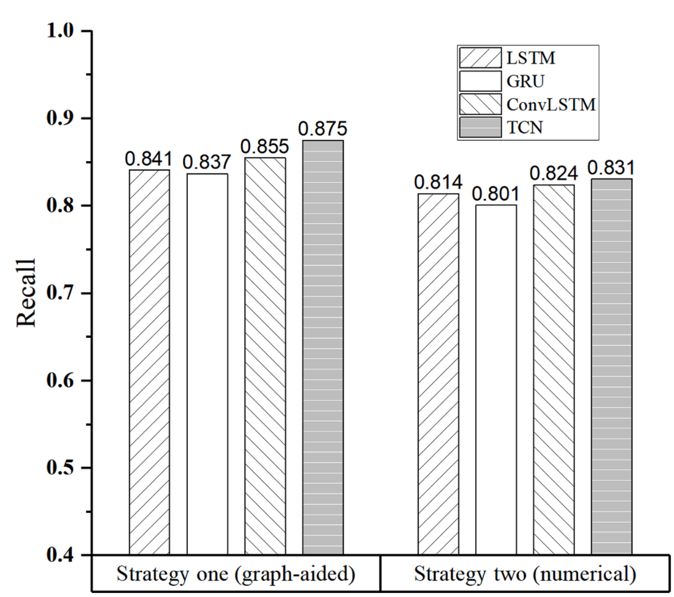

Fig. 9. Modelling performance in terms of Recalls in different strategies.

inadequate in furnishing a thorough evaluation of one's health condition. Despite their capabilities in extracting temporal characteristics, RNNs are not designed to consider the relationships between features themselves. Fully utilising the temporal dependency can result in the creation of a dependable concept modelling model. Acquiring a more profound comprehension of the concept will be facilitated by this. The RNN-CNN-based model, in the meantime, can seize the temporal features of the process data. In contrast to RNN-based models, its convolution layer allows for the extraction of abundant information from it. Our proposed CKG-TCN method does not match the accuracy levels demonstrated by the Transformer architecture. The self-attention mechanism utilised in the transformer can be credited for this. By dynamically weighing the input features, the mechanism is proficient in capturing long-range dependencies in the data. This allows the model to focus better on the most relevant features, leading to improved accuracy in fault prediction. However, our CKG-TCN method outperforms the transformer in terms of precision, recall, and F1 score. This can be explained by the fact that the CKG-TCN method specifically leverages the knowledge graph-aided fusion approach, which captures both temporal feature embedding and attributes relationship embedding. This enables the model to effectively handle the multisource heterogeneous data and model the complex relationships between the attributes, leading to better performance in distinguishing true positive cases from false positives and false negatives. As a result, the CKG-TCN method achieves higher precision, recall, and F1 scores, which are important metrics for assessing the overall quality of fault prediction. These findings highlight the advantages of our approach in handling multisource heterogeneous data and utilising knowledge graph-aided fusion for multi-faceted modelling tasks. They also suggest that there is room for further research to explore the integration of the strengths of the Transformer architecture, such as its self-attention mechanism, into our proposed framework to enhance its overall performance.

Thirdly, a comparison of graph-aided and conventional data fusion experiments reveals that there are possible 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. Specifically, since the GCN is utilised for fitting temporal and attribute relationship feature representations, it is imperative that graph features are extracted by the GCN. By using GCN extraction and aggregation, a node can aggregate more information without experiencing excessive noise. Through the combination of the features learned by each layer of the GCN with the characteristics of the node, CKG-TCN can provide a feature representation with rich information and less noise. Therefore, CKG-TCN-noGCN may produce less noise in feature representations than CKG-TCN-noGCN.

Lastly, as shown in this case study, the proposed framework for fault diagnosis was applied to a real-world cold-rolling mill to demonstrate its effectiveness and applicability in practical settings. The mill was a multi-stage process consisting of several rolling stands, where the thickness of the strip was gradually reduced. Beyond our previous work [45], the proposed framework leveraged a knowledge graph constructed from the data to model the relationships between different types of data and diagnose strip breakage faults. Specifically, when there is a fault frequently occurs in the production process, the proposed method not only improves the accuracy and convergence speed of fault diagnosis but also enables constructing a domain map of equipment fault diagnosis by combining mechanism knowledge and data-driven methods of multi-faceted conceptual modelling.

For future work, 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. In addition, the experiments show that the Transformer baseline achieves better performance than our proposed CKG-TCN method. This highlights the importance of further refining our method to better capture the inherent characteristics of the problem. Potential improvements could include exploring alternative graph representation learning techniques, incorporating attention mechanisms from Transformer models into the CKG-TCN framework, or adapting the CKG-TCN architecture for better feature fusion and temporal modelling. Additionally, investigating the combination of our method with the Transformer architecture may lead to a more effective approach for multi-faceted modelling. Furthermore, it would be beneficial to explore a more precise adjacency matrix and a more effective spatial-temporal structure.

5. Conclusions

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 ML models for fault concept modelling. The experimental results show: (1) the KG-aided fusion strategy outperforms 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.

CRediT authorship contribution statement

Zheyuan Chen: Conceptualization, Methodology, Writing – original draft, Validation. **Yuwei Wan:** Methodology, Writing – original draft, Validation. **Ying Liu:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Agustin Valera-Medina:** Writing – review & editing, Validation.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author, Prof Ying Liu, is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). Prof Liu is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from LiuY81@cardiff.ac.uk.

Data availability

The authors do not have permission to share data.

Acknowledgement

None.

References

[1] S.C.J. Lim, Y. Liu, W.B. Lee, A methodology for building a semantically annotated multi-faceted ontology for product family modelling, Adv. Eng. Inf. 25 (2) (2011) 147–161.
[2] H. Fang, G. Guo, J. Zhang, Multi-faceted trust and distrust prediction for recommender systems, Decis. Support Syst. 71 (2015) 37–47.
[3] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, M.M. Hoffman, Machine learning for integrating data in biology and medicine: principles, practice, and opportunities, Inf. Fusion 50 (2019) 71–91.
[4] W. Ding, X. Jing, Z. Yan, L.T. Yang, A survey on data fusion in internet of things: towards secure and privacy-preserving fusion, Inf. Fusion 51 (2019) 129–144.
[5] M. Muzammal, R. Talat, A.H. Sodhro, S. Pirbhulal, A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks, Inf. Fusion 53 (2020) 155–164.
[6] S. Stillman, I. Essa, Towards reliable multimodal sensing in aware environments, in: Proceedings of the Workshop on Perceptive User Interfaces, 2001, pp. 1–6.
[7] B. Zhou, et al., Semantic-aware event link reasoning over industrial knowledge graph embedding time series data, Int. J. Prod. Res. 61 (12) (2023) 4117–4134.
[8] Y. Lu, X. Xu, A semantic web-based framework for service composition in a cloud manufacturing environment, J. Manuf. Syst. 42 (2017) 69–81.
[9] P. Zheng, L. Xia, C. Li, X. Li, B. Liu, Towards Self-X cognitive manufacturing network: an industrial knowledge graph-based multi-agent reinforcement learning approach, J. Manuf. Syst. 61 (2021) 16–26.
[10] X. Li, M. Lyu, Z. Wang, C.H. Chen, P. Zheng, Exploiting knowledge graphs in industrial products and services: a survey of key aspects, challenges, and future perspectives, Comput. Ind. 129 (2021), 103449.
[11] Q. Li, Z. Cao, M. Tanveer, H.M. Pandey, C. Wang, A semantic collaboration method based on uniform knowledge graph, IEEE Internet Things J. 7 (5) (2019) 4473–4484.
[12] B. Abu-Salih, et al., Relational learning analysis of social politics using knowledge graph embedding, Data Min. Knowl. Discov. 35 (4) (2021) 1497–1536.
[13] M. Kejriwal, Domain-Specific Knowledge Graph Construction, Springer, 2019.
[14] L. Xia, P. Zheng, X. Li, R.X. Gao, L. Wang, Toward cognitive predictive maintenance: a survey of graph-based approaches, J. Manuf. Syst. 64 (2022) 107–120.
[15] S. Ji, S. Pan, E. Cambria, P. Marttinen, S.Y. Philip, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE transactions on neural networks and learning systems 33 (2) (2021) 494–514.
[16] N.F. Noy and D.L. McGuinness, "Ontology development 101: a guide to creating your first ontology," ed: Stanford knowledge systems laboratory technical report KSL-01-05 and …, 2001.
[17] C. Choi, M. Cho, J. Choi, M. Hwang, J. Park, P. Kim, Travel ontology for intelligent recommendation system, in: Proceedings of the Third Asia International Conference on Modelling & Simulation, IEEE, 2009, pp. 637–642.
[18] Y. Liu, S.C.J. Lim, W.B. Lee, Product family design through ontology-based faceted component analysis, selection, and optimization, J. Mech. Des. 135 (8) (2013), 081007.
[19] S.C.J. Lim, Y. Liu, W.B. Lee, Faceted search and retrieval based on semantically annotated product family ontology, in: Proceedings of the WSDM'09 Workshop on Exploiting Semantic Annotations in Information Retrieval, 2009, pp. 15–24.
[20] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, P. Merialdo, Knowledge graph embedding for link prediction: a comparative analysis, ACM Trans. Knowl. Discov. Data (TKDD) 15 (2) (2021) 1–49.
[21] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, S. Yang, Community preserving network embedding, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[22] B. Abu-Salih, Domain-specific knowledge graphs: a survey, J. Netw. Comput. Appl. 185 (2021), 103076.
[23] W. Wang, M. Zhang, Tensor deep learning model for heterogeneous data fusion in Internet of Things, IEEE Trans. Emerg. Top. Comput. Intell. 4 (1) (2018) 32–41.
[24] S. Jabbar, et al., A methodology of real-time data fusion for localized big data analytics, IEEE Access 6 (2018) 24510–24520.
[25] X. Li, C.H. Chen, P. Zheng, Z. Wang, Z. Jiang, Z. Jiang, A knowledge graph-aided concept–knowledge approach for evolutionary smart product–service system development, J. Mech. Des. 142 (10) (2020), 101403.
[26] K.K. Ng, C.H. Chen, C.K. Lee, J.R. Jiao, Z.X. Yang, A systematic literature review on intelligent automation: aligning concepts from theory, practice, and future perspectives, Adv. Eng. Inf. 47 (2021), 101246.
[27] M. Wang, Y. Li, Y. Zhang, L. Jia, Spatio-temporal graph convolutional neural network for remaining useful life estimation of aircraft engines, Aerosp. Syst. 4 (1) (2021) 29–36.
[28] Y. Zhang, Y. Li, X. Wei, L. Jia, Adaptive spatio-temporal graph convolutional neural network for remaining useful life estimation, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–7.
[29] N. Xie, W. Wang, B. Ma, X. Zhang, W. Sun, F. Guo, Research on an agricultural knowledge fusion method for big data, Data Sci. J.l 14 (7) (2015) 1–9.
[30] X. Xiao, C. Li, J. Huang, T. Yu, Fault diagnosis of rolling bearing based on knowledge graph with data accumulation strategy, IEEE Sens. J. 22 (19) (2022) 18831–18840.
[31] Z. Li, Y. Li, Q. Sun, B. Qi, Bearing fault diagnosis method based on convolutional neural network and knowledge graph, Entropy 24 (11) (2022) 1589.
[32] C. Zhang, J. Yu, S. Wang, Fault detection and recognition of multivariate process based on feature learning of one-dimensional convolutional neural network and stacked denoised autoencoder, Int. J. Prod. Res. 59 (8) (2021) 2426–2449.
[33] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (7) (2018) 5525–5534.
[34] A. Holzinger, B. Ofner, M. Dehmer, Multi-touch graph-based interaction for knowledge discovery on mobile devices: state-of-the-art and future challenges. Interactive Knowledge Discovery and Data Mining in Biomedical Informatics, Springer, 2014, pp. 241–254.
[35] Y. Zheng, Methodologies for cross-domain data fusion: an overview, IEEE Trans. Big Data 1 (1) (2015) 16–34.
[36] Y. Oulabi, R. Meusel, C. Bizer, Fusing time-dependent web table data, in: Proceedings of the 19th International Workshop on Web and Databases, 2016, pp. 1–7.
[37] S. Bahrami, A. Bosaghzadeh, F. Dornaika, Multi similarity metric fusion in graph-based semi-supervised learning, Computation 7 (1) (2019) 15.
[38] X. Wang, L. Zhu, Z. Zheng, M. Xu, Y. Yang, Align and tell: boosting text-video retrieval with local alignment and fine-grained supervision, IEEE Trans. Multimed. 14 (8) (2022) 1–11.
[39] H.L. Nguyen, D.T. Vu, J.J. Jung, Knowledge graph fusion for smart systems: a survey, Inf. Fusion 61 (2020) 56–70.
[40] J. Chen, A. Zhang, HGMF: heterogeneous graph-based fusion for multimodal data with incompleteness, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 1295–1305.
[41] L. Zhang, et al., Deep fusion of brain structure-function in mild cognitive impairment, Med. Image Anal. 72 (2021), 102082.
[42] R.M. Damhuis, P.L. Roux, C.J. Fourie, The identification and mitigation of geohazards using shallow airborne engineering geophysics and land-based geophysics for brown-and greenfield road investigations, Q. J. Eng. Geol. Hydrogeol. 53 (2) (2020) 321–332.
[43] S. Chen, J. Wang, H. Li, Z. Wang, F. Liu, S. Li, Top-down human-cyber-physical data fusion based on reinforcement learning, IEEE Access 8 (2020) 134233–134245.
[44] O. Corcho, F. Priyatna, D. Chaves-Fraga, Towards a new generation of ontology based data access, Semant. Web 11 (1) (2020) 153–160.
[45] Z. Chen, Y. Liu, A. Valera-Medina, F. Robinson, M. Packianather, Multi-faceted modelling for strip breakage in cold rolling using machine learning, Int. J. Prod. Res. 59 (21) (2021) 6347–6360.
[46] C. Lea, M.D. Flynn, R. Vidal, A. Reiter, G.D. Hager, Temporal convolutional networks for action segmentation and detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 156–165.
[47] S. Bai, J.Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint ,* 2018.
[48] Y. Jiang, P. Dai, P. Fang, R.Y. Zhong, X. Cao, Electrical-STGCN: an electrical spatio-temporal graph convolutional network for intelligent predictive maintenance, IEEE Trans. Ind. Inf. 18 (12) (2022) 8509–8518.
[49] F. Hou, J. Zhang, J. Cao, X. Shi, Review of chatter studies in cold rolling, Gangtie Yanjiu Xuebao(J. Iron Steel Res.) 19 (10) (2007) 6–10.
[50] M. Mashayekhi, N. Torabian, M. Poursina, Continuum damage mechanics analysis of strip tearing in a tandem cold rolling process, Simul. Model. Pract. Theory 19 (2) (2011) 612–625.
[51] W. Johnson, A. Mamalis, A survey of some physical defects arising in metal working processes, in: Proceedings of the Seventeenth International Machine Tool Design and Research Conference, Springer, 1977, pp. 607–621.
[52] K. Takami, J. Mahmoudi, E. Dahlquist, M. Lindenmo, Multivariable data analysis of a cold rolling control system to minimise defects, Int. J. Adv. Manuf. Technol. 54 (5) (2011) 553–565.
[53] Z. Chen, Y. Liu, A. Valera-Medina, F. Robinson, Multi-sourced modelling for strip breakage using knowledge graph embeddings, Procedia CIRP 104 (2021) 1884–1889.
[54] Y. Wan, Z. Chen, F. Hu, Y. Liu, M. Packianather, R. Wang, Exploiting knowledge graph for multi-faceted conceptual modelling using GCN, Procedia Comput. Sci. 200 (2022) 1174–1183.
[55] Z. Jin, Environment Modeling-Based Requirements Engineering For Software Intensive Systems, Morgan Kaufmann, 2018.
[56] L. Zhao, et al., T-GCN: a temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. 21 (9) (2019) 3848–3858.
[57] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint ,* 2017.