

# A survey on graph neural networks, machine learning and deep learning techniques for time series applications in industry

Muhammad Jamal Ahmed, Alberto Mozo and Amit Karamchandani

E.T.S. de Ingeniería de Sistemas Informáticos, Universidad Politécnica de Madrid, Madrid, Spain

## ABSTRACT

Extensive studies have been conducted to investigate Artificial Intelligence (AI) in the context of time series data. In this article, we investigate the complex domain of industrial time series, from the dimensions of classical machine learning (ML), deep neural networks (DNNs) and graph neural networks (GNNs). Current surveys often focus on a specific methodology or oversee the connection of diverse approaches; our article bridges this gap by providing an all-inclusive interpretation across numerous techniques. In addition, the aim of this article is to focus on the core areas of time series such as forecasting, classification, and anomaly detection. From traditional methodologies like Autoregressive Integrated Moving Average (ARIMA) and support vector machine (SVM) methods, the advancements of DNNs, for instance long-short-term memory (LSTMs), convolutional neural networks (CNNs), attention mechanisms, and transformers, describe how temporal information is used for forecasting, anomaly detection, and classification. Then the article discusses the advances and limitations in ML, DNN, and GNN in order to improve the different methods in either category. Lastly, we outline future directions and open research questions with the different methodologies used in time series.

**Subjects** Artificial Intelligence, Data Mining and Machine Learning, Data Science

**Keywords** Time series, Industrial, Graph neural networks, Machine learning, Deep learning, Neural networks, Temporal information, Anomaly detection, Time series classification, Forecasting

## INTRODUCTION

The advent of industry and the rapid expansion of today's industry 4.0 contribute to the increased complexity and automation of industrial processes. Although industrial processes on a large scale bring significant economic benefits to the country, the high level of interconnectivity and complexity within the structure means that even a minor disruption can result in serious fiscal setbacks and possibly loss of life. Therefore, to guarantee safe and reliable implementation of the manufacturing idea, fault prediction, detection, and diagnostic technology are crucial for all planned critical industrial productions. In *Palit & Popovic (2006)*, we explore the evolving network of use cases like autonomous robots, system integration, the Internet of Things, simulations, additive manufacturing, cloud computing, augmented reality, big data, and cyber security that outline the industrial quarter. As our environment becomes increasingly interconnected, the synthesis of digital know-how with conventional industrial methods has led to a

Submitted 10 July 2024  
Accepted 10 July 2025  
Published 25 September 2025

Corresponding author  
Muhammad Jamal Ahmed,  
muhammadjamal.a@upm.es

Academic editor  
Claudio Ardagna

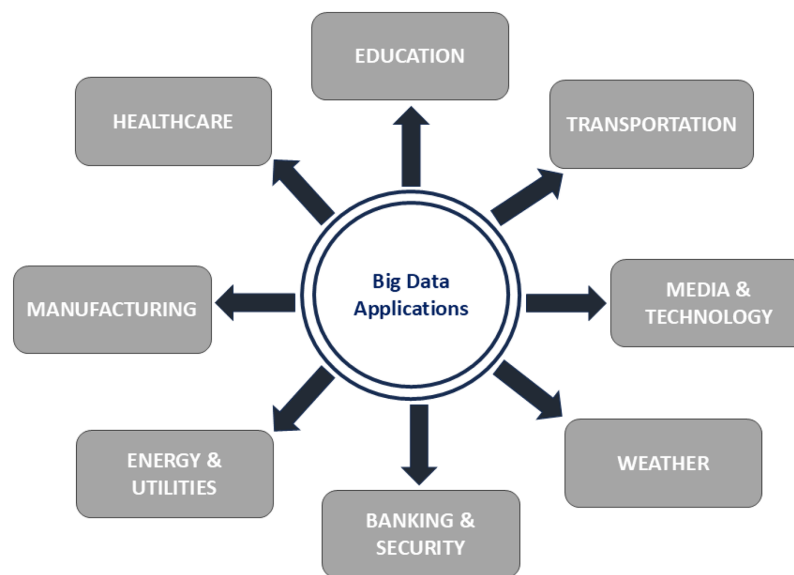
Additional Information and  
Declarations can be found on  
page 40

DOI 10.7717/peerj-cs.3097

© Copyright  
2025 Ahmed et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**



**Figure 1** Applications of big data in industrial sector. Full-size [DOI: 10.7717/peerj-cs.3097/fig-1](https://doi.org/10.7717/peerj-cs.3097/fig-1)

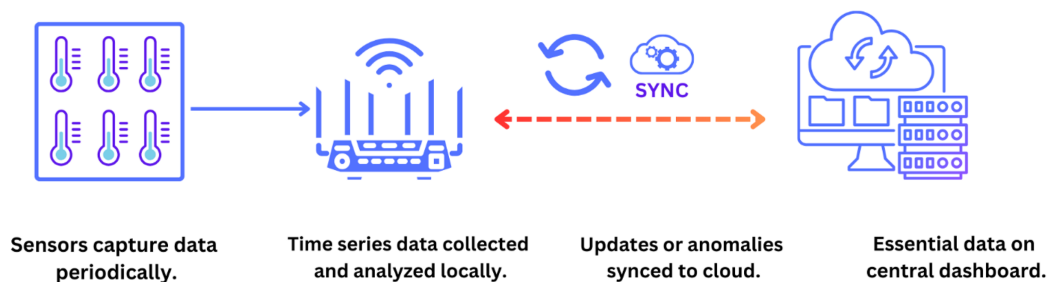
standard shift. [Figure 1](#) condenses the diverse range of applications within Industry 4.0, where data analytics, smart systems, and automation converge to transform manufacturing technology and beyond.

The central driving force behind this advanced technological progress is the transformative power of big data. [Figure 1](#) serves as a comprehensive guide to the numerous applications of big data within the industry 4.0, showcasing how data-driven insights are revolutionizing the global manufacturing sector. The integration of big data extends from predicting maintenance and optimizing supply chains to providing instant analytics and facilitating intelligent decision making. This incorporation of big data is pushing the boundaries of what is achievable in the industrial sector.

In the context of industrial environments, temporal analysis has become crucial for assessing operational reliability and efficiency and ensuring the safety and longevity of critical infrastructure. With the influx of temporal data from operational monitoring processes, the industrial sector is challenged to develop and integrate analytical approaches to effectively leverage this information to improve decision making and optimize its operational processes ([Bertolini et al., 2021](#)). [Figure 2](#) shows a multistage approach to time series data analysis, where data is generated at the sensor level, undergoes initial processing and analysis, and is then transferred to the cloud to improve its accessibility and scalability.

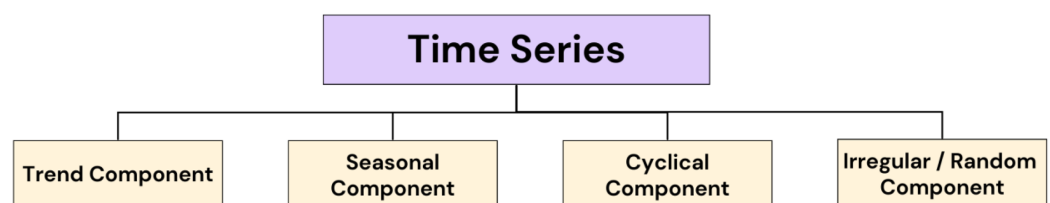
[Figure 3](#) illustrates the core components of the time series, each playing its own role in unraveling the anonymity of temporal drifts. From the all-encompassing course of trends to the periodic patterns of periods, to the surging sequences of cyclical effects, to the impulsive irregularities that add to the complexity, this figure guides the important workings that form the backbone of time series.

Classical machine learning, deep learning, and graph neural networks are established as standards for interpreting diverse patterns deeply rooted in time series data in industrial



**Figure 2** The above figure depicts a setup of time series starting from data collection to storing the data to be analysed.

Full-size DOI: 10.7717/peerj-cs.3097/fig-2



**Figure 3** Components that constitute the backbone of time series.

Full-size DOI: 10.7717/peerj-cs.3097/fig-3

settings. In industrial settings, the origins of time series analysis can be traced back to the revolutionary efforts of [Box et al. \(2015\)](#), who introduced the Autoregressive Integrated Moving Average (ARIMA) methods. These statistical techniques patented an elementary landmark in time series forecasting, providing the industrial sector with a robust structure to forecast demand, optimize production schedules, and maintain inventories. The advent of ARIMA and similar approaches represented a paradigm shift, allowing industry 4.0 to use historical data for informed policy making. Classical machine learning approaches, apparently signified by ARIMA, demonstrated an essential role in founding the early architectures for analyzing time series. These techniques, rooted in statistical fundamentals, propose real-world insights, and the interpretability that compiles them is especially valued for industries that require a vivid understanding of primary dynamics. In addition, traditional methods such as ARIMA have been shown to be effective in setups where data have strong temporal shapes and can be represented by means of nonlinear or linear associations ([Makridakis & Hibon, 2000](#)).

Recent advances in conventional machine learning practices have supported the incorporation of ensemble approaches for the analysis of time-series data. In particular, gradient-boosting machines and random forests have gained prominence. When dealing with nonlinear relationships and complex dependencies, ensemble models ([Karamchandani et al., 2023](#)) have seen progress within time series. By assembling predictions from different methods, these practices increase predictive capabilities and enhance robustness, making them convenient for various industrial frameworks ([Mozo et al., 2024](#)). Furthermore, ensemble models, conventional machine learning techniques, have progressed to integrate feature engineering methods, licensing improved illustration

of temporal dependencies and patterns. Feature engineering, coupled with typical processes such as support vector machine (SVM) and k-Nearest Neighbors (k-NN), has led to accurate improvements in predicting time-series tasks. These methods illustrate the adaptability of classical machine learning practices to face the shade limitations produced by practical industrial sector data in time series contexts. Although classical machine learning has made significant progress in the time series domain, it is important to acknowledge its limitations. The insights and interpretation in the real world offered by these techniques come at the cost of making inferences about the linearity and stationarity of the data (*Iglesias et al., 2023*). In situations where time series have complicated non-linear shapes or experience structural variations, typical machine learning approaches may fade.

Inscribing these boundaries requires a subtle method that fuses the power of conventional methods with evolving practices, certifying an adaptive and comprehensive framework for analyzing industrial time-series data. The field of time-series forecasting has undergone a groundbreaking metamorphosis with the arrival of deep learning techniques, with neural networks at the forefront. *Graves (2014)* presented an influential complement to this change with the construction of long short-term memories (LSTMs). These structures were designed to address the shortcomings of conventional techniques and to capture complicated temporal dependencies by allowing the time-series model to capture long-term dependencies in the data. The introduction of LSTM techniques has opened new avenues for dynamic and accurate prediction of time-series data, representing an innovation in the field of industrial environments. An essential milestone in the direction of deep learning for the prediction and analysis of time series data was manifested by the research of *Wang et al. (2018)*, who proposed sequence-to-sequence methods. This novelty was licensed to industries to predict failure detection and optimize complicated supply chain processes. Sequence-to-sequence methods had the power to channel input sequences and produce their output sequences, transforming predictive analytics and enlightening further understanding and comprehension of temporal dependencies and patterns. The metamorphic power of deep learning, as demonstrated by these developments, has made it an essential implement for addressing encounters modeled by practical time series data in industries (*Cutler et al., 2007*).

Current research by *Yang et al. (2021)* has explored attention architectures within deep learning methods for time series analysis. These mechanisms increase interpretability by allowing the technique to focus on explicit parts of the input sequence, thereby gaining an understanding of the aspects that drive the prediction. Approaches such as transfer learning, in which pre-trained methods are revised for precise industrial setups and adhere successfully to *Schölkopf et al. (2001)*. This method influences information from one domain to boost production in the other, an advantageous approach in situations where there is a limited amount of labeled data. Nevertheless, it is critical to admit the trials associated with deep learning, together with the need for considerable computational resources, the amount of labeled data, and possible overfitting in high-dimensional, multifaceted datasets.

Anomaly detection is a key player in maintaining the safety and reliability of industrial settings, certifying the early detection of faults before they intensify into serious problems. Traditional techniques, such as SVM and isolation forests of [Liu, Ting & Zhou \(2012\)](#), have been the followers in this field, contributing active solutions for the recognition of eccentricities of usual behavior. However, the field has undergone a transformative change with the incorporation of deep learning practices, notably variational autoencoders (VAEs) ([Kingma & Welling, 2022](#)). These techniques have played and continue to play a critical role in improving anomaly detection capabilities in time series data for industrial frameworks. VAEs could learn the probabilistic distribution of standard data and, therefore, it can successfully classify subtle deviations that might indicate anomalous behavior. The current stream of anomaly detection studies has validated the use of generative adversarial networks (GANs) in this area, as verified by [Akçay, Atapour-Abarghouei & Breckon \(2019\)](#), embodies the ability to integrate different deep learning practices to achieve more robust anomaly detection models. Furthermore, the integration of deep learning and attention mechanisms has shown potential in anomaly detection. A recent study by [Vaswani et al. \(2017\)](#) discovers the incorporation of attention approaches into time series, highlighting their role in illuminated anomaly detection performance in industrial settings.

[Li et al. \(2018\)](#) comes up with significant arguments by revealing the success of graph neural networks (GNNs) in the domain of time-series forecasting, exclusively in situations where time series data in an industrial context have characteristic graphical assemblies. GNNs go beyond traditional machine learning (ML) and thus captures these associations, which provides a more comprehensive understanding of the hidden dynamics particularly valuable in industries. A study by [Zhang et al. \(2022a\)](#) explores the integration of reinforcement learning with GNN, offering a pioneering exploration of their collective potential. Reinforcement learning, coupled with the ability of GNNs to handle complicated dependencies, generates a coactive method that embraces potential for augmenting decision approaches in response to the dynamic environments of industrial sectors. GNNs can successfully capture anomalous patterns that are evident in interconnected structures, giving a robust result for detecting deviations from normal behavior. The research in [Zhang et al. \(2022a\)](#) illustrates the positive incorporation of GNNs in the detection of anomalies, demonstrating the applicability of GNNs in several domains, taking into account the industrial context. Although GNNs offer exceptional capabilities, challenges remain in terms of interpretability and scalability. Addressing these challenges is crucial to certifying the unified incorporation of GNNs into the frameworks of industrial sectors ([De Livera, Hyndman & Snyder, 2011](#)). However, with the change in criteria towards neural networks, interpretability has emerged as a major concern in the industrial field. [De Livera, Hyndman & Snyder \(2011\)](#) highlighted the importance of interpretability in these networks and identified its implications for trusting and understanding the conclusions drawn by these approaches in crucial industrial situations.

The overview of attentional structures by [Vaswani et al. \(2017\)](#) proved to be an essential development, allowing practices to focus on serious time sections within time series. This mechanism not only improved overall accuracy, but also demonstrated good

interpretability for classification practices, confirming that important information is not ignored. The evolution of classification practices, as observed in the studies of [Taheri, Gimpel & Berger-Wolf \(2019\)](#), indicates a constant determination to make even these methods compatible with the subtle requirements of classification of industrial activity. In pursuit of filtering classification duties for time series data in industrial frameworks, current exploration by [Li, Wu & Xu \(2022\)](#) discovers the life-changing potential success of transformer-based structures. The use of these architectures signifies a retreat from classical recurrent neural networks and presents a new viewpoint on the classification of different sequences in the industrial background. The exploration of transformer-based techniques highlights the adaptability and compliance of neural networks and their volume to evolve in aggregation with the exclusive challenges modeled by time series data in Industry 4.0. As industries more and more hold the incorporation of different practices, from traditional SVMs to front-line transformer techniques, the survey will steer precisely through apiece standard. The key aspects of our survey are as follows.

- **Interdisciplinary synthesis:** By linking conventional machine learning approaches, cutting-edge deep learning frameworks, and evolving graph neural network practices, this review forms an interdisciplinary description. This combination enables the reader to grasp the evolution of analyzing time-series data across different technological domains, providing a broader understanding.
- **Comprehensive exploration:** This survey not only explores the theoretical underpinnings of each architype but also offers practical intuitions into their real-world applications. This all-encompassing exploration provides the reader with a nuanced perspective that enables informed decision-making about the variety of methodologies built on explicit application or industry requirements.
- **Emerging trends and challenges** The advantages and challenges associated with each standard are addressed; this survey enunciates to the testimony of potential research directions and emerging trends. This progressive technique is helpful for practitioners, decision makers, and researchers, looking for a stay to take the modern advances in time series analysis.
- **Applicability across industries:** The survey's emphasis on industrial data adds a layer of distinction, making it especially relevant to industries that are transforming to the digital side. By illuminating how each area can be useful in industrial environments, this study helps as an applied guide for experts navigating the convolutions of time series within various segments. By being expressive, the data can be applied in industrial frameworks.
- **Guidance for future research:** The fusion of classical machine learning, deep learning, and graph neural networks, coupled with a thorough study of challenges, circles the phase for future research activities. This study not only fuses the current knowledge but also addresses the branded limitations.

In [Table 1](#), we compare our survey with other surveys. Survey one ([Bertolini et al., 2021](#)) has just investigated ML techniques, the second survey ([Drakaki et al., 2022](#)) thoroughly focuses on ML and DL methods, the third one ([Waikhom & Patgiri, 2023](#)) clearly focuses

Table 1 Comparisons of different survey articles with ours.			
	ML techniques	DL techniques	GNNs
Bertolini et al. (2021)	✓	✗	✗
Drakaki et al. (2022)	✓	✓	✗
Waikhom & Patgiri (2023)	✗	✗	✓
Our survey	✓	✓	✓

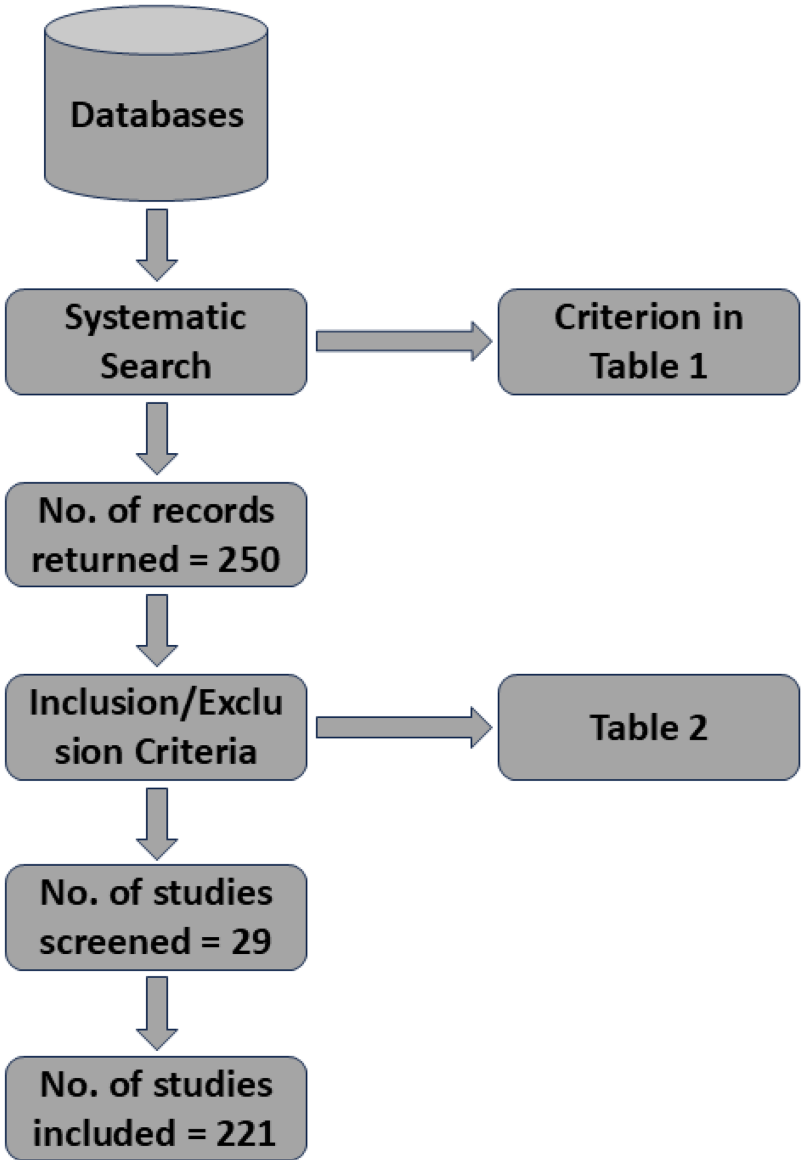
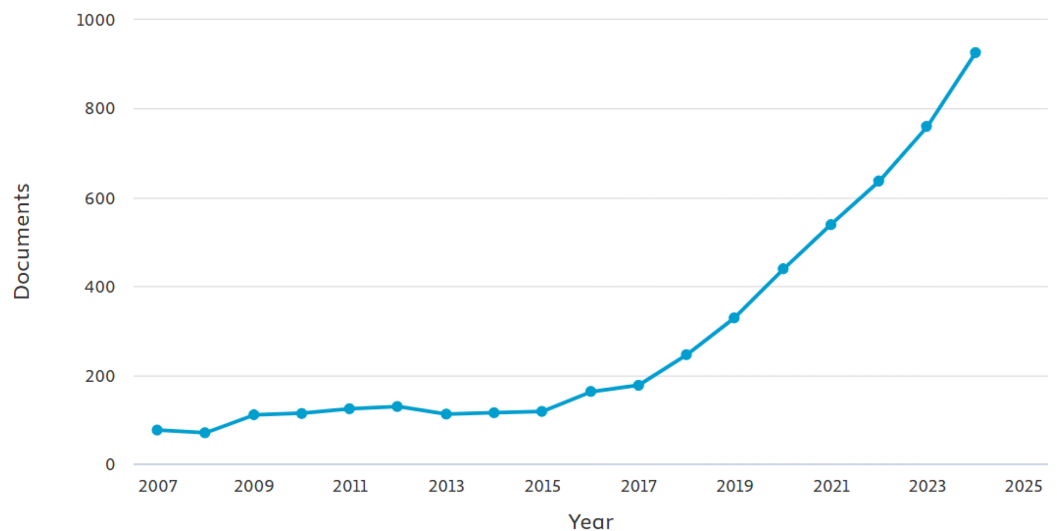


Figure 4 PRISMA diagram to illustrate the study selection process.
 [Full-size !\[\]\(b7bb2c9869ff794e0138733720a5c7ad\_img.jpg\) DOI: 10.7717/peerj-cs.3097/fig-4](#)





**Figure 5** The number of documents *i.e.*, articles and conferences published yearly. This search was performed on the Scopus platform. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674\_img.jpg\) DOI: 10.7717/peerj-cs.3097/fig-5](https://doi.org/10.7717/peerj-cs.3097/fig-5)

on GNN, compared to these surveys we have deep dived into all three techniques investigated in the surveys mentioned, which is a unique part of our survey.

## METHODOLOGY

This section provides a detailed description of the systematic approach used to conduct this study and the survey on time series analysis for industrial frameworks using graph neural networks, machine learning, and deep learning.

The literature initially collected was a total of 250 manuscripts. The literature was collected from different academic databases and sources, including Google Scholar, ArXiv, PubMed, ACM Digital Library, and IEEE Xplore. Various keywords such as ('time series analysis' OR 'forecast') AND ('machine learning' OR 'deep learning' OR 'graph neural networks') AND ('fault detection' OR 'anomaly detection'). Databases like Google Scholar are known for their broader coverage, ensuring that gray literature and interdisciplinary studies are not overlooked. The inclusion of ArXiv was to capture latest and cutting edge research, which may not be available in peer-reviewed literature. The other databases such as IEEE Xplore and PubMed were given priority because of peer-reviewed and high-quality articles in engineering and computational fields. The selection of unsupervised and semi-supervised learning (SSL) models, was their emphasis on limited labeled data scenarios most common in industry. The exclusion consists of studies that lacked practical validation and focused solely theories in industry. In Fig. 4 we show the PRISMA diagram to illustrate the study selection process. In Fig. 5 we illustrates the yearly studies conducted and how these studies have rose in all these years. Tables 2 and 3 shows the systematic search of different articles and the inclusion as well as exclusion criteria of articles.



**Table 2** Systematic search on different academic databases (title, abstract, keywords).

Criterion	Search terms
ML/DL	("Machine learning" OR "Deep learning" OR "Artificial Intelligence" OR "AI" OR "Neural Network")
TSC	(Classif* OR "Prediction" OR "Forecasting") AND ("Time series")
GNN	("Graph neural network" OR "GNN" OR "Graph convolutional network" OR "GCN" OR "Graph attention network" OR "GAT")
Industry	("Industrial Applications" OR "Manufacturing" OR "Smart Industry" OR "Industry 4.0")
Type	Journal article OR Conference article
Language	English

**Table 3** Inclusion and exclusion criteria of our survey article (INCL stands for inclusion and EXCL stands for exclusion).

Criterion	Search terms
INCL 1	Time series classification
INCL 2	Time series Forecasting OR Prediction
INCL 3	Time series Anomaly Detection
INCL 4	Only articles that precisely address explainability or interpretability
INCL 5	Only articles that show their approach for industrial time series
EXCL 1	Articles those are published prior to 2007
EXCL 2	Articles without any citations
EXCL 3	Articles that solely conduct analyses purely on the statistics
EXCL 4	Articles those do not involve industry

## APPLICATIONS OF TIME-SERIES ANALYSIS

### Time series forecasting

A critical feature of data exploration involves predicting or forecasting future values based on historical information. This method finds wide applications in various fields, including economics, finance, weather forecasting, energy utilization, and industrial manufacturing. The intention is to get an idea of the elemental trends and patterns within a timely methodical arrangement of data points to generate accurate predictions. In order to make effective predictions, we need to understand the different components of time series data, and these components are necessary to make compelling predictions. The three main components are: (a) Trends, the long-term direction or movement in the data. Trends can have no direction, can be down and can be up, (b) Seasonality—the fluctuations or frequent patterns in the time series data that occur at regular intervals, repeatedly, by aspects such as the time of the day, a few days, a week, a month, or a season, (c) Noise—the data that have arbitrary irregularities or fluctuations that are not described by the seasonality or the trend. It means the inherent unpredictability of pragmatic values.

### Time series anomaly detection

This effort focuses on identifying unexpected events and irregularities in an event within time-series or sequential data. Anomaly detection involves determining the timing of the abnormal event, while anomaly identification involves gaining significant insight into why,

where, and how the anomaly occurred. Because anomaly events are often difficult to obtain, recent research often approaches this task as an unsupervised task, which involves building a model that characterizes non-anomalous or normal data. This trained model is then used to identify anomalies by producing a good score each time an anomalous event is detected.

### Time series imputation

The focus is on the estimation and completion of missing or partial data points within a time series. Ongoing studies in this area can generally be divided into two primary methods: out-of-sample and in-sample imputation. Out-of-sample imputation deals with the imputation of missing data that are not available in the training data file, whereas in-sample imputation involves the imputation of missing values within a given time series data.

### Time series classification

The main goal of classification in time series data is to assign a categorical label to a given data by considering its characteristics or underlying patterns. Rather than capturing different sequences within individual time series samples, classification primarily involves recognizing discriminative patterns that help to separate samples based on their own class marker, the label.

## FOUNDATIONAL MODELS OF TIME SERIES ANALYSIS

Traditional machine learning models are often characterized by their shallowness. In the context of artificial neural networks (ANN), standard neural networks (NN) typically have no more than two layers and have limited data processing capabilities in their original state ([Choudhary et al., 2022](#)). In various operations, the need for prior feature engineering arises when examining extensive data, as well as data with high dimensionality, with traditional neural networks. Typically, data are processed using dimensionality reduction techniques, including data mapping techniques such as SOM or Principal Component Analysis (PCA) ([Gewers et al., 2021](#)). As a result, a hybrid intelligent system formed by combining two or more models becomes essential to effectively analyze complex data. Deep learning, as indicated by the use of many hidden layers within the architecture of an artificial neural network (ANN) ([Bui et al., 2020](#)), differs fundamentally from conventional machine learning models in the way in which it learns representations from raw data ([Angelov & Gu, 2019](#)). Several levels of abstraction ([Bengio, 2009](#)) are acquired for data representation in a deep learning model. Essentially, the learning process derives increased meaning from the data using advanced data abstraction at a higher level ([Yassine et al., 2019](#)). Deep learning methods defeat outdated machine learning techniques in the number of hidden layers. However, the distinctive and defining aspect of deep learning that distinguishes it from traditional machine learning is the advanced feature engineering capabilities of deep learning models. [Table 4](#), presents the use cases, strengths, and limitations of different models in industrial sector. [Table 5](#), shows the comparative analysis of different approaches of ML, DNN, and GNN in time series tasks. [Table 6](#) shows a

**Table 4** Use cases, strengths, and limitations of different models in industrial sector.

Model type	References	Use Cases	Strengths	Weaknesses
ARIMA	<a href="#">Xu et al. (2022)</a>	Time-series forecasting (e.g., sales, energy demand)	Interpretable, effective for linear trends and seasonality	Limited to linear dependencies, struggles with non-stationary data
LSTM	<a href="#">Xu et al. (2022)</a>	Predictive maintenance, anomaly detection in time-series	Captures long-term dependencies, handles sequential data	Computationally expensive, requires large datasets
CNN	<a href="#">Wang et al. (2020a)</a>	Quality control using visual data, fault detection	Excels at spatial and hierarchical feature extraction	Struggles with sequential data, not inherently interpretable
GNN	<a href="#">Chen, Feng &amp; Wirjanto (2023)</a>	Supply chain optimization, industrial IoT, network fault diagnosis	Models relational and topological data, scales to complex networks	Computationally demanding, requires domain knowledge for graph construction

**Table 5** Comparative analysis of ML, DNN, and GNN approaches in time series tasks.

Approach	Reference	Specific applications in time series	Computational complexity	Accuracy	Scalability	Strengths	Limitations
Machine learning (ML)	<a href="#">Goldsteen et al. (2022)</a>	Forecasting: ARIMA, Classification: SVM, Anomaly Detection: Isolation Forest.	Relatively low for basic algorithms (e.g., Linear Regression, SVM), but increases with ensemble methods (e.g., Random Forest).	Moderate for simple models; varies with data preprocessing and engineering.	High, suitable for small to large datasets.	Easy to implement, interpretable results, low resource requirements.	Performance depends heavily on feature engineering; limited for highly complex datasets.
Deep neural networks (DNNs)	<a href="#">Zintgraf et al. (2017)</a>	Forecasting: LSTMs, GRUs; Classification: CNNs, MLPs; Anomaly Detection: Autoencoders.	High, due to training and optimization of deep architectures.	High, especially for large and complex datasets.	Moderate, depending on hardware and parallelization.	Ability to model complex relationships, no need for feature engineering, effective for large-scale data.	Resource-intensive, requires significant labeled data and computational power; prone to overfitting.
Graph neural networks (GNNs)	<a href="#">Zambon, Livi &amp; Alippi (2022)</a>	Classification: Node or graph-level predictions; Anomaly Detection: Structural patterns in graphs.	High, involves graph construction and message-passing operations.	High, particularly for data with inherent graph structure or inter-dependencies.	Moderate, scalability depends on graph size and sparsity.	Captures spatial and relational data, handles irregular data structures, interprets complex dependencies.	Computationally expensive, challenging for dense graphs or datasets without clear graph structure.

continuation of [Table 5](#), depicts the different key performance metrics to better understand the strengths and limitations.

This uniqueness is characterized by intricate feature abstraction and construction, which is integrated into the assembly of the technique within the learning process. Robustness to data variation is a key strength of deep learning models, attributed to their

**Table 6** Analyzing key performance metrics to better understand strength and limitations.

Approach	Reference	Errors	Interpretability	Computational cost	Suitability for sequential data
ML	<a href="#">Goldsteen et al. (2022)</a>	High RMSE	High	Low	Poor (non-stationary issues)
DNNs	<a href="#">Zintgraf et al. (2017)</a>	Low MAE and RMSE, high F1-score	Moderate and Low	High	Excellent (long-term dependencies)
GNNs	<a href="#">Zambon, Livi &amp; Alippi (2022)</a>	Low MAE and RMSE, high F1-score	Moderate	High	Moderate (topological data)

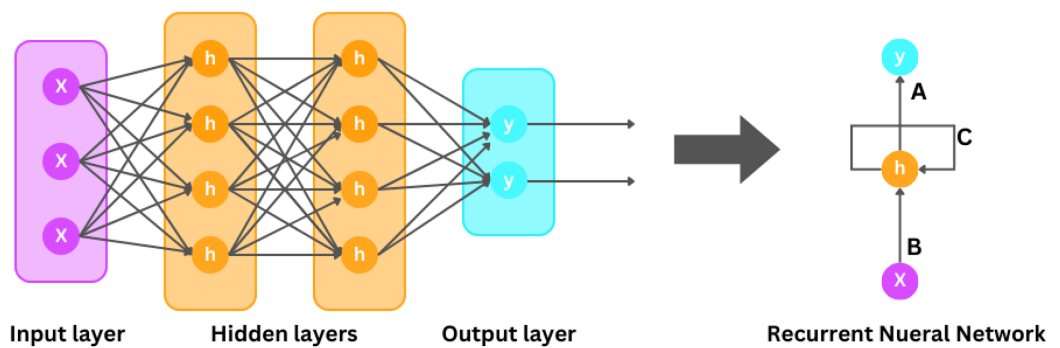
advanced intangible representation and feature engineering capabilities ([Wang et al., 2019](#)). In addition, the hierarchical structure inherent in deep networks enables them to effectively model the intricate nonlinear relationships present in large datasets. In contrast, traditional ML models often face challenges when dealing with large and high-dimensional data sets. As a result, ML models rely on feature selection as a dimensionality reduction method to facilitate the processing of large datasets more efficiently. The difficulty arises in real-world industrial applications because the large data sets collected are often contaminated with noise, outliers, and various types of anomaly, which poses a formidable challenge in feature selection. In the reign of intelligent engineering, autoencoders and their diversity, recurrent neural networks (RNNs), deep belief networks (DBNs), and convolutional neural networks (CNNs) stand out as the most widely used DL networks.

### Deep belief network

A deep belief network is formed by assembling several restricted Boltzmann machines ([Hinton, 2007](#)). It is worth noting that in a DBN there are connections between layers, whereas within a layer there are no connections between neurons. The construction of the network, organized layer by layer, facilitates the development of a classification feature representation ([Lagunas & Garcés, 2017](#)). This representation is instrumental in constructing an advanced representation of the input information. The DBN achieves input reconstruction by learning a probability distribution in the unsupervised training process. The Restricted Boltzmann Machine is a propagative stochastic feedforward artificial neural network known for its effectiveness in feature engineering. Training a DBN involves training multiple RBMs. The lower RBM, the hidden layer, is considered the network's training data, and the output of that RBM serves as the training data for the higher RBM. Once all RBMs are trained, a fine-tuning procedure is performed using a backpropagation process with the training information as the output ([Lee et al., 2018](#)).

### RNNs for time series analysis

There are different classes of neural networks, one of which is RNNs illustrated in [Fig. 6](#), which are designed to handle sequential and continuous data, making them particularly useful for analyzing time-series data. Unlike classical feedforward neural networks, RNNs have networks that formulate cycles, allowing them to maintain a hidden state that captures information about previous inputs into the system. RNNs take advantage of the hidden state to capture both temporal and contextual dependencies in sequential data. The



**Figure 6** Recurrent neural network.

Full-size DOI: 10.7717/peerj-cs.3097/fig-6

scientific design of an RNN involves computing the hidden states at each time interval based on the input information at that time step and also from the previous time step. Suppose that the input sequence is denoted as:

$$X = \{x^{(1)}, x^{(2)}, \dots, x^{(T)}\}$$

where  $x(t)$  characterizes the input at each time step  $t$ . Similarly, let

$H = \{h^{(1)}, h^{(2)}, \dots, h^{(T)}\}$  denote the hidden state sequence. The hidden state which is  $h(t)$  is calculated using the following equation.

$$h(t) = \sigma(W_{hx}x(t) + W_{hh}h(t-1) + b_h).$$

While  $W_{hx}$  denotes the weight matrix for the input  $x(t)$ ,  $W_{hh}$  denotes the weight matrix for  $h(t-1)$ , which is again the hidden state,  $b_h$  shows us the bias term, and  $\sigma$  finally corresponds to the activation function, generally tanh or ReLU are the most commonly used activation functions. The hidden state is then used to generate the output at this time step:

$$y(t) = \sigma(W_{yh} \cdot h(t) + b_y).$$

While  $W_{yh}$  represents the weight matrix for the connection between the output and the hidden state,  $b_y$  represents the bias term. The RNN can be trained by adjusting the hyperparameters ( $W_{hx}$ ,  $W_{hh}$ ,  $b_h$ ,  $W_{yh}$ ,  $b_y$ ) to reduce the loss, which shows the amplitude of the variance between the actual target and the predicted output  $y(t)$ .

However, traditional RNNs suffer from the limitations of exploding and vanishing gradients. The gradient vanishing trick happens when gradients become very small in the time of backpropagation, making it difficult for the model to learn long-range dependencies. Gradient exploding occurs when gradients are too large, causing unpredictability during training.

### Long short-term memory

To address the issues discussed in RNN, other advanced RNN designs, known as gated recurrent units (GRUs) and long short-term memory (LSTMs) illustrated in Fig. 7, have been brought into the light. LSTMs were designed to overcome the problems of traditional RNNs, namely the vanishing gradient dilemma, to better learn the long-term dependencies

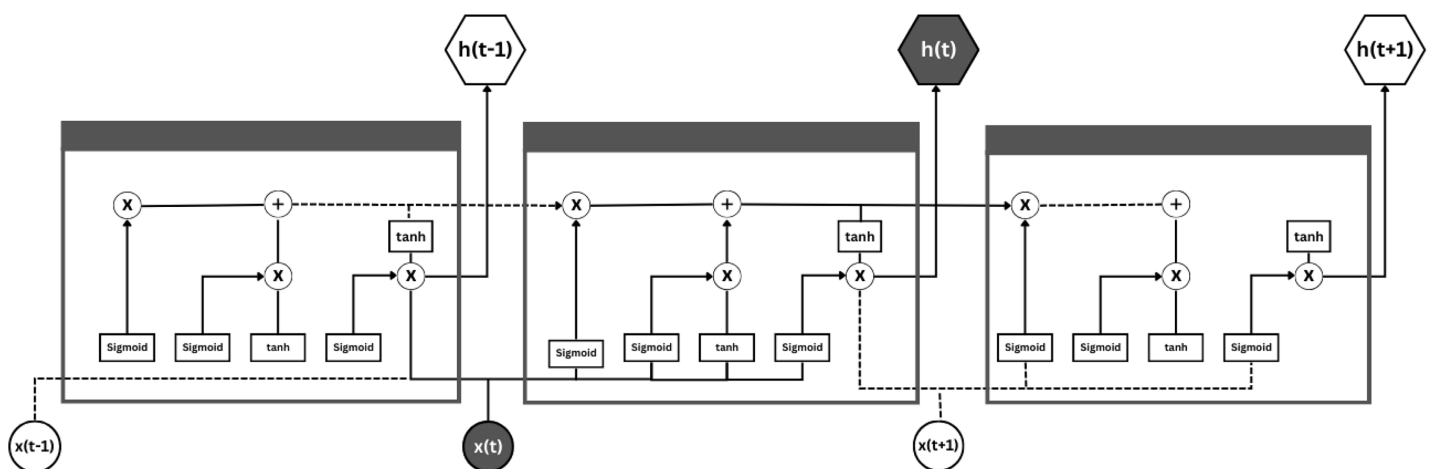


Figure 7 Long short-term memory.

Full-size DOI: 10.7717/peerj-cs.3097/fig-7

in sequential time series data. In LSTMs, this is achieved by introducing gating mechanisms and memory cells to regulate the flow of information between networks.

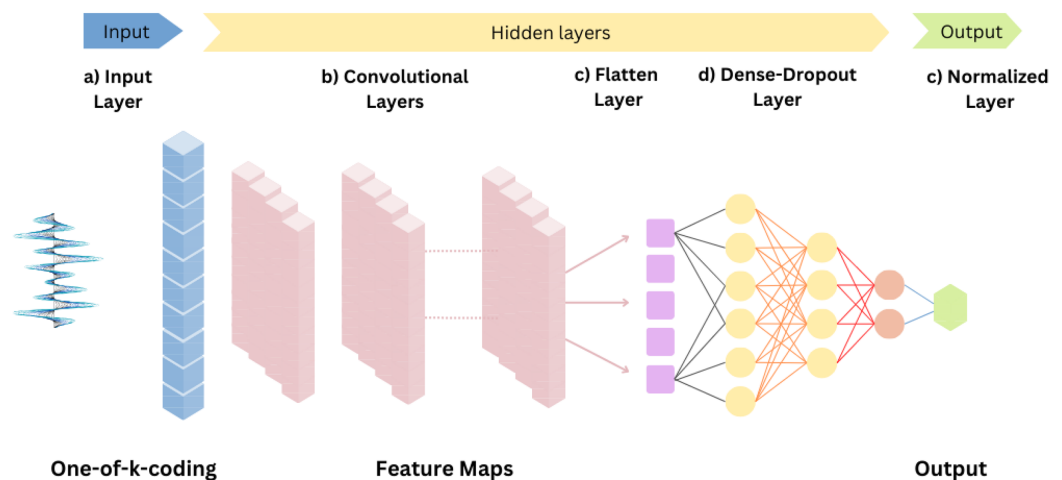
The key idea is the memory cell, which collects and holds information for long periods of time. The main building blocks of the memory cell are structured by three gates: the input gate ( $i_t$ ), the forgetting gate ( $f_t$ ) and the output gate ( $o_t$ ). The purpose of these three gates is to manage the input and output of information. Suppose that at time step  $t$ , the input is  $X_t$ , at time step  $t_1$ , the hidden state is  $H_{t-1}$ , and the memory cell state is  $C_{t-1}$  at time step  $t-1$ . Furthermore, the input value at time step  $t$  is assumed to be  $X_t$ , while at time step  $t-1$  the hidden state value is  $H_{t-1}$  and the memory cell state is  $C_{t-1}$  at time step  $\{t-1\}$ .

The input gate  $i_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i)$ , the forget gate  $f_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f)$ , the candidate cell state  $\tilde{C}_t = \tanh(W_C \cdot [H_{t-1}, X_t] + b_C)$ , the updated memory cell state  $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$ , the output gate  $o_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o)$ , and the hidden state  $H_t = o_t \cdot \tanh(C_t)$ . In the training process, these parameters ( $W_f, W_i, W_C, W_o, b_f, b_i, b_C, b_o$ ) are adjusted to reduce a preferred loss function, of course, using backpropagation over time.

LSTMs have been extraordinarily successful in several time series data analysis tasks, including natural language processing, speech recognition, weather forecasting, industrial process forecasting, and financial forecasting. LSTMs are enabled by their gating mechanisms to use and carefully update information from earlier time steps, making them particularly well suited to modeling sequential time series data with complicated dependencies.

## Convolutional neural networks

In particular, CNNs as displayed in Fig. 8, designed for image processing problems have been revised for time-series analysis by incorporating one-dimensional convolutions. This design revision allows CNNs to capture hierarchies and local patterns within sequential time series data. We have also tried to explain how CNNs work on sequential time series



**Figure 8** Convolutional neural networks.

Full-size DOI: 10.7717/peerj-cs.3097/fig-8

data, including the mathematical underpinnings. In the case where we need to analyze time series data, CNNs use their one-dimensional convolutions to distinguish features or patterns within the sequential data. In the convolution layers, filters or kernels are applied to the input sequence of their local segments, capturing relevant patterns, called feature maps.

Suppose that  $X$  is the input time series data of length  $L$ , and the number of applied filters is  $K$ . The  $i^{th}$  filter at position  $j$ , which is the convolution operation, is given by  $(X * W_i)_j = \sum_{k=0}^{M-1} X_{j+k} W_{i,k} + b_i$ . Where  $X_{j+k}$  is the value of this input sequence at position  $j + k$ , the weight of the filter is  $W_{i,k}$  for position  $k$ , for the  $i^{th}$  filter  $b_i$  is the bias term and the variable  $M$  represents the filter size. The result of the convolution operation is a feature map for each filter individually. After this process, an activation function, mostly ReLU, is applied element-wise to familiarize with the nonlinearity:  $\text{ReLU}(X * W_i + b_i)$ . The reason for this is to learn more complex patterns. To reduce the computational load and also to reduce the dimensionality, *i.e.*, to down-sample the extracted feature maps, pooling layers are widely used. Within a certain window, this is  $\text{MaxPooling}(X) = \max(X)$ , which shows that the maximum or average value is taken by the pooling operation. Further fully connected layers can be added for advanced abstraction and prediction after one or more pooling and convolutional layers have been run. The result is classically convolved over one or more fully connected layers, followed by an activation function as required by the task. In addition, for effective training and learning, the network uses backpropagation and various hyperparameter optimizations so that the biases and weights are carefully tuned. 1D CNNs for time series data offer the convenience of dropping dimensionality, learning hierarchical representations, and capturing local patterns. 1D-CNNs have been effectively applied in various domains, including weather forecasting, health monitoring, financial prediction, and industrial process systems.



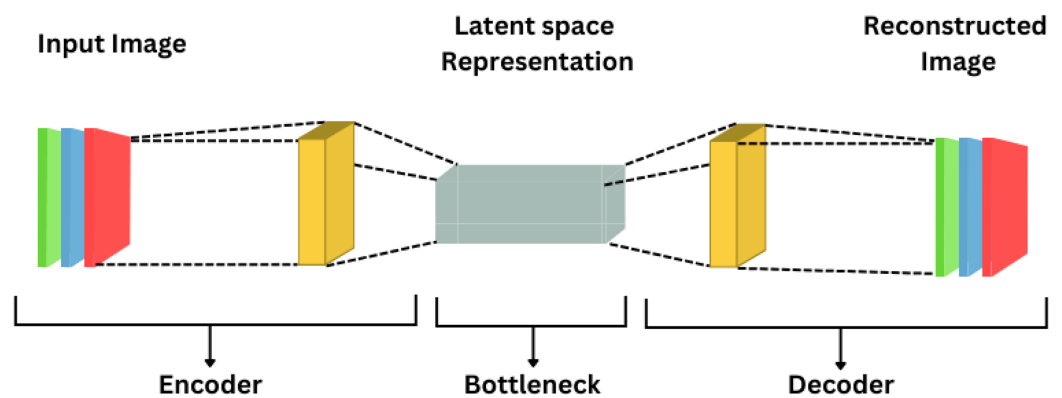


Figure 9 Autoencoder.

Full-size DOI: 10.7717/peerj-cs.3097/fig-9

## Autoencoders

Unsupervised feedforward neural networks are represented as autoencoders shown in Fig. 9, designed with the goal of having their output closely match the input data. This architecture consists of encoder and decoder steps, where the encoder transforms the input information into a latent representation and the decoder reconstructs the input based on this representation. Typically, gradient descent methods are used to optimize the hyperparameters of the methods to minimize the reconstruction error. Notable alternatives to AEs include sparse autoencoders and denoising autoencoders.

## Transformers

The dominance of the transformer architecture has emerged as a spotlight in various domains such as computer vision tasks, natural language processing tasks, but their employment is not limited to only text or image data. Transformers also play and have proven competent in the field of time series, presenting an attention-based and parallelized style. We will briefly discuss the basic and fundamental approaches of transformers.

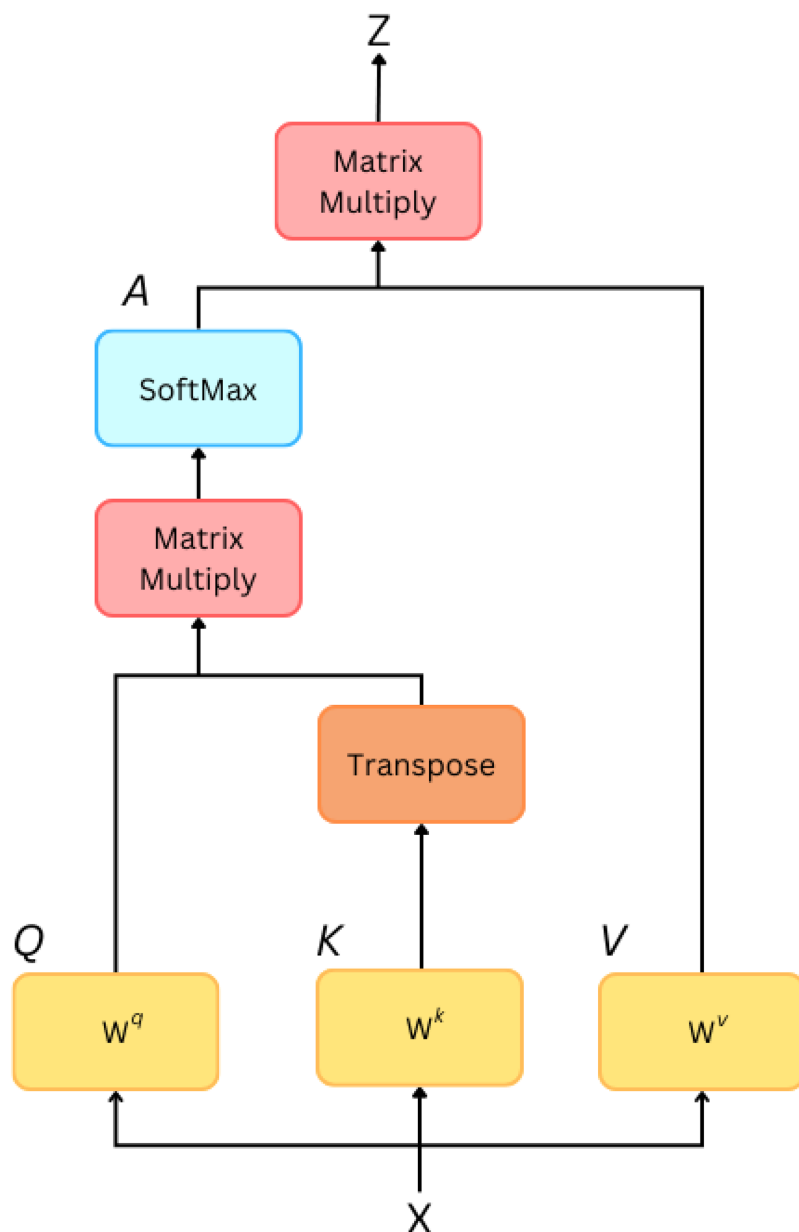
**Self-attention mechanism (Fig. 10):** A mechanism that allows each component in the input sequence to emphasize additional elements in order to capture dependencies regardless of their locality. Given an input sequence  $X = \{x_1, x_2, \dots, x_n\}$ , the self-attention score  $e_{ij}$  between elements  $x_i$  and  $x_j$  is calculated as follows  $e_{ij} = \text{softmax}(d_k(W_q x_i)^T (W_k x_j))$ .

Where  $W_q$  and  $W_k$  are learned weight matrices for queries and keys, and  $d_k$  is the dimensionality of the key vectors. **Scaled dot product attention:** A weighted sum of the values  $(W_v x_j)$  calculated by the self-attention scores is used to obtain the output for each position. The weighted sum for each item  $i$  is calculated as follows

$$\text{Attention}(Q, K, V) = \text{softmax}(d_k QK^T)V.$$

**Multi-head attention (Fig. 11):** To increase the capacity of the model, multi-head attention uses multiple sets of learned linear projections in parallel. Let denote the number of heads. The output of each head is concatenated and linearly transformed:  $\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)$ ,  $W_o$  Here,  $W_o$  is another learned weight matrix.

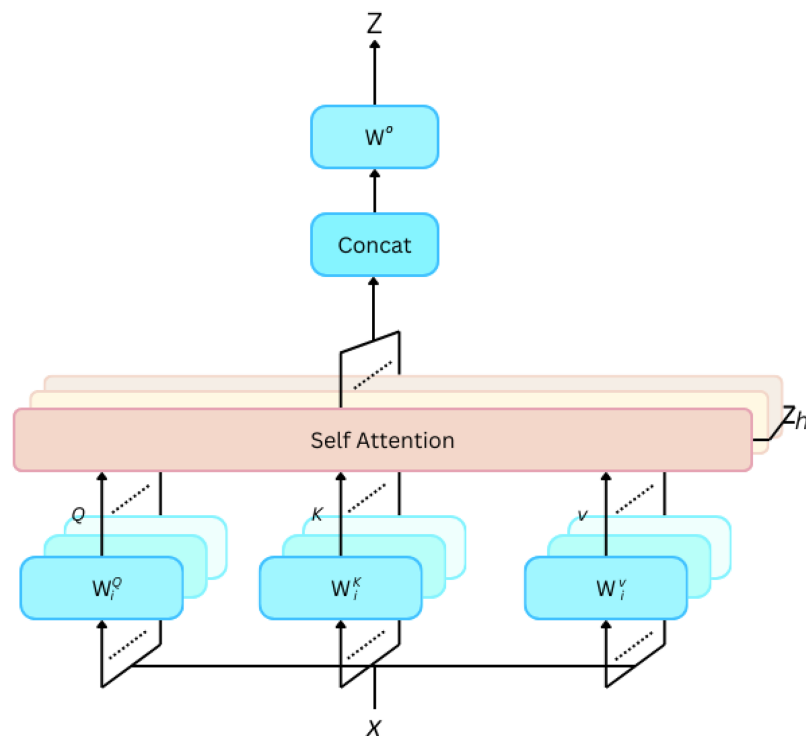
**Encoder and decoder structure (Fig. 12):** A transformer typically consists of an encoder and a decoder. The encoder processes the input sequence, and the decoder generates the



**Figure 10** Self-attention.

Full-size  DOI: [10.7717/peerj-cs.3097/fig-10](https://doi.org/10.7717/peerj-cs.3097/fig-10)

output sequence. The encoder consists of several layers, each of which contains a multihead self-attention mechanism followed by feedforward neural networks.  $\text{EncoderLayer}(X) = \text{FFN}(\text{MultiHead}(X))$ . The decoder has an additional multi-head attention layer that monitors the output of the encoder and allows it to focus on relevant parts of the input sequence.  $\text{DecoderLayer}(X, \text{EncoderOutput}) = \text{FFN}(\text{MultiHead}(X) + \text{MultiHead}(X, \text{EncoderOutput}))$ . These basics set the stage for understanding how transformers can be applied to time series data. In particular, the self-attention mechanism allows the model to consider global dependencies and capture long-range relationships, making it suitable for analyzing sequential information.



**Figure 11** Multi-head self attention.

Full-size DOI: 10.7717/peerj-cs.3097/fig-11

## MACHINE LEARNING: INSIGHTS INTO FORECASTING, ANOMALY DETECTION, IMPUTATION, AND CLASSIFICATION STRATEGIES

Exploring the multifaceted field of time series analysis, this section delves into various works that shed light on prediction, anomaly detection, and classification. By reviewing a spectrum of methodologies, we unravel the intricacies of temporal data and provide a comprehensive view of advances in predictive modeling, anomaly detection, and event classification within the dynamic landscape of time series analysis. Accurate forecasting is imperative for improving industrial actions, moving towards machine learning, and two models taking dominance, Autoregressive Integrated Moving Average (ARIMA) and LSTM network techniques, play a critical and significant role. The study of [Xu et al. \(2022\)](#) begins a reasonable investigation of ARIMA and LSTM methods, exfoliating light on the model's advantages and limitations in prediction accuracy and computational competence. LSTM prove effective due to their ability to capture long-term dependencies in circumstances where future outcomes or predictions are significantly influenced by historical ones. However, it is expected that limitations may arise which can lead to overfitting, particularly when training with very inadequate data. Although techniques such as ARIMA are suitable for capturing linear dependencies, a situation may arise where they fail to adapt quickly to unexpected changes. In the research of [Binkowski, Marti & Donnat \(2018\)](#), the researchers investigate the analogy of ARIMA and LSTM methods, highlighting the volume of LSTMs to capture nonlinear dependencies. This study commits

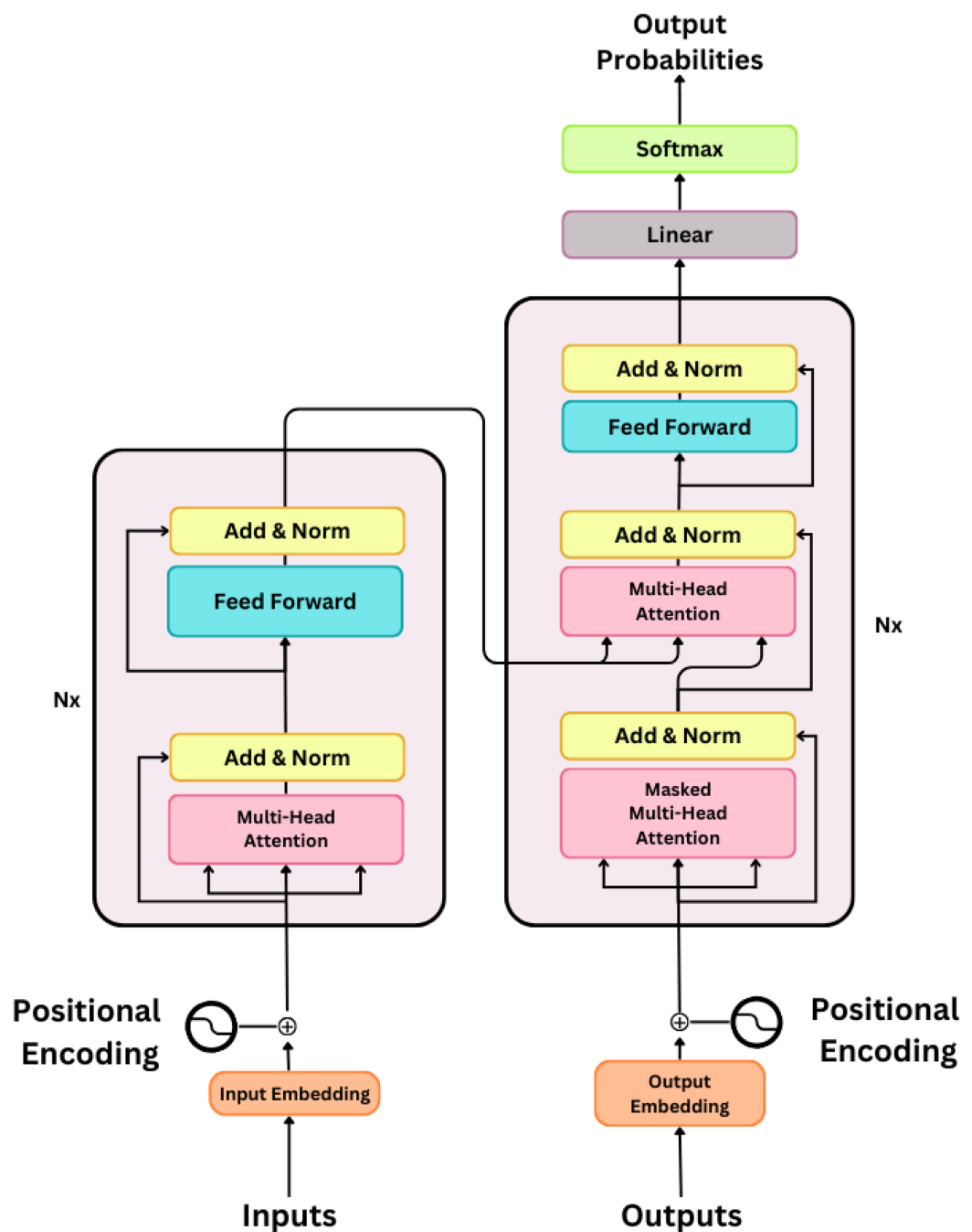


Figure 12 Encoder-decoder structure.

Full-size DOI: 10.7717/peerj-cs.3097/fig-12

to intuitions in the choice of appropriate techniques based on data characteristics. [Hewage et al. \(2020\)](#), proposed temporal convolutional networks (TCNs), exploring the functions of TCNs to demonstrate their success in capturing long-term dependencies, increasing the accuracy and efficiency to predict time series in industrial progressions.

Techniques such as ensemble learning are further explored in [Du \(2019\)](#). The study discusses the prediction of wind power generation using ensemble methods and highlights

the advantages of combining several techniques to improve accuracy. Probabilistic prediction is an essential prerequisite, and the study by [Pinson \(2013\)](#), emphasizes dependent (conditional) autoregressive methods for probabilistic prediction. This study offers insights into how to deal with uncertainty in prediction, especially in the context of renewable energy. In industrial time series data, the detection of anomalies is very important and critical for avoiding machine failures and ensuring system safety. A study conducted by [Jiang et al. \(2021\)](#) educates us on anomaly detection which is a very beneficial insight. The research discovers the use of SVM and isolation forests to detect outliers. SVM, which excels in cases where there are unbalanced classes, nevertheless requires great care in hyperparameter tuning. On the other hand, isolation forests, which take advantage of the isolation principle, show competence in training large data sets, making them suitable and favorable for industrial situations. However, limitations and threats arise when outliers are deeply rooted in condensed clusters. The research of [Ullah & Mahmoud \(2021\)](#) extends the argument to use different algorithms for anomaly detection in sensor data in the industrial Internet of Things (IIoT) environment.

The study provides a proportional analysis and comparison of algorithmic performance in the context of industrial IoT. Another study by [Schubert, Gupta & Wever \(2023\)](#) explores scalable anomaly detection. They develop techniques that are scalable for anomaly detection, keeping in mind the suitability for large dynamic systems, emphasizing the adaptability in handling growing anomalies in industrial developments. [Olivieri, Colleoni & Bonaccorso \(2023\)](#) argues on the impression of developing tendencies in predictive analytics in Industry 4.0, also taking into account anomaly detection. While she is more focused on the survey of tendencies, she sheds light on the insights of integrating machine learning models for real-time anomaly detection. Regarding streaming data, [Simkute et al. \(2021\)](#) discusses techniques for real-time anomaly detection. It also explores approaches that are best suited to streaming data, a critical requirement in industrial environments. In the classification orbit, machine learning methods such as neural networks, k-Nearest Neighbors (k-NN), and Random Forests dominate as key players. [Fahim & Sillitti \(2019\)](#) shares his thoughts on Random Forests, highlighting its efficiency in determining valued perceptions and educating decision making. The robustness of methods such as Random Forests, as opposed to overfitting, demonstrates their suitability for scenarios and claims where consideration of the decision-making process is very important. However, due to their ease and simplicity, they may struggle to capture complicated relationships in complicated time series data. [Shrestha, Krishna & von Krogh \(2021\)](#) explore k-NN precisely for the classification domain in time series. They highlighted the success of k-NN in capturing local patterns, assembling it to be a valued tool in very dynamic environments of the industry. However, we keep in mind the high dimensionality of time series data, which will have its challenges and will require careful attention. [Wang et al. \(2021b\)](#) sheds light on different classification models in the time series domain, which include neural networks, k-Nearest Neighbors and Random Forests. The investigation provides insight into the advantages and disadvantages of different approaches. In investigating the classification domain in time series with respect to machine learning methods, [Guigou \(2019\)](#) provides an overview that summarizes the

growing landscape of machine learning models for classification in the context of time series. The comparative analysis of [Wang et al. \(2022\)](#) enhances the perception of classification in time series data.

## ADVANTAGES AND LIMITATIONS OF MACHINE LEARNING MODELS IN INDUSTRIAL TIME SERIES ANALYSIS

In the domain of industrial time series investigation, machine learning techniques suggest several advantages, but they contend with integral limitations. With considerable advantages, machine learning models in the industrial time series domain, it is inescapable to endorse their limitations and constraints. [Idowu, Strüber & Berger \(2021\)](#) The scalability of ML techniques is discussed in detail; they provide useful insights into the application of ML at scale, focusing on the benefits of appropriate handling of large industrial datasets using distributed computing. However, there is the possibility that, considering the technological essence of scalability, constructive adaptation may be required for explicit industrial practice. Speaking of the interpretability of ML methods, [Zhang et al.'s \(2022b\)](#) research highlights the importance of constructing black-box methods that are interpretable. Although it provides direction on how to achieve interpretability, the hypothetical nature of the argument may require a real-world transformation for industrial frameworks. Furthermore, the [Shalev-Shwartz, Shammah & Shashua \(2018\)](#) investigation of generalization highlights encounters with generalisation in industrial settings. They also impart practices to promote model robustness, highlighting the need for techniques to adapt to different operating environments.

However, the hypothetical nature of the argument requires practical investigation for certain industrial settings. The ethical considerations associated with the production and use of ML methods are addressed in [Jaton \(2021\)](#). The research discovers the possible social effects of automation and artificial intelligence in industrial settings, highlighting the importance of accountable machine learning protocols. While the emphasis on ethical considerations may be crucial, it may require additional research for practical implementation in various industrial settings. Adapting to the evolving nature of industrial operations is one of the essential qualities of the volume of ML methods, allowing immediate integration of changes in data arrangements ([Tseng et al., 2021](#)). This conformance is particularly valuable in industries where operational situations change vigorously, polishing the competence of models to maintain accuracy in the seam of changing conditions. However, realizing the generalizability of these robust techniques across diverse industrial settings requires additional investigation. ML methods have established substantial success in predicting sustainable use within industrial operations ([Koste & Malhotra, 1999](#)). By examining time series or sequential data, these techniques can predict machine deterioration, enabling positive maintenance interventions and limiting unexpected breakdowns.

The predictive skills of ML promise to extend the maintenance route, thus significantly improving functional productivity. On the other hand, the practical application of predictive maintenance approaches requires meticulous validation in contrast to concrete machine settings to guarantee effectiveness and reliability. Furthermore, the combination

of ML techniques promises to improve automation and optimize complex processes in industrial control systems (*Jiang et al., 2021*). These techniques, skilled enough to adapt data patterns and learn these patterns, are likely to advance real-time decision-making and asset allocation efficiency. Nevertheless, the placement of ML in control systems requires an exhaustive inspection of the consistency and safety of the inferences, since in industrial frameworks any failure could have unadorned penalties. Regardless of these advantages, in industrial frameworks, time series analysis lacks the interpretability of ML techniques and its residue is a substantial encounter (*Carvalho, Pereira & Cardoso, 2019*).

The intrinsic complication of certain algorithms can obscure their decision processes, making it difficult to extract meaningful information from time-series data. Addressing this encounter requires advanced methods to improve the interpretability of these techniques without compromising their predictive accuracy. Furthermore, the energy competence of these methods in environments where the industry is resource constrained is a significant concern (*Aldhaheri et al., 2024*). Deploying and training skeptic ML methods requires extensive computational power, and ensuring the energy competence of these methods is imperative for sustainable industrial operations. Discovering optimization strategies and energy-efficient ML designs is the key to mitigating the environmental burden of extensive ML implementation in industrial frameworks. These techniques offer considerable support in the analysis of industrial time series, but their implementation is not free from limitations and restrictions.

The consideration of unbalanced data sets, trivial in industrial settings, is crucial to avoid biased methods. *Qian, Vaddiraju & Khan (2023)* Focal points to the obligation of developing models to grasp imbalances occurring in data classes, guaranteeing reliable technique results. An important concern that poses adversarial attacks, particularly in dangerous industrial arrangements (*Aryal, Gupta & Abdelsalam, 2022*). Shielding ML techniques from premeditated data handling involves vigorous security benchmarks that protect decision processes persuaded by model productivities. The resource-intensive nature of the training of complicated machine learning techniques presents a threat to smaller industrial units with inadequate computational frameworks (*Abdeldayem et al., 2022*). Regularizing the implementation of machine learning in different industrial settings requires the advancement of resource-efficient training approaches and lightweight model architectures. Interpretability remains a crucial challenge, especially in applications that are very critical in safety critical aspects where the decision-making method is crucial (*Papenmeier, Englebienne & Seifert, 2019*). The generation of interpretable and transparent machine learning techniques, customized to industrial requirements, is crucial for fostering trust and facilitating operational decision making.

In ML implementation, ethical considerations deserve careful consideration, as emphasized by *Boardman (2022)*. Talking of socio-economic impacts and biases in ML education requires the expansion of explicit ethical contexts for the use of industrial ML (*Boddington, 2023*). Exploring the multifaceted sociotechnical perspective is significant for supporting responsible AI protocols in the industrial sector. In conclusion, while ML techniques hold excessive potential for analyzing time series in industrial environments, certifying security, mitigating imbalance, regularizing access, reviving interpretability, and



acquainting with ethical issues are dynamic steps towards their active and responsible implementation in industrial frameworks.

## DEEP LEARNING: INSIGHTS INTO FORECASTING, ANOMALY DETECTION, IMPUTATION, AND CLASSIFICATION STRATEGIES

Starting from the pioneering frontier of time series investigation, this segment explores the competence of deep learning techniques in reshaping the field of prediction, anomaly detection, and classification. From extracting multifaceted temporal patterns to improving interpretability, these cutting-edge methods demonstrate the changing latent of deep learning across diverse applications within the domain of time series analysis. The increasing diffusion of technology has led to a gradual intensification in the use of soft measurement techniques for monitoring and analyzing industrial systems. The complexity inherent in thermal-hydraulic processes, heat-mass transfer, and physicochemical reactions within various industrial frameworks, including those in the petroleum, nuclear, thermal, chemical, and power sectors, poses challenges in establishing highly accurate mechanism models for predicting operational states (*Kadlec, Gabrys & Strandt, 2009*). Therefore, unlike methods based on mechanism models, time series prediction using a data-driven approach proves to be more appropriate and offers improved development and applicability to process systems within the industrial sector (*Kano & Ogawa, 2010; Karl Pearson, 1901*).

Recently, the term deep learning has been applied to methods that use different layers to capture latent features at advanced and more abstract levels. Deep learning attempts to select complex abstract features employing both non-linear and linear changes, providing a viable approach to address this problem. In the historical literature, four typical deep learning models have been defined and are widely used in time series forecasting tasks in many process systems in industrial environments. These models include deep belief networks, recurrent neural networks, encoder-decoder, and convolutional neural networks. These well-known models have shown successful applications in a variety of predictive or forecasting tasks. Recurrent neural networks stand out as a popular choice for achieving high performance in forecasting and prediction tasks. Examples include prediction of remaining useful life (RUL) for aircraft engines designed by *Yan, Shao & Wang (2004)*, prediction of solar power level proposed by *Miao et al. (2019)*, prediction of soft sensor industrial data established by *Wang et al. (2024)*, and a system to predict the temperature of long-term steam generator (SG) in a nuclear facility reported by *Yuan et al. (2020)*.

In addition, some scholars are trying to model hybrid methods, either in their unique forms or different forms, to effectively address the forecasting task. For example, *Nguyen, Liu & Zio (2020)* designed an approach that combines a variational recurrent unit with a Bayesian network in an IoT framework to achieve time series forecasting in meteorological data (*Das & Ghosh, 2019*). *Wang et al. (2021a)* proposed a CNN-LSTM network as a solution to forecast the levels in electrical power structures (*Zheng et al., 2023*). *Zheng et al. (2023)* used a hybrid CNN-LSTM model to capture 2D

link information and retain long-term memory to predict different traffic scenarios. [Lima et al. \(2021\)](#) integrated LSTM-CNN with an autoencoder to predict time series in metal packaging plants. Another study ([Zhang et al., 2023](#)) used principal component analysis for the direct extraction of multivariate time series features and combined it with a CNN-GRU method, which is again a hybrid model. A study by [Wang et al. \(2020a\)](#) introduced the variational mode decomposition integrated with a GRU technique. Furthermore, [Yu et al. \(2022a\)](#) supported a method for predicting wind speed using an LSTM-based method with empirical wavelet transform. Recently, there has been a growing passion and interest in deep neural networks ([Zintgraf et al., 2017](#)), driven by numerous demonstrated successes in various tasks ([Jain, Nandakumar & Ross, 2016](#)).

The proven ability of these methods to detect advanced correlations in complex data, often characterized by significant volume and dimensionality, contributes to their popularity ([Vakaruk et al., 2021](#)). This trend has not spared anomaly detection from multivariate time series data, which has witnessed a surge in DNN-based methods. These methods have proposed systematic developments and demonstrated improved performance, as highlighted in [Rana et al. \(2020\)](#). Methods based on deep neural networks seek to acquire deep latent representations of time series in multivariate data, allowing the inference of a model of variability. This learned model is then used to classify anomalies in previously unseen data. The surge in the adoption of deep neural network architectures is driven by the need to understand potentially complex data patterns underlying the temporal evolution of multivariate time series data. Building on the aforementioned rationale and inspired by the laudable achievements of DNNs in various domains, researchers have shifted their focus away from direct contrasts with other traditional approaches, such as conventional techniques or statistical methods, including machine learning ([Avci et al., 2021](#)).

This tendency has motivated researchers to develop increasingly complex models to improve the performance of DNN-based models. However, this is often done without substantial theoretical or empirical evidence demonstrating their superiority over more established techniques documented in research ([Vakaruk et al., 2023](#)). Training DNN-based techniques is a complex process that involves the evaluation of a significant number of hyperparameters, requiring large training model sizes and significant computational characteristics. Furthermore, the complexity of these models continues to increase with the ongoing development of larger and more complex architectures. On the contrary, traditional models tend to be simpler, easier to understand, lighter in weight, and often well adapted to the practical constraints of real-world applications.

Therefore, it is essential to determine whether the complication introduced by deep neural network methods is an essential trade-off for performance gains, or whether the advances reported in recent years are misleading ([Al-Garadi et al., 2020](#)) and advocate a preference for traditional methods. Several studies have investigated the actual benefits of DNN-based methods in various application domains. An initial study by [Jiao et al. \(2020\)](#) demonstrated that methods using conventional linear regression techniques outperformed methods based on deep neural networks in blind reconstruction in single-pixel imaging and an optical cryptosystem outbreak. [Zhang et al. \(2018\)](#) introduced a constancy test and

showed that image reconstruction methods based on deep neural networks have a high sensitivity to small perturbations in the input images during training, leading to unbalanced results.

In addition, [Maron et al. \(2021\)](#) has shown that even small changes in the input to a deep neural network, often imperceptible to an individual, can disrupt even the most proficient neural networks. This underscores the lack of robustness of methods based on deep neural networks and the fact that these models rely heavily on large amounts of data. The concept of a digital factory envisages a highly digitized and networked integration of equipment and machinery, designed to increase production and value through self-optimization and automation. In a computerized industrial process, equipment settings are closely linked to both productivity and quality. A stable process contributes to an increase in value, while an efficient process reduces assembly time and supports production. Therefore, the timely detection of faults or the anticipation of potential anomalies in the equipment is critical. The equipment used in digital factories includes manufacturing equipment, structure equipment, and logistics digitization equipment.

Manufacturing equipment is responsible for ensuring that products are manufactured efficiently and to excellent standards. Meanwhile, infrastructure plays a key role in supplying gas, water, chemicals, and electricity to the manufacturing system. It is also involved in the management of chemical waste and wastewater treatment. Logistics digitalization facilities are tasked with transporting goods from the point of origin to the other point. Although various machine learning methods are used to identify anomalies, faults, and damage in such industrial systems ([Fernandes, Corchado & Marreiros, 2022](#)), deep learning models have shown significant potential in this regard. Data-driven models play a key role in improving system operation in large manufacturing industries, as they have the ability to identify potential failures without requiring extensive knowledge of the domain. [Chevrot, Vernotte & Legeard \(2022\)](#) used an LSTM-based autoencoder to study the standard state of the equipment and identify anomalies in various multivariate time series streams related to production equipment machines. The LSTM-based autoencoder consists of an encoder and a decoder, each of which incorporates long- and short-term memory networks, which are modifications of recurrent neural networks. Following the revolutionary impact of convolutional neural networks in computer vision ([Smith, Smith & Hansen, 2021](#)), scientists have extended their application to the analysis of time series data ([Shumway & Stoffer, 2017](#)). CNN-based diagnostic and fault detection models have demonstrated their effectiveness in managing multivariate time series data from semiconductor manufacturing developments in studies such as [Hsu & Liu \(2021\)](#), [Zhu et al. \(2021\)](#).

## ADVANTAGES AND LIMITATIONS OF DEEP NEURAL NETWORKS

In the dominion of time series investigation, multifarious practices and methods address definite challenges. A study [Gaugel & Reichert \(2023b\)](#) proposes an efficient labeling technique for time series, using a segmental strategy with three dedicated modules for probable performance developments. The challenges continue in rapid fashion

engineering, where a deep learning architecture in [Wang et al. \(2021b\)](#) integrates visual, time series, and textual data to improve the detection of different trends in the fashion industry, regardless of the constraints imposed by the lack of historical data. For fault diagnosis, a technique integrating LSTM and the residual convolutional neural network validates the efficiency of extracting characteristics from time series data with compact training effort ([Yao, Yang & Li, 2021](#)).

Challenges include understanding the details of the model, a lack of data to evaluate, and the effect of small data proportions. In industry and engineering, an optimization and machine learning method highlights waste reduction and resource savings, encountering limited set of certain instances, and data accessibility complications ([Dambros et al., 2019](#)). Feature extraction based on scalable hypothesis tests ([Weichert et al., 2019](#)) presents feature extraction with great efficiency, integrating meta-information with parallelization and time series. It lacks to further address the prerequisite of domain knowledge and also very incomplete points on evaluation metrics. Another study ([Christ, Kempa-Liehr & Feindt, 2017](#)), proposes a feature selection based on principal component analysis, confronted with extensive execution work and a lack of standard datasets.

Reasonable research in anomaly detection techniques for industrial control systems highlights the effect of size set in training, illustrates implementation variances, and reveals limitations in method diversity. Some studies such as [Wang, Yang & Li \(2020\)](#), which introduces how to form an image for signals, also illuminate practical applications, but weaken in the discussion of network architecture and in-depth comparative analysis. An approach ([Tziolas et al., 2022](#)) for unbalanced data using a GAN and feature extractor limits the argument on the in-depth description and research on hyperparameter tuning details and how they built the feature extractor. [Liu, Hsaio & Tu \(2019\)](#) introduces process optimization, a tensor structure for ensemble modeling, and data transformation. Constraints arise from the installation of frequent techniques in production and a limited argument on unbalanced data encounter. Incorporating data analytics and sensors to improve process manufacturing, they propose an ARIMA-based predictive maintenance technique, deficient in detail of key metrics and speaking of security as well as privacy concerns. In tissue engineering, the combination of time series technique and dielectric spectroscopy demonstrate improved prediction with LSTM, but deficit precision on external validation, dataset size, and pertinence to longer-term value situations. [Kim et al. \(2023\)](#) presents Gated CNN-based Transformer, which demonstrates effective soft-sensor modeling but shortcomings in comprehensive analysis of trade-offs and component effects. [Liu et al. \(2021\)](#) proposes contrastive predictive coding (CPC) to label effectiveness but has limitations in real-world scenarios and comparisons.

[Jiang et al. \(2019\)](#) investigates multi-head CNN-RNN for anomaly detection, which uses independent CNNs for different sensors, but finds limitations in baseline comparison. [Zhang et al. \(2021\)](#) introduces a hybrid model of bidirectional deep inner partial least squares (DiPLS) with LSTM, which shows better developments in interpretability but requires careful consideration of the extended computational task. The transfer probability of Markov chains ([Kanawaday & Sane, 2017](#)), a feature extraction technique, expresses limitations in lack of intuition in information security, insufficient comparison debate, and

sensitivity to hyperparameters. The attention-based architecture (Jatti, Sekhar & Shah, 2021) for soft sensor simulation presents spatiotemporal attention, but has a deficit in exploring the model interpretability and sensitivity tasks in industrial frameworks. Giving manufacturing process data to a branched LSTM as a spatial sequence investigates fault detection, but limitations arise in terms of tuning hyperparameters essentials applicability. Transfer learning setups for time series partitioning provide insights into training speed, but deep investigation of layer-frozen residues is a challenge. Ali Nemer et al. (2022) introduces a linear-discriminative generative-discriminative method, this technique competently discovers deep neural network pipelines, highlighting productivity over precision. Limitations consist of the trade-off between accuracy and efficiency, limited validation, and domain specificity. Geng et al. (2022) presents a TC-GATN architecture that employs temporal causal graph attention networks that engage nonlinear relationships, which also faces challenges in dependence on causality information and complexity.

The fused techniques of Gamage, Klopper & Samarabandu (2022) integrate XGBoost and GRU, to extract features and timing knowledge, highlighting key feature outputs but demanding additional investigation of security and scalability fears. A method of Schockaert (2020) employing continuous learning reports a nonstationary challenge in data distribution, but has deficiency in strict evaluation and neglects resource and privacy concerns. A study by Fährmann et al. (2022) introduces a segmental data driven system to predict failures in different machines, it adapts to variations but lacks on how to handle data correctness, class imbalance, and lack of feature facts. Canizo et al. (2019) propose a hybrid prototype for anomaly detection in industrial IoT frameworks, which emphasizes real-time intuitions but requires additional investigation of privacy and scalability issues. The research by Wang, Bao & Qin (2023) presents a technique to predict multistep time series data in industrial processes, proposing a self-motivated transfer learning architecture that integrates multitasking and domain adaptation. However, there is a clear lack of detailed elucidation of a number of workings, including neural network construction and hidden layer selection principles, which limits the robustness of the technique. Limiting generalization to real-world industrial settings with a dataset of 2,859 samples also raises questions about model scalability. A new method by Zang, Liu & Wang (2018), for the sole purpose of maintenance prediction in the case of industrial drying hoppers, uses a deep learning technique that targets unbalanced data in time series scenarios. However, the lack of a physics-based prototype poses challenges in terms of interpretability. Limitations in accurately characterizing abnormal procedures and bias in event classification influence the reliability of the technique in various industrial settings.

The use of generative techniques to integrate time series with image coding for fault detection is an innovative method. But it faces challenges in reconstructing the image and obsessiveness with data size, complementing concerns about versatility in different industrial settings. Neural network using Bayesian knowledge introduced by Yuan et al. (2021), which is a stochastic device to detect anomaly refrigeration engineering. Uncertainty actions for curative activities depend on inadequate labor comfort, questioning practical relevance in real industrial setups. A deep network-based hybrid



architecture of [Li \(2021\)](#), for short-term multivariate time series prediction in the industrial IoT setting, merges several mechanisms. However, its precise suitability to the Industrial IoT framework and limited insight into selected features affect its interpretability. The preliminary hyperparameter service and its reworking by means of Bayesian optimization license additional investigation ([Gaugel & Reichert, 2023a](#)). The method for fault detection using 1D CNNs ([Shojaee et al., 2021](#)) in time series scenarios adapted to architectural improvements. Lack of highlighting how to handle a small training size, balancing fault event intricacy, overfitting concerns with appropriate sequence length for flawless performance.

## GRAPH NEURAL NETWORKS: INSIGHTS INTO FORECASTING, ANOMALY DETECTION, IMPUTATION, AND CLASSIFICATION STRATEGIES

As we navigate the complicated realm of time series, this unit sheds light on the contributory character of GNNs in transforming forecasting, anomaly detection, and classification. From interpreting complicated dependencies within systems that are highly interconnected, to demonstrating an all-encompassing use of temporal information, GNNs position themselves at the forefront of modernization, restructuring the changing aspects of time series submissions through their exceptional graph-oriented method. Time-series forecasting involves predicting future values in a time sequence by relying on past observations. The roots of time series forecasting can be identified in models ([Li et al., 2018](#)) that are statistically autoregressive. Their sole purpose is to use a linear combination of past values within a time series to predict its future values. In recent years, deep learning-based approaches have shown significant efficacy in predicting time series results, demonstrating a superior ability to capture non-linear temporal and spatial patterns ([Gao & Ribeiro, 2022](#)). Various methods have been used for time series prediction, including attention-based neural networks, CNNs, and RNNs. However, several approaches, including LSTNet ([Pan et al., 2019](#)) and temporal pattern attention-long short-term memory (TPA-LSTM) ([Zambon et al., 2019](#)), indirectly model or tend to overlook the complex dynamic spatial correlations present in time series. To effectively and explicitly model the spatial and temporal dependencies within the multivariate time series data, there have been notable advances in the use of methods based on GNNs. This has led to improved forecast performance. Forecasting models based on GNNs can be categorized and studied from different perspectives ([Ahmed et al., 2021](#)). In terms of forecasting tasks, many models focus on multistep forecasting—predicting several successive future steps using historical data.

In contrast, a minority deals with single-step forecasting, *i.e.*, predicting only a single or next arbitrary step ahead. The analytical dissection of these time series forecasting models involves three perspectives: exploring spatial (*i.e.*, intervariable) dependencies, examining intertemporal dependencies, and architecturally integrating temporal and spatial modules. The relationships between spatial elements, also known as intertime series connections, have a significant impact on a model's ability to forecast ([Guo et al., 2019](#)). Some of the methods such as (1) spatial GNNs, (2) spectral GNNs, or (3) a fusion of both are

commonly used by contemporary research in some cases where they are faced with accompanying graph structures and time series data that can be used to illustrate the strength of connections in them to model and capture these spatial dependencies in time series. In the early stages, GNN-based forecasting models mainly used ChebConv ([Wu et al., 2019](#)) to approximate graph convolution using Chebyshev polynomials, effectively modeling and capturing dependencies between time series. As an example, ChebConv layers and temporal convolution were combined to effectively capture temporal and spatial patterns. Proposing temporal and spectral graph neural networks, StemGNN ([Chen et al., 2020](#)) aims to use the frequency domain and ChebConv convolutional neural networks to extract complex patterns from time series.

Subsequent relevant studies have mostly followed this approach, using ChebConv to introduce inventive modifications and to model spatial time-series dependencies. Such adaptations include multigraph construction ([Wang et al., 2020b](#); [Zheng et al., 2020](#)), attention mechanisms ([Zhang et al., 2020a](#); [Guo et al., 2019](#)), and hybrid combinations of the two ([Yu et al., 2020](#)). Based on StemGNN, theoretical evidence has recently been provided by [Paassen et al. \(2020\)](#) that demonstrates the advantages of using spectral GNNs to model different signed time series relationships. This includes variables with strong negative and positive correlations in a multivariate time series. It was also found that for such tasks, any orthonormal family of polynomials can achieve comparable expressiveness, albeit with different empirical performance and convergence rates. Inspired by the recent successes of spatial GNNs ([Chen, Segovia & Gel, 2021](#)), an alternative line of research model dependencies between time series through graph diffusion ([Lan et al., 2022](#); [Liu et al., 2022](#)) or message passing ([Choi et al., 2022](#)). Looking at the graph, it becomes clear that these approaches are specific simplifications in contrast to those based on spectral GNNs, which highlight robust local homophilic tendencies ([Shao et al., 2022](#); [Chauhan et al., 2022](#)). In early techniques such as DCRNN ([Yu et al., 2022b](#)) and Graph WaveNet ([Cui et al., 2021](#)), GRU ([Zhao et al., 2020](#)), temporal convolution or graph diffusion layers are integrated to effectively capture patterns in time series data. In later studies, graph diffusion has been used by some approaches such as Spatio-Temporal Graph Diffusion Network (ST-GDN) ([Deng & Hooi, 2021](#)) and graph time series (GTS) ([Grattarola et al., 2020](#)). On the contrary, spatio-temporal graph convolutional networks (STGCN) (1st), which is a later iteration of ST-MetaNet ([Dai & Chen, 2022](#)) and STGCN ([Zambon, Livi & Alippi, 2022](#)), adopted graph convolutional network (GCN) ([Han & Woo, 2022](#)) and graph attention network (GAT) ([Chen et al., 2022c](#)) to characterize spatial dependencies, allowing the aggregation of information from neighboring time series. Other relevant studies, including multi-range attentive bicomponent GCN (MRA-BGCN) ([Wu et al., 2021](#)), Spatio-Temporal Graph Neural Networks (STGNN) ([Chen et al., 2022a](#)), graph multi-attention network (GMAN) ([Guan et al., 2022](#)) and adaptive graph convolutional recurrent network (AGCRN) ([Srinivas, Kumar Sarkar & Runkana, 2022](#)), introduced adaptations to capture relationships between time series through message passing. The spatial-temporal synchronous graph convolution was introduced by Spatio-Temporal Synchronous Graph Convolutional Network (STSGCN) ([Chen, Feng & Wirjanto, 2023](#)) to improve learning efficiency.



This extension of GCN is specifically designed to capture both spatial and temporal dependencies within localized spatio-temporal graphs. Before using graphs and temporal convolutions, Spatio-Temporal Fusion Graph Neural Network (STFGNN) ([Chen et al., 2022b](#)) formed fusion graphs that integrated spatial and temporal aspects using dynamic time wrapping (DTW). Zero-Inflated Graph Convolutional Networks (Z-GCNETs) ([Zhou, Zeng & Li, 2022](#)) improved current approaches by integrating prominent time-dependent topological data, with particular emphasis on zigzag persistence images. Multiscale temporal graphs were introduced by METRO ([Duan et al., 2022](#)) to represent dynamic temporal and spatial relationships within time series data. This was achieved by incorporating cross-scale graph fusion modules and single-scale graph updates, with the aim of integrating the modeling of spatial-temporal dependencies. An alternative avenue of extension is the integration of graph propagation, which allows the merging of substructures and higher-order relationships within the network. Examples of this include spatial graph propagation (SGP) ([Aldhaferi et al., 2024](#)) and multivariate time-series graph neural network (MTGNN) ([Chauhan et al., 2022](#)). Specifically, SGP uses multihop spatial processing and reservoir computing to generate precomputed spatio-temporal representations, resulting in efficient and scalable predictive models. Graph propagation and continuous graph propagation were introduced by subsequent research efforts such as multivariate time-series graph ordinary differential equation (MTGODE) ([Ao & Fayek, 2023](#)) and temporal pattern graph neural network (TPGNN) ([Srinivas, Kumar Sarkar & Runkana, 2022](#)) based on temporal polynomial coefficients. Comparable contributions in this area include spatio-temporal graph ordinary differential equation (STGODE) ([Chauhan et al., 2022](#)) and spatio-temporal graph neural controlled differential equation (STG-NCDE) ([Shao et al., 2022](#)).

Unlike GAT-centric approaches, methods based on graph transform ([Zambon, Livi & Alippi, 2022](#)) can capture extended spatial dependencies due to their extensive coverage, which puts them in a distinct category of improved techniques. The goal of time series anomaly detection is to detect data observations that deviate from the typical pattern established by the data generation process. Any data point that falls outside this pattern is characterized as an anomaly, while data points within the established pattern are referred to as normal data. It is worth noting that in the literature, terms such as novelty and outlier are often used almost interchangeably with the term anomaly. Deviations from standard conditions may manifest themselves as a single data point or as a series of observations. Unlike regular time-series data, anomalies present a challenge in characterization due to two main factors. Firstly, they tend to be associated with infrequent events, making the collection and labeling process difficult. Second, defining the full range of possible anomalous events is typically unattainable, compromising the effectiveness of supervised learning methods. As a result, unsupervised detection techniques have been extensively explored as a pragmatic approach to solving complex real-world problems. Historically, approaches such as distance-based methods and distributional techniques have been widely used to identify anomalies in time series data. Using distance metrics, the former category measures the deviation of observations from representative data points, while the latter focuses on identifying anomalies by targeting points with low probability. As the

complexity of the data generation process increases and the dimensionality of the multivariate time series expands, these approaches become less effective.

With the advancement of deep learning, initial studies have suggested the use of recurrent models that employ both reconstruction and prediction approaches for improved anomaly detection in multivariate time series data. Prediction and reconstruction strategies rely on prediction and reconstruction errors as measures of the discrepancy between expected and actual signals. The idea is that if a model trained on normal data has difficulty predicting or reconstructing certain data, there is a greater likelihood that such data are indicative of an anomaly. However, recurrent models (Aldhaheeri et al., 2024) have been found to do a poor job in explicitly modeling pairwise interdependence between pairs of variables, thereby limiting their effectiveness in identifying complex anomalies (Angelov & Gu, 2019). Recently, GNNs have shown significant potential in bridging this gap by skillfully capturing both spatial and temporal dependencies between variable pairs (Ali Nemer et al., 2022). An unsupervised approach to anomaly detection requires models to develop a comprehensive understanding of the characteristics that define normality within a given dataset (Canizo et al., 2019). Advances in anomaly detection and diagnosis have led to the introduction of more comprehensive backbone and scoring modules (Cheng et al., 2023), largely influenced by the adoption of GNN methods (Choi et al., 2022).

The goal of time series classification is to assign a categorical label to a given time series, and this assignment is based on the inherent features or patterns within the time series. As highlighted in a recent review (Ali Nemer et al., 2022), the initial literature on time series classification focused mainly on distance-based methods for assigning class labels to time series (Liu, Ting & Zhou, 2012) and ensemble techniques such as hierarchical vote collective of transformation-based ensembles (HIVE-COTE) (Lines, Taylor & Bagnall, 2018). However, despite their state-of-the-art performance, both approaches face scalability limitations when applied to high-dimensional or large datasets. Exploring the possibilities of deep learning techniques, experts have initiated efforts to improve the effectiveness and scalability of time series classification approaches in response to these limitations. The potential of deep learning lies in its ability to capture complex patterns and hierarchies of attributes, and it has proven effective in addressing the challenges of time series classification, especially when dealing with datasets that contain a large number of training labels. For an in-depth exploration of time series classification using deep learning, we recommend the recent survey by Herrmann & Webb (2021). An aspect of this field that adds an element of intrigue, and is not covered in the aforementioned (Ali Nemer et al., 2022), is the use of GNN for time series classification. By transforming time-series data into graph representations, the robust capabilities of GNNs can be used to capture patterns at both localized and overarching levels. Capturing intricate relationships between different samples of time series data within a given dataset is a strength of GNNs. The conversion of a single variable time series into a graph using the Series2Graph methodology aims to recognize unique patterns that enable precise classification using a GNN. In this method, each time series is considered as a graph, where the graph acts as input to the GNN and produces outputs for classification.

First, the decomposition of each series into subsequences produces nodes, and these nodes are connected by edges to represent their relationships. After this transformation, a GNN is used for graph classification. The Time2Graph+ approach first introduced the Series2Graph perspective ([Zhang et al., 2020b](#)). The Time2Graph + modeling procedure follows a two-step approach: first, a time series is transformed into a graph of shape lets, and then a GNN is used to understand the connections between these shape lets. The Time2Graph algorithm divides each time series into sequential segments when creating a shape-let graph. It then uses data mining techniques to assign characteristic shape lets to these subsequences. These shapes let you act as nodes within the graph, establishing connections through edges. The formation of edges between nodes is determined by the likelihood of a shape to occur sequentially in a given time series.

As a result, the transformation of each time series produces a graph in which the nodes are shape lets, and the edges are the transition probabilities between these shape lets. Following the construction of the graph, [Zerveas et al. \(2021\)](#) uses a graph attention network in conjunction with a graph pooling operation to extract the overall representation of the time series. The resulting representation is then fed into a classifier to assign class labels to the time series.

## ADVANTAGES AND LIMITATIONS OF GRAPH NEURAL NETWORKS

GNNs are greatly successful in the industrial sector due to their competence to model dense relational structures, compelling them standard for applications such as industrial IoT (IIoT), supply chain optimization, and predictive maintenance. Contrasting to traditional techniques, which strain interconnected data, GNNs certainly attain dependencies involving nodes, permitting more anomaly detection and precise predictions in industrial systems. For instance, in predictive maintenance, GNNs can investigate sensor data from interconnected technology to identify immediate signs of failure, curtailing expensive downtime. In supply chain management, they optimize logistics by modeling relationships between suppliers, warehouses, and distributors, reducing disruptions and developing efficiency. Furthermore, GNNs scale successfully to large networks, such as power grids, where they can predict failures and improve energy distribution. Although computationally demanding, they are substantially more competent than fully connected deep networks when dealing with structured data. Their capability to learn from graph-based dependencies, pooled with enhanced interpretability, makes GNNs a needed tool for adopting in industrial challenges that concern complex, interconnected systems.

[Li et al. \(2018\)](#) proposes a novel technique for modeling spatial dependencies in traffic flows. It fuses sequence-to-sequence framework, diffusion convolution, and scheduled sampling practice to capture both temporal and spatial dependency. However, the limitations consist of the lack of interpretability and complexity of deep neural networks, likely costly computation, and the dependence on a precise representation of the road network graph. [Gao & Ribeiro \(2022\)](#) introduces a time-then-graph architecture for time-based graphs and demonstrates a fluency advantage over time and graph. Although

better performance is achieved for certain tasks, limitations include an emphasis on certain events, possible inadequacies for edge-level prediction, and the need for further theoretical investigation. Research in [Pan et al. \(2019\)](#) presents a common house of attributed graphs, but expresses limitations in treating significant generalization as a result of lack of elementary operators of mathematics. The sequences generated by graphs have vital complexity and are constrained by hidden variables, posing challenges in grasping complicated dependencies. The study of [Zambon et al. \(2019\)](#) introduces a technique to predict traffic flow using the spatio-temporal graph CNN, a characteristic of this method is the adaptability to predict traffic flow. While successfully capturing active features, there are limitations when dealing with ordinary 2D or 3D network input and an explicit lack of attention to apparent inducing aspects such as weather. [Guo et al. \(2019\)](#) presents Graph WaveNet, an integration of dilated causal convolution with graph convolution for spatiotemporal graph simulation. Limitations arise from issues such as scalability and ambiguities in the training of hidden dependencies.

The research in [Wu et al. \(2019\)](#) focuses on traffic prediction using the Multi-Range Attentive Bicomponent, which successfully captures both the node and the edge relationships. Interpretability is challenged due to the complexity of the model and the study is deficient in showing training-time insights. In addition, [Chen et al. \(2020\)](#) proposes a model for traffic prediction that implements a structure learning convolution, involving convolutional neural networks in graph domains. Limitations arise in the areas of hyperparameter sensitivity and computational complexity. Moving forward, the research of [Wang et al. \(2020b\)](#) presents synchronized graph convolutional neural networks that can capture both spatial and temporal dependencies in different network data predictions. Despite consistently outperforming benchmarks, the challenges lie in performance differences due to dataset specificity and a lack of detail on internal mechanisms. The research proposed by [Zheng et al. \(2020\)](#), data adaptive graph generation node adaptive and parameter learning modules for traffic prediction.

The technique achieves high-tech effectiveness, but challenges include a lack of thorough model evaluations and a comprehensive investigation of internal mechanisms. [Zhang et al. \(2020a\)](#) presents a mechanism to dwell on the issues of long-term graph convolutional networks and the differentiable architecture search for both spatial and temporal traffic forecasting. The study lacks a comprehensive analysis of the model, including the need to compare the model with existing state-of-the-art models. On the other hand, research by [Song et al. \(2020\)](#) proposes a spatio-temporal graph based on transformers for crowd trajectory prediction, relying exclusively on attention techniques. Although successful, limitations include potential degradation in predicting impulsive activity and challenges in handling constant time series sequential data. In summary, each study presents novel methods and frameworks for various time series prediction tasks, but these studies also express common limitations, such as problems of interpretability, model complexity, and limitations in dealing with external aspects or diverse datasets.

More comprehensive and comprehensive surveys are needed to improve the applicability and understanding of these techniques. The study in [Yu et al. \(2020\)](#) introduces a novel graph neural network with an output layer to focus on predicting graph

controls for graph-to-graph training. Graph controls or edits offer a formative and comprehensive illustration for innumerable changes in graph constructions, reviving computational competence and interpretability compared to prevailing approaches. Graph edits are well thought out and valuable in terms of sparsity. However, the method recognizes the limitation of the Markovian postulation, which implies imminent research to discover alternative frameworks. The need to fuse nonstructural data through alternative designs is also recognized. The scope of evaluation is inadequate and requires development in complex and diverse situations. The study by [Paassen et al. \(2020\)](#) presents graph convolutional networks with a time-aware zigzag topological layer for time-dependent graph neural networks.

The addition of zigzag trajectories increases the system's ability to predict time series on graphs, which means higher performance on different datasets. Nevertheless, its efficiency may be inadequate for time-dependent structures, and limitations related to mathematical context and generalizability are not systematically discovered. In [Chen, Segovia & Gel \(2021\)](#), a framework for predicting traffic flow has been developed using a data-driven approach to build dynamic graphs with spatio-temporal dependencies. Although the techniques are successively applied, their generalizability to other domains or tasks is not categorically established. The reliance on past data to make the spatio-temporal distance hyperparameter sensitive is illustrative and heralds possible robustness concerns. Research conducted in [Lan et al. \(2022\)](#) introduced a graph module known as a temporal polynomial to capture the correlation representation of dynamic variables in the prediction of multivariate multidimensional time series. It supports limitations associated to real-world procedures, causal construction, and causal knowledge. The emphasis on adhering to these limitations is projected for future work.

Furthermore, in [Liu et al. \(2022\)](#) the authors introduce a hybrid model that integrates spatio-temporal graphs with neural controlled differential equations for traffic prediction. While superior in performance to reference models, it has a clear deficiency in detailing its comparative study and intuitions into the hyper-parameter sensitivity and the interface between neural-controlled differential equations and graph-controlled networks. [Choi et al. \(2022\)](#) presents a pre-training technique called TSFormer for long-term evidence learning in GNN. It successfully addresses the method's limitation in capturing long-term dependencies. However, scalability concerns remain and the characterization of "long-term" needs further investigation. A study by [Shao et al. \(2022\)](#) discusses the prediction of variable subsets and presents a technique that uses a non-parametric wrapper.

Although competent, its performance is inclined by hyperparameters, and the postulation of variables preoccupied at casual may not embrace continuously ([Jin et al., 2024](#)). Scalability constraints associated with large datasets are not widely discovered. Graph structures in a regularized framework by [Chauhan et al. \(2022\)](#), integrate implicit and explicit graphical structures in a regularized framework to handle multivariate time series prediction. The technique's understanding of hyper-parameters and the configuration of cultivated graph structures with interpretability are accredited. While scalability remains unexplored. In the research conducted by [Yu et al. \(2022b\)](#), they introduce a framework that learns temporal graphs from multivariate time series and that



too at manifold scales ([Chai et al., 2024](#)), overcoming static variable importance modeling constraints. However, it lacks further analysis of interpretability, scalability, and computational complexity properties ([Li et al., 2024](#)). The work of [Cui et al. \(2021\)](#) presents an innovative structure for detecting anomalies in multivariate time series using the graph attention network. In the run through, each research presents novel methods to detect anomalies in multivariate time series ([Qin et al., 2024](#)), addressing numerous features such as relationship learning ([Guo et al., 2024](#)), graph-based modeling, and anomaly metric strategy. Although all the work presented acknowledges positive limitations, additional analysis and detailed exploration is needed to improve the reasoning behind these innovative approaches and their applicability in the real world.

## ENHANCING MODEL INTERPRETABILITY AND EXPLAINABILITY

Interpretability and explainability are important to ensure understanding and trust in AI models ([Bell et al., 2022](#)), particularly in the framework of convoluted constructions such as DNNs and GNNs. In this segment, we explore approaches intended to enhance the interpretability and explainability of such models, in view of current progressions and dedicated tools custom-made for DNNs and GNNs.

### Model simplification and feature importance

An approach to improving interpretability is to simplify the model techniques. For DNNs, practices such as distillation, feature reduction, and pruning can support the generation of more interpretable techniques without significantly affecting performance ([Gosiewska, Kozak & Biecek, 2021](#)). Likewise, in GNNs, custom-made generalization plans for graph structures can be helpful in understanding the model's decision-making method. Furthermore, feature importance methods such as SHapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) can deliver understanding into the role of distinct features to the model's predictions, by this means setting off interpretability ([Gosiewska, Kozak & Biecek, 2021](#)).

### Post-hoc explanation methods

The objective of these methods is to provide interpretable intuitions into model predictions without altering the fundamental architecture. Methods such as occlusion analysis, saliency maps, and gradient-based techniques can provide significant explanations for DNNs by stressing vital sections of input data that affect output ([Han, Srinivas & Lakkaraju, 2022](#)). In the same way, in GNNs, techniques like graph visualization mechanisms and graph attention techniques can clarify the rank of nodes and edges in decision-making methods.

### Attention techniques and layer-wise relevance propagation

Attention techniques have appeared as commanding tools to set the stage for interpretability in DNN models. By robust weighting of input features built on their importance to assignment, attention techniques offer spontaneous enlightenment for model predictions ([Achtibat et al., 2024](#)). With reference to GNNs, attention techniques

can help categorize significant nodes and edges in the graph, enabling an improved understanding of the model's performance. Another technique, layer-wise relevance propagation, points relevance scores to separate layers of the neural network, subject to understanding of how input features assist to the output.

### Counterfactual explanations

These explanations provide an exclusive perspective on model interpretability by supplying explanations for distinct predictions planted on hypothetical states. By making counterfactual instances that advance to diverse results, these practices help users recognize the decision limitations of the model and categorize actionable understandings to develop fairness and model performance ([Guidotti, 2022](#)).

### Tools for model interpretability

Quite a few dedicated libraries and tools have been established to enable model interpretability for DNNs and GNNs. For example, libraries including Captum and TensorFlow Explainability offer the use of many customized explanation methods for DNN models ([Kaur et al., 2020](#)). Similarly, tools such as the Deep Graph Library provide services for interpreting and visualizing GNNs, making them more accessible to practitioners and researchers. Integrating these tools and approaches into the evaluation and development of DNNs and GNNs can meaningfully enhance their explainability, interpretability, nurturing trust, and adoption in applications associated with the real world ([Bell et al., 2022](#)). As the field progresses, more research into customized innovative frameworks and techniques for multipart models will be important to endorse accountability and transparency in AI systems.

## ETHICAL CONSIDERATIONS, SECURITY AND DATA PRIVACY CHALLENGES

Although improving model interpretability and scalability is decisive, it is also imperative to consider ethical issues, security challenges, and data privacy, particularly in the context of time series analysis in industrial frameworks ([Albahri et al., 2023](#)). This section dives into the ethical inferences of implementing AI practices, the prerequisite for preserving data privacy, and recommends practices and best strategies to direct these compound issues reliably. Implementing AI techniques in industrial contexts can have reflective ethical consequences, mainly when decisions based on these techniques influence entities or communities ([de Almeida, dos Santos & Farias, 2021](#)). Biases in the unintended consequences, algorithmic fairness, and training data of model predictions are some serious ethical issues that require careful consideration ([Beil et al., 2019](#)). It is important to certify that models are accountable, fair and transparent to moderate latent harms and safeguard impartial consequences for all participants. Modern AI techniques, exclusively DNN architectures, regularly need immense data for training, raising concerns about data privacy and security ([Zareen, Akram & Ahmad Khan, 2020](#)). Time series in industrial frameworks can contain sensitive data, such as confidential business, personal identification, or actions of proprietary processes ([Díaz-Rodríguez et al., 2023](#)). Protecting



these data against breaches, misuse, or unlicensed access is paramount in ensuring compliance and trust with control frameworks such as the General Data Protection Regulation and the California Consumer Privacy Act ([Alexander, 2019](#)).

Best practices and guidelines that integrate data privacy, security, and ethical considerations in the deployment and development of AI models require a positive adherence and method to the best practices. For example: Documentation of data sources, model architecture, and decision-making processes needed to be present to support accountability and transparency. Reduce the retention and collection of personally identifiable information to minimize the risk of unauthorized access and privacy breaches ([Goldsteen et al., 2022](#)). Consider methods such as differential privacy and data anonymization to protect sensitive data while preserving the utility of the information for model training and analysis ([González-Sendino, Serrano & Bajo, 2024](#)). Ensure that these AI models are free from interpretable and explainable biases to encourage fairness and reduce the risk of accidental consequences ([Omotunde & Ahmed, 2023](#)). Implement robust security actions, including encryption, auditing, and access controls, to protect data against unauthorized access and cyber threats ([Borgman, 2018](#)). Acquire up-to-date consent from users concerning sharing practices, data collection, and usage, also make available transparency about how their data is being used to deploy and build AI models.

## OPEN QUESTIONS, CHALLENGES, AND PROPOSED SOLUTIONS

From our investigation, it is apparent that although the progress in time-series research is still quite extensive, there are still loads of open issues and research questions to be addressed. It is evident that further studies are needed in most of the industrial approaches in the context of time series. Hence, adding together to the already research work, the following are still complementary to open gaps in research.

### High-dimensional anomaly detection

Identifying anomalies in high dimensional data is challenging because of not identifying precise local correlations, thus using techniques such as UMAP, t-SNE, or autoencoders, can be employed to convert high-dimensional data into lower-dimensional illustrations, preserving important correlations whereas dropping computational complexity. In addition, algorithms like graph-based anomaly detection that take advantage from the correlations between dimensions can be investigated. Also, incorporating attention mechanisms with current anomaly detection techniques can vigorously spotlight significant features, focusing on the challenge of unsuitable feature selection.

### Ensemble methods

Traditional ML techniques such as clustering usually fail to provide robust outcomes for anomaly detection. Ensemble techniques, which merge outputs from multiple methods, deliver a favorable alternative. Hence, finding methods to choose the base learners and accurate subspaces is valuable. In addition, deciding on the right combination and quantity policies is still a challenging and open research problem to adopt. Implementing

heterogeneous ensemble learning a combination of diverse methods (*e.g.*, SVMs, decision trees, and neural networks) to leverage their original strengths. Stacking techniques or voting mechanisms can additionally improve anomaly detection accuracy. By developing adaptive ensemble techniques that dynamically choose subspaces and base learners constructed on data characteristics, enhancing both detection quality and runtime.

### **Evolving data dimensions**

The loss or arrival of existing or new dimensions during the period is yet another stimulating open research area for the future. In prospective application areas such as anomaly detection in IoT or industrial IoT devices, sensors can be on or off periodically, and innovative techniques are required to detect anomalies or outliers more competently in this challenging and stimulating scenario. By developing incremental learning algorithms that can easily adapt to modifications in data structure over different time periods. Methods such as adaptive LSTMs and online GNNs can deal with dynamic time-series data and evolving graph structure. Furthermore, incorporating techniques such as transfer learning can be adapted so that models trained on one set of dimensions to new data, reducing retraining struggles.

### **Scarcity of public real-world datasets**

Unlike data set sources for conventional ML problems for time series, publicly available real-world data sets for industrial data are very scarce. One key reason is that, in industry, data is believed to be extremely confidential—the data hold wear and tear or faults in their products reasonably for predictive maintenance. This has serious significance in the research work in this domain, because of public non availability of real-world data, research is classically steered by academia in association with industry, manufacturers, or suppliers, and in companies themselves. Hence, a non-efficiency in follow-up research and lack of reproducible research. Encouraging industry-academia cooperation to initiate anonymized datasets however making certain and preserving data privacy. Also, employing data synthesis approaches, such as generative adversarial networks (GANs), to make realistic industrial data for research. Exploring domain adaptation techniques that can transfer knowledge from accessible datasets to unseen, new industrial domains, minimizing reliance on explicit datasets.

### **Lack of labeled data**

There is a lack of labeled data in industry, operating with datasets of real-world industry propose the prospect of being capable of assessing developed techniques in practice. Consequently, real-world data is frequently not or only incompletely labelled, since to annotate this volume of data requires expert knowledge and is hugely time-consuming. However, employing self-supervised learning and semi-supervised learning approaches to take advantage of unlabeled data. As well, techniques like pseudo-labeling and contrastive learning can improve model performance without the need of extensive labeling. In

addition, developing annotation bases employing active learning, where models rank first the most useful data points for skilled labeling, optimizing annotation attempts.

### Representation of training data

Finding representative training datasets is quite challenging due to the scale and variety of industrial applications and its products. Using simulated data from software-in-the-loop systems to accompany real-world data. Exploring domain-specific data augmentation and GANs approaches to improve dataset representativeness and diversity. Designing algorithms that are competent of leveraging synthetic data while easing domain differences using methods like domain adversarial neural networks (DANNs).

### Model interpretability

The issue with non-interpretability of complex ML models demands details for the replacement of components. With maintenance that is interval-based the justification is insignificant; conversely, with utilizing advanced ML techniques, *e.g.*, deep learning models, interpretation and explanation of the techniques and their choices turn into a challenge. By utilizing Explainable AI (XAI) methods such as LIME, SHAP, and attention visualization to present understandings into model decisions. Aim on acquiring interpretable designs like prototype-based learning models. Coming to GNNs, it is quite challenging to interpret them critical industrial systems because of opaque decision process, regulatory compliance, and trust in automation. For instance, to trace certain decisions input will be quite difficult in GNNs, in healthcare or energy industry they demand XAI for accountability, and industrial stakeholders require transparency to trust GNNs. However, we can use attention mechanisms for influential nodes and edges in GNNs, also the use of XAI techniques as discussed earlier is advisable, and employ graph visualization tools. Furthermore, exploring hybrid techniques which combines interpretable ML approaches with deep learning to balance explainability and accuracy. For example, incorporate rule-based systems or decision trees with neural networks to increase interpretability.

By focusing on these challenges with far-forward techniques and algorithms, the research society can substantially augment the scalability, robustness, and applicability of time-series in industrial frameworks.

## CONCLUSION

In summary, this comprehensive survey acts as a guide to the vast field of time-series analysis within industrial data. It starts with the basic principles of traditional machine learning, which provided insights for forecasting, demand management, and production optimization. As the industrial landscape evolved, the integration of deep learning architectures led to a transformative phase. ML, DL, and GNN approaches have shown great promise in various fields, as well as in time series. To run through the past, investigate the present, and analyze the future, in this survey, we made several contributions to the application of ML, DL, and GNNs theory in the domain of time series. The merits and

limitations of DL and GNN were evaluated in detail compared to traditional ML approaches. In addition, the tendencies of industrial frameworks were considered in detail to explain the significance and necessity of AI algorithms for time-series modeling. Key ML, DL, and GNN models and frameworks were summarized and reviewed to facilitate the use and development of AI-based tools. Despite the contributions of ML, DL, and GNNs limitations, exits that open doors for further investigation. With DL and GNNs interpretability remains the issue, their “black-box” nature remains a significant challenge, methods should focus on integrating XAI techniques. GNNs hold promise, but it has issues in adaptability and scalability in various industrial scenarios, the focus should be shifted to domain specific adaptations. The scarcity of labeled, publicly available and representative industrial data hinders benchmarking and reproducibility; addressing this problem requires collaborative efforts between industry and academia. Promising research questions and points for future run were explored shortly. However, challenges remain, including concerns about the interpretability of deep learning models and the need for graph-based solutions tailored to different industrial applications.

## ACKNOWLEDGEMENTS

We have used ChatGPT for sentence structuring and any grammatical mistakes.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The work received funding from the European Union Horizon 2020 research innovation, the SDGine program under the Marie Skłodowska-Curie grant agreement number 945139. This work was also funded by the European Commission through the HORIZON-JU-SNS-2022 ACROSS project with Grant Agreement number 101097122. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

European Union Horizon 2020 Research Innovation.

SDGine Program under the Marie Skłodowska-Curie: 945139.

European Commission through the HORIZON-JU-SNS-2022 ACROSS: 101097122.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Muhammad Jamal Ahmed conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Alberto Mozo conceived and designed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Amit Karamchandani analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

This is a literature review.

## REFERENCES

- Abdeldayem OM, Dabbish AM, Habashy MM, Mostafa MK, Elhefnawy M, Amin L, Al-Sakkari EG, Ragab A, Rene ER. 2022. Viral outbreaks detection and surveillance using wastewater-based epidemiology, viral air sampling, and machine learning techniques: a comprehensive review and outlook. *Science of the Total Environment* **803**(23):149834 DOI [10.1016/j.scitotenv.2021.149834](https://doi.org/10.1016/j.scitotenv.2021.149834).
- Achtibat R, Hatefi SMV, Dreyer M, Jain A, Wiegand T, Lapuschkin S, Samek W. 2024. AttnLRP: attention-aware layer-wise relevance propagation for transformers. ArXiv DOI [10.48550/arXiv.2402.05602](https://doi.org/10.48550/arXiv.2402.05602).
- Ahmed MJ, Saeed F, Paul A, Jan S, Seo H. 2021. A new affinity matrix weighted k-nearest neighbors graph to improve spectral clustering accuracy. *PeerJ Computer Science* **7**:e692 DOI [10.7717/peerj-cs.692](https://doi.org/10.7717/peerj-cs.692).
- Akay S, Atapour-Abarghouei A, Breckon TP. 2019. GANomaly: semi-supervised anomaly detection via adversarial training. In: Jawahar CV, Li H, Mori G, Schindler K, eds. *Computer Vision—ACCV 2018, Lecture Notes in Computer Science*. Cham: Springer International Publishing, 622–637.
- Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoodi AH, Bai J, Salhi A, Santamaría J, Ouyang C, Gupta A, Gu Y, Devci M. 2023. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. *Information Fusion* **96**(10):156–191 DOI [10.1016/j.inffus.2023.03.008](https://doi.org/10.1016/j.inffus.2023.03.008).
- Aldaheri A, Alwahedi F, Ferrag MA, Battah A. 2024. Deep learning for cyber threat detection in IoT networks: a review. *Internet of Things and Cyber-Physical Systems* **4**(3):110–128 DOI [10.1016/j.iotcps.2023.09.003](https://doi.org/10.1016/j.iotcps.2023.09.003).
- Alexander CB. 2019. The general data protection regulation and California consumer privacy act: the economic impact and future of data privacy regulations. *Loyola Consumer Law Review* **32**(Issue 2):199 DOI [10.54648/gplr2020080](https://doi.org/10.54648/gplr2020080).
- Al-Garadi MA, Mohamed A, Al-Ali AK, Du X, Ali I, Guizani M. 2020. A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials* **22**(3):1646–1685 DOI [10.1109/COMST.2020.2988293](https://doi.org/10.1109/COMST.2020.2988293).
- Ali Nemer M, Azar J, Demerjian J, Makhoul A, Bourgeois J. 2022. A review of research on industrial time series classification for machinery based on deep learning. In: *2022 4th IEEE Middle East and North Africa COMMUNICATIONS Conference (MENACOMM)*. Piscataway: IEEE, 89–94.
- Angelov PP, Gu X. 2019. Anomaly detection—empirical approach. In: Angelov PP, Gu X, eds. *Empirical Approach to Machine Learning, Studies in Computational Intelligence*. Cham: Springer International Publishing, 157–173.

- Ao S-I, Fayek H. 2023. Continual deep learning for time series modeling. *Sensors* 23(16):7167 DOI 10.3390/s23167167.
- Aryal K, Gupta M, Abdelsalam M. 2022. A survey on adversarial attacks for malware analysis. ArXiv DOI 10.48550/arXiv.2111.08223.
- Avci O, Abdeljaber O, Kiranyaz S, Hussein M, Gabbouj M, Inman DJ. 2021. A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing* 147(2):107077 DOI 10.1016/j.ymssp.2020.107077.
- Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. 2019. Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental* 7:70 DOI 10.1186/s40635-019-0286-6.
- Bell A, Solano-Kamaiko I, Nov O, Stoyanovich J. 2022. It's just not that simple: an empirical study of the accuracy-explainability trade-off in machine learning for public policy. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22*. New York, NY, USA: Association for Computing Machinery, 248–266.
- Bengio Y. 2009. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning* 2(1):1–127 DOI 10.1561/22000000006.
- Bertolini M, Mezzogori D, Neroni M, Zammori F. 2021. Machine learning for industrial applications: a comprehensive literature review. *Expert Systems with Applications* 175(6):114820 DOI 10.1016/j.eswa.2021.114820.
- Binkowski M, Marti G, Donnat P. 2018. Autoregressive convolutional neural networks for asynchronous time series. In: *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 580–589.
- Boardman J. 2022. A new kind of data science: the need for ethical analytics. Published and Grey Literature from PhD Candidates. Available at <https://digitalcommons.kennesaw.edu/dataphdgreylit/33>.
- Boddington P. 2023. *AI ethics: a textbook*. Cham: Springer Nature.
- Borgman CL. 2018. Open data, grey data, and stewardship: universities at the privacy frontier. *Berkeley Technology Law Journal* 33(2):365–412 DOI 10.48550/arXiv.1802.02953.
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015. *Time series analysis: forecasting and control*. Hoboken: John Wiley & Sons.
- Bui DT, Tsangaratos P, Nguyen V-T, Liem NV, Trinh PT. 2020. Comparing the prediction performance of a deep learning neural network model with conventional machine learning models in landslide susceptibility assessment. *CATENA* 188(18):104426 DOI 10.1016/j.catena.2019.104426.
- Canizo M, Triguero I, Conde A, Onieva E. 2019. Multi-head CNN–RNN for multi-time series anomaly detection: an industrial case study. *Neurocomputing* 363(3):246–260 DOI 10.1016/j.neucom.2019.07.034.
- Carvalho DV, Pereira EM, Cardoso JS. 2019. Machine learning interpretability: a survey on methods and metrics. *Electronics* 8(8):832 DOI 10.3390/electronics8080832.
- Chai S, Wang S, Liu C, Liu X, Liu T, Yang R. 2024. A visual measurement algorithm for vibration displacement of rotating body using semantic segmentation network. *Expert Systems with Applications* 237:121306 DOI 10.1016/j.eswa.2023.121306.
- Chauhan J, Raghuveer A, Saket R, Nandy J, Ravindran B. 2022. Multi-variate time series forecasting on variable subsets. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22*. New York, NY, USA: Association for Computing Machinery, 76–86.



- Chen W, Chen L, Xie Y, Cao W, Gao Y, Feng X. 2020. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(4):3529–3536 DOI 10.1609/aaai.v34i04.5758.
- Chen Z, Chen D, Zhang X, Yuan Z, Cheng X. 2022c. Learning graph structures with transformer for multivariate time-series anomaly detection in IoT. *IEEE Internet of Things Journal* 9(12):9179–9189 DOI 10.1109/jiot.2021.3100509.
- Chen K, Feng M, Wirjanto TS. 2023. Multivariate time series anomaly detection via dynamic graph forecasting. ArXiv DOI 10.48550/arXiv.2302.02051.
- Chen Y, Segovia I, Gel YR. 2021. Z-GCNETs: time zigzags at graph convolutional networks for time series forecasting. In: *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 1684–1694.
- Chen W, Tian L, Chen B, Dai L, Duan Z, Zhou M. 2022a. Deep variational graph convolutional recurrent network for multivariate time series anomaly detection. In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 3621–3633.
- Chen W, Zhou Z, Wen Q, Sun L. 2022b. Time series subsequence anomaly detection via graph neural networks. Available at [https://openreview.net/forum?id=73U\\_NIKaNx](https://openreview.net/forum?id=73U_NIKaNx).
- Cheng Z, Yang Y, Jiang S, Hu W, Ying Z, Chai Z, Wang C. 2023. Time2Graph+: bridging time series and graph representation learning via multiple attentions. *IEEE Transactions on Knowledge and Data Engineering* 35(2):2078–2090 DOI 10.1109/tkde.2021.3094908.
- Chevrot A, Vernotte A, Legeard B. 2022. CAE: contextual auto-encoder for multivariate time-series anomaly detection in air transportation. *Computers & Security* 116(5):102652 DOI 10.1016/j.cose.2022.102652.
- Choi J, Choi H, Hwang J, Park N. 2022. Graph neural controlled differential equations for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 36(6):6367–6374 DOI 10.1609/aaai.v36i6.20587.
- Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, Park CW, Choudhary A, Agrawal A, Billinge SJL, Holm E, Ong SP, Wolverton C. 2022. Recent advances and applications of deep learning methods in materials science. *NPJ Computational Materials* 8:59 DOI 10.1038/s41524-022-00734-6.
- Christ M, Kempa-Liehr AW, Feindt M. 2017. Distributed and parallel time series feature extraction for industrial big data applications. ArXiv DOI 10.48550/arXiv.1610.07717.
- Cui Y, Zheng K, Cui D, Xie J, Deng L, Huang F, Zhou X. 2021. METRO: a generic graph neural network framework for multivariate time series forecasting. *Proceedings of the VLDB Endowment* 15(2):224–236 DOI 10.14778/3489496.3489503.
- Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. *Ecology* 88(11):2783–2792 DOI 10.1890/07-0539.1.
- Dai E, Chen J. 2022. Graph-augmented normalizing flows for anomaly detection of multiple time series. ArXiv DOI 10.48550/arXiv.2202.07857.
- Dambros JWV, Trierweiler JO, Farenzena M, Kloft M. 2019. Oscillation detection in process industries by a machine learning-based approach. *Industrial & Engineering Chemistry Research* 58(31):14180–14192 DOI 10.1021/acs.iecr.9b01456.
- Das M, Ghosh SK. 2019. FB-STEP: a fuzzy Bayesian network based data-driven framework for spatio-temporal prediction of climatological time series data. *Expert Systems with Applications* 117(12):211–227 DOI 10.1016/j.eswa.2018.08.057.
- de Almeida PGR, dos Santos CD, Farias JS. 2021. Artificial intelligence regulation: a framework for governance. *Ethics and Information Technology* 23(3):505–525 DOI 10.1007/s10676-021-09593-z.

- De Livera AM, Hyndman RJ, Snyder RD. 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* 106(496):1513–1527 DOI 10.1198/jasa.2011.tm09771.
- Deng A, Hooi B. 2021. Graph neural network-based anomaly detection in multivariate time series. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(5):4027–4035 DOI 10.1609/aaai.v35i5.16523.
- Díaz-Rodríguez N, Del Ser J, Coeckelbergh M, López de Prado M, Herrera-Viedma E, Herrera F. 2023. Connecting the dots in trustworthy artificial intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion* 99(2):101896 DOI 10.1016/j.inffus.2023.101896.
- Drakaki M, Karnavas YL, Tzifafettas IA, Linardos V, Tzionas P. 2022. Machine learning and deep learning based methods toward industry 4.0 predictive maintenance in induction motors: state of the art survey. *Journal of Industrial Engineering and Management (JIEM)* 15(1):31–57 DOI 10.3926/jiem.3597.
- Du P. 2019. Ensemble machine learning-based wind forecasting to combine NWP output with data from weather station. *IEEE Transactions on Sustainable Energy* 10(4):2133–2141 DOI 10.1109/tste.2018.2880615.
- Duan Z, Xu H, Wang Y, Huang Y, Ren A, Xu Z, Sun Y, Wang W. 2022. Multivariate time-series classification with hierarchical variational graph pooling. *Neural Networks* 154(Nov):481–490 DOI 10.1016/j.neunet.2022.07.032.
- Fahim M, Sillitti A. 2019. Anomaly detection, analysis and prediction techniques in IoT environment: a systematic literature review. *IEEE Access* 7:81664–81681 DOI 10.1109/access.2019.2921912.
- Fährmann D, Damer N, Kirchbuchner F, Kuijper A. 2022. Lightweight long short-term memory variational auto-encoder for multivariate time series anomaly detection in industrial control systems. *Sensors* 22(8):2886 DOI 10.3390/s22082886.
- Fernandes M, Corchado JM, Marreiros G. 2022. Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. *Applied Intelligence* 52(12):14246–14280 DOI 10.1007/s10489-022-03344-3.
- Gamage S, Kloppe B, Samarabandu J. 2022. Experiences with contrastive predictive coding in industrial time-series classification. *ACM SIGKDD Explorations Newsletter* 24(2):114–123 DOI 10.1145/3575637.3575654.
- Gao J, Ribeiro B. 2022. On the equivalence between temporal and static equivariant graph representations. In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 7052–7076.
- Gaugel S, Reichert M. 2023a. Industrial transfer learning for multivariate time series segmentation: a case study on hydraulic pump testing cycles. *Sensors* 23(7):3636 DOI 10.3390/s23073636.
- Gaugel S, Reichert M. 2023b. PrecTime: a deep learning architecture for precise time series segmentation in industrial manufacturing operations. *Engineering Applications of Artificial Intelligence* 122(5):106078 DOI 10.1016/j.engappai.2023.106078.
- Geng Z, Chen Z, Meng Q, Han Y. 2022. Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes. *IEEE Transactions on Industrial Informatics* 18(3):1521–1529 DOI 10.1109/tii.2021.3086798.
- Gewers FL, Ferreira GR, Arruda HFD, Silva FN, Comin CH, Amancio DR, Costa LDF. 2021. Principal component analysis: a natural approach to data exploration. *ACM Computing Surveys* 54(4):70:1–70:34 DOI 10.1145/3447755.

- Goldsteen A, Ezov G, Shmelkin R, Moffie M, Farkash A. 2022.** Anonymizing machine learning models. In: Garcia-Alfaro J, Muñoz-Tapia JL, Navarro-Arribas G, Soriano M, eds. *Data Privacy Management, Cryptocurrencies and Blockchain Technology*. Cham: Springer International Publishing, 121–136.
- González-Sendino R, Serrano E, Bajo J. 2024.** Mitigating bias in artificial intelligence: fair data generation via causal models for transparent and explainable decision-making. *Future Generation Computer Systems* 155:384–401 DOI 10.1016/j.future.2024.02.023.
- Gosiewska A, Kozak A, Biecek P. 2021.** Simpler is better: lifting interpretability-performance trade-off via automated feature engineering. *Decision Support Systems* 150:113556 DOI 10.1016/j.dss.2021.113556.
- Grattarola D, Zambon D, Livi L, Alippi C. 2020.** Change detection in graph streams by learning graph embeddings on constant-curvature manifolds. *IEEE Transactions on Neural Networks and Learning Systems* 31(6):1856–1869 DOI 10.1109/tnnls.2019.2927301.
- Graves A. 2014.** Generating sequences with recurrent neural networks. ArXiv DOI 10.48550/arXiv.1308.0850.
- Guan S, Zhao B, Dong Z, Gao M, He Z. 2022.** GTAD: graph and temporal neural network for multivariate time series anomaly detection. *Entropy* 24(6):759 DOI 10.3390/e24060759.
- Guidotti R. 2022.** Counterfactual explanations and how to find them: literature review and benchmarking. *Data Mining and Knowledge Discovery* 38(5):2770–2824 DOI 10.1007/s10618-022-00831-6.
- Guigou F. 2019.** The artificial immune ecosystem: a scalable immune-inspired active classifier, an application to streaming time series analysis for network monitoring. PhD Thesis. Université de Strasbourg.
- Guo S, Lin Y, Feng N, Song C, Wan H. 2019.** Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 33(1):922–929 DOI 10.1609/aaai.v33i01.3301922.
- Guo T, Yuan H, Hamzaoui R, Wang X, Wang L. 2024.** Dependence-based coarse-to-fine approach for reducing distortion accumulation in G-PCC attribute compression. *IEEE Transactions on Industrial Informatics* 20(9):11393–11403 DOI 10.1109/tii.2024.3403262.
- Han T, Srinivas S, Lakkaraju H. 2022.** Which explanation should I choose? A function approximation perspective to characterizing post hoc explanations. *Advances in Neural Information Processing Systems* 35:5256–5268.
- Han S, Woo SS. 2022.** Learning sparse latent graph representations for anomaly detection in multivariate time series. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22*. New York, NY, USA: Association for Computing Machinery, 2977–2986.
- Herrmann M, Webb GI. 2021.** Amercing: an intuitive, elegant and effective constraint for dynamic time warping. ArXiv DOI 10.48550/arXiv.2111.13314.
- Hewage P, Behera A, Trovati M, Pereira E, Ghahremani M, Palmieri F, Liu Y. 2020.** Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing* 24(21):16453–16482 DOI 10.1007/s00500-020-04954-0.
- Hinton GE. 2007.** Boltzmann machine. *Scholarpedia* 2(5):1668 DOI 10.4249/scholarpedia.1668.
- Hsu C-Y, Liu W-C. 2021.** Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. *Journal of Intelligent Manufacturing* 32(3):823–836 DOI 10.1007/s10845-020-01591-0.

- Idowu S, Strüber D, Berger T. 2021.** Asset management in machine learning: a survey. In: *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. Piscataway: IEEE, 51–60.
- Iglesias G, Talavera E, González-Prieto Á, Mozo A, Gómez-Canaval S. 2023.** Data augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications* **35(14)**:10123–10145 DOI [10.1007/s00521-023-08459-3](https://doi.org/10.1007/s00521-023-08459-3).
- Jain AK, Nandakumar K, Ross A. 2016.** 50 years of biometric research: accomplishments, challenges, and opportunities. *Pattern Recognition Letters* **79(2)**:80–105 DOI [10.1016/j.patrec.2015.12.013](https://doi.org/10.1016/j.patrec.2015.12.013).
- Jaton F. 2021.** Assessing biases, relaxing moralism: on ground-truthing practices in machine learning design and application. *Big Data & Society* **8(1)**:205395172110135 DOI [10.1177/20539517211013569](https://doi.org/10.1177/20539517211013569).
- Jatti VS, Sekhar R, Shah P. 2021.** Machine learning based predictive modeling of ball nose end milling using exogeneous autoregressive moving average approach. In: *2021 IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*. Piscataway: IEEE, 68–72.
- Jiang W, Cheng C, Zhou B, Ma G, Yuan Y. 2019.** A novel GAN-based fault diagnosis approach for imbalanced industrial time series. *IEEE Access* **7**:143608–143619 DOI [10.1109/access.2019.2944689](https://doi.org/10.1109/access.2019.2944689).
- Jiang Y, Yin S, Dong J, Kaynak O. 2021.** A review on soft sensors for monitoring, control, and optimization of industrial processes. *IEEE Sensors Journal* **21(11)**:12868–12881 DOI [10.1109/jsen.2020.3033153](https://doi.org/10.1109/jsen.2020.3033153).
- Jiao S, Gao Y, Feng J, Lei T, Yuan X. 2020.** Does deep learning always outperform simple linear regression in optical imaging? *Optics Express* **28(3)**:3717–3731 DOI [10.1364/oe.382319](https://doi.org/10.1364/oe.382319).
- Jin W, Tian X, Shi B, Zhao B, Duan H, Wu H. 2024.** Enhanced UAV pursuit-evasion using boids modelling: a synergistic integration of bird swarm intelligence and DRL. *Computers, Materials & Continua* **80(3)**:3523–3553 DOI [10.32604/cmc.2024.055125](https://doi.org/10.32604/cmc.2024.055125).
- Kadlec P, Gabrys B, Strandt S. 2009.** Data-driven soft sensors in the process industry. *Computers & Chemical Engineering* **33(4)**:795–814 DOI [10.1016/j.compchemeng.2008.12.012](https://doi.org/10.1016/j.compchemeng.2008.12.012).
- Kanawaday A, Sane A. 2017.** Machine learning for predictive maintenance of industrial machines using IoT sensor data. In: *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. Piscataway: IEEE, 87–90.
- Kano M, Ogawa M. 2010.** The state of the art in chemical process control in Japan: good practice and questionnaire survey. *Journal of Process Control* **20(9)**:969–982 DOI [10.1016/j.jprocont.2010.06.013](https://doi.org/10.1016/j.jprocont.2010.06.013).
- Karamchandani A, Mozo A, Vakaruk S, Gómez-Canaval S, Sierra-García JE, Pastor A. 2023.** Using N-BEATS ensembles to predict automated guided vehicle deviation. *Applied Intelligence* **53(21)**:26139–26204 DOI [10.1007/s10489-023-04820-0](https://doi.org/10.1007/s10489-023-04820-0).
- Karl Pearson FRS. 1901.** LIII. On lines and planes of closest fit to systems of points in space. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. Oxfordshire: Taylor & Francis Group.
- Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J. 2020.** Interpreting interpretability: understanding data scientists’ use of interpretability tools for machine learning. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI ’20*. New York, NY, USA: Association for Computing Machinery.

- Kim B, Alawami MA, Kim E, Oh S, Park J, Kim H. 2023. A comparative study of time series anomaly detection models for industrial control systems. *Sensors* **23**(3):1310 DOI 10.3390/s23031310.
- Kingma DP, Welling M. 2022. Auto-encoding variational Bayes. ArXiv DOI 10.48550/arXiv.1312.6114.
- Koste LL, Malhotra MK. 1999. A theoretical framework for analyzing the dimensions of manufacturing flexibility. *Journal of Operations Management* **18**(1):75–93 DOI 10.1016/s0272-6963(99)00010-8.
- Lagunas M, Garcés E. 2017. Transfer learning for illustration classification. In: *Spanish Computer Graphics Conference (CEIG)*, 9 DOI 10.48550/arXiv.1806.02682.
- Lan S, Ma Y, Huang W, Wang W, Yang H, Li P. 2022. DSTAGNN: dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 11906–11917.
- Lee K, Lee H, Lee K, Shin J. 2018. Training confidence-calibrated classifiers for detecting out-of-distribution samples. ArXiv DOI 10.48550/arXiv.1711.09325.
- Li Y. 2021. A fault prediction and cause identification approach in complex industrial processes based on deep learning. *Computational Intelligence and Neuroscience* **2021**:e6612342 DOI 10.1155/2021/6612342.
- Li T, Hui S, Zhang S, Wang H, Zhang Y, Hui P, Jin D, Li Y. 2024. Mobile user traffic generation via multi-scale hierarchical GAN. *ACM Transactions on Knowledge Discovery from Data* **18**(8):1–19 DOI 10.1145/3664655.
- Li R, Wu X-J, Xu T. 2022. Video is graph: structured graph module for video action recognition. ArXiv DOI 10.48550/arXiv.2110.05904.
- Li Y, Yu R, Shahabi C, Liu Y. 2018. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. ArXiv DOI 10.48550/arXiv.1707.01926.
- Lima ALdCD, Aranha VM, Carvalho CJdL, Nascimento EGS. 2021. Smart predictive maintenance for high-performance computing systems: a literature review. *The Journal of Supercomputing* **77**(11):13494–13513 DOI 10.1007/s11227-021-03811-7.
- Lines J, Taylor S, Bagnall A. 2018. Time series classification with HIVE-COTE: the hierarchical vote collective of transformation-based ensembles. *ACM Transactions on Knowledge Discovery from Data* **12**(5):52:1–52:35 DOI 10.1145/3182382.
- Liu C-L, Hsiao W-H, Tu Y-C. 2019. Time series classification with multivariate convolutional neural network. *IEEE Transactions on Industrial Electronics* **66**(6):4788–4797 DOI 10.1109/tie.2018.2864702.
- Liu Y, Liu Q, Zhang J-W, Feng H, Wang Z, Zhou Z, Chen W. 2022. Multivariate time-series forecasting with temporal polynomial graph neural networks. *Advances in Neural Information Processing Systems* **35**:19414–19426.
- Liu FT, Ting KM, Zhou Z-H. 2012. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data* **6**(1):3:1–3:39 DOI 10.1145/2133360.2133363.
- Liu G, Zhu L, Yu W, Yu W. 2021. Image formation, deep learning, and physical implication of multiple time-series one-dimensional signals: method and application. *IEEE Transactions on Industrial Informatics* **17**(7):4566–4574 DOI 10.1109/tii.2020.3018496.
- Makridakis S, Hibon M. 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* **16**(4):451–476 DOI 10.1016/s0169-2070(00)00057-1.
- Maron RC, Hagggenmüller S, von Kalle C, Utikal JS, Meier F, Gellrich FF, Hauschild A, French LE, Schlaak M, Ghoreschi K, Kutzner H, Heppt MV, Haferkamp S, Sondermann W,



- Schadendorf D, Schilling B, Hekler A, Krieghoff-Henning E, Kather JN, Fröhling S, Lipka DB, Brinker TJ. 2021. Robustness of convolutional neural networks in recognition of pigmented skin lesions. *European Journal of Cancer* 145:81–91 DOI 10.1016/j.ejca.2020.11.020.
- Miao H, Li B, Sun C, Liu J. 2019. Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks. *IEEE Transactions on Industrial Informatics* 15(9):5023–5032 DOI 10.1109/tii.2019.2900295.
- Mozo A, Vakaruk S, Sierra-García JE, Pastor A. 2024. Anticipatory analysis of AGV trajectory in a 5G network using machine learning. *Journal of Intelligent Manufacturing* 35(4):1541–1569 DOI 10.1007/s10845-023-02116-1.
- Nguyen H-P, Liu J, Zio E. 2020. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen estimator and applied to time-series data of NPP steam generators. *Applied Soft Computing* 89(16–18):106116 DOI 10.1016/j.asoc.2020.106116.
- Olivieri M, Colleoni E, Bonaccorso G. 2023. How have travelers’ needs evolved because of the COVID-19 pandemic? Corporate reputation building in tourism industry on digital platforms. In: Rialti R, Kvítková Z, Makovník T, eds. *Online Reputation Management in Destination and Hospitality*. Leeds: Emerald Publishing Limited, 51–71.
- Omotunde H, Ahmed M. 2023. A comprehensive review of security measures in database systems: assessing authentication, access control, and beyond. *Mesopotamian Journal of CyberSecurity* 2023:115–133 DOI 10.58496/mjcs/2023/016.
- Paassen B, Grattarola D, Zambon D, Alippi C, Hammer BE. 2020. Graph edit networks. Available at [https://openreview.net/forum?id=dIEJsyHGGeL&trk=public\\_post\\_comment-text](https://openreview.net/forum?id=dIEJsyHGGeL&trk=public_post_comment-text).
- Palit AK, Popovic D. 2006. *Computational intelligence in time series forecasting: theory and engineering applications*. Cham: Springer Science & Business Media.
- Pan Z, Liang Y, Wang W, Yu Y, Zheng Y, Zhang J. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD ’19*. New York, NY, USA: Association for Computing Machinery, 1720–1730.
- Papenmeier A, Englebienne G, Seifert C. 2019. How model accuracy and explanation fidelity influence user trust. ArXiv.
- Pinson P. 2013. Wind energy: forecasting challenges for its operational management. *Statistical Science* 28(4):564–585 DOI 10.1214/13-sts445.
- Qian Y, Vaddiraju S, Khan F. 2023. Safety education 4.0—a critical review and a response to the process industry 4.0 need in chemical engineering curriculum. *Safety Science* 161(12):106069 DOI 10.1016/j.ssci.2023.106069.
- Qin C, Shi G, Tao J, Yu H, Jin Y, Xiao D, Liu C. 2024. RCLSTMNet: a residual-convolutional-LSTM neural network for forecasting cutterhead torque in shield machine. *International Journal of Control, Automation and Systems* 22(2):705–721 DOI 10.1007/s12555-022-0104-x.
- Rana N, Latiff MSA, Abdulhamid SM, Chiroma H. 2020. Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments. *Neural Computing and Applications* 32(20):16245–16277 DOI 10.1007/s00521-020-04849-z.
- Schockaert C. 2020. A causal-based framework for multimodal multivariate time series validation enhanced by unsupervised deep learning as an enabler for industry 4.0. ArXiv DOI 10.48550/arXiv.2008.02171.
- Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. 2001. Estimating the support of a high-dimensional distribution. *Neural Computation* 13(7):1443–1471 DOI 10.1162/089976601750264965.



- Schubert D, Gupta P, Wever M. 2023.** Meta-learning for automated selection of anomaly detectors for semi-supervised datasets. In: Crémilleux B, Hess S, Nijssen S, eds. *Advances in Intelligent Data Analysis XXI, Lecture Notes in Computer Science*. Cham: Springer Nature Switzerland, 392–405.
- Shalev-Shwartz S, Shammah S, Shashua A. 2018.** On a formal model of safe and scalable self-driving cars. ArXiv DOI [10.48550/arXiv.1708.06374](https://doi.org/10.48550/arXiv.1708.06374).
- Shao Z, Zhang Z, Wang F, Xu Y. 2022.** Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22*. New York, NY, USA: Association for Computing Machinery, 1567–1577.
- Shojaee P, Zeng Y, Chen X, Jin R, Deng X, Zhang C. 2021.** Deep neural network pipelines for multivariate time series classification in smart manufacturing. In: *2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*. Piscataway: IEEE, 98–103.
- Shrestha YR, Krishna V, von Krogh G. 2021.** Augmenting organizational decision-making with deep learning algorithms: principles, promises, and challenges. *Journal of Business Research* **123**(2):588–603 DOI [10.1016/j.jbusres.2020.09.068](https://doi.org/10.1016/j.jbusres.2020.09.068).
- Shumway RH, Stoffer DS. 2017.** Characteristics of time series. In: Shumway RH, Stoffer DS, eds. *Time Series Analysis and Its Applications: With R Examples, Springer Texts in Statistics*. Cham: Springer International Publishing.
- Simkute A, Luger E, Jones B, Evans M, Jones R. 2021.** Explainability for experts: a design framework for making algorithms supporting expert decisions more explainable. *Journal of Responsible Technology* **7–8**(3):100017 DOI [10.1016/j.jrt.2021.100017](https://doi.org/10.1016/j.jrt.2021.100017).
- Smith ML, Smith LN, Hansen MF. 2021.** The quiet revolution in machine vision—a state-of-the-art survey paper, including historical review, perspectives, and future directions. *Computers in Industry* **130**(2):103472 DOI [10.1016/j.compind.2021.103472](https://doi.org/10.1016/j.compind.2021.103472).
- Song C, Lin Y, Guo S, Wan H. 2020.** Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* **34**(1):914–921 DOI [10.1609/aaai.v34i01.5438](https://doi.org/10.1609/aaai.v34i01.5438).
- Srinivas SS, Kumar Sarkar R, Runkana V. 2022.** Hypergraph learning based recommender system for anomaly detection, control and optimization. In: *2022 IEEE International Conference on Big Data (Big Data)*. Piscataway: IEEE, 1922–1929.
- Taheri A, Gimpel K, Berger-Wolf T. 2019.** Learning to represent the evolution of dynamic graphs with recurrent models. In: *Companion Proceedings of the 2019 World Wide Web Conference, WWW '19*. New York, NY, USA: Association for Computing Machinery, 301–307.
- Tseng M-L, Tran TPT, Ha HM, Bui T-D, Lim MK. 2021.** Sustainable industrial and operation engineering trends and challenges toward industry 4.0: a data driven analysis. *Journal of Industrial and Production Engineering* **38**(8):581–598 DOI [10.1080/21681015.2021.1950227](https://doi.org/10.1080/21681015.2021.1950227).
- Tziolas T, Papageorgiou K, Theodosiou T, Papageorgiou E, Mastos T, Papadopoulos A. 2022.** Autoencoders for anomaly detection in an industrial multivariate time series dataset. *Engineering Proceedings* **18**(1):23 DOI [10.3390/engproc2022018023](https://doi.org/10.3390/engproc2022018023).
- Ullah I, Mahmoud QH. 2021.** A framework for anomaly detection in IoT networks using conditional generative adversarial networks. *IEEE Access* **9**:165907–165931 DOI [10.1109/access.2021.3132127](https://doi.org/10.1109/access.2021.3132127).
- Vakaruk S, Karamchandani A, Sierra-García JE, Mozo A, Gómez-Canaval S, Pastor A. 2023.** Transformers for multi-horizon forecasting in an industry 4.0 use case. *Sensors* **23**(7):3516 DOI [10.3390/s23073516](https://doi.org/10.3390/s23073516).

- Vakaruk S, Sierra-García JE, Mozo A, Pastor A. 2021.** Forecasting automated guided vehicle malfunctioning with deep learning in a 5G-based industry 4.0 scenario. *IEEE Communications Magazine* **59(11)**:102–108 DOI [10.1109/mcom.221.2001079](https://doi.org/10.1109/mcom.221.2001079).
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser u, Polosukhin I. 2017.** Attention is all you need. In: *Advances in Neural Information Processing Systems*. Vol. 30, New York: Curran Associates, Inc.
- Waikhom L, Patgiri R. 2023.** A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. *Artificial Intelligence Review* **56(7)**:6295–6364 DOI [10.1007/s10462-022-10321-2](https://doi.org/10.1007/s10462-022-10321-2).
- Wang Y, Bao D, Qin SJ. 2023.** A novel bidirectional DiPLS based LSTM algorithm and its application in industrial process time series prediction. *Chemometrics and Intelligent Laboratory Systems* **240(11)**:104878 DOI [10.1016/j.chemolab.2023.104878](https://doi.org/10.1016/j.chemolab.2023.104878).
- Wang Q, Bu S, He Z, Dong ZY. 2021a.** Toward the prediction level of situation awareness for electric power systems using CNN-LSTM network. *IEEE Transactions on Industrial Informatics* **17(10)**:6951–6961 DOI [10.1109/tii.2020.3047607](https://doi.org/10.1109/tii.2020.3047607).
- Wang J, Chen Y, Hao S, Peng X, Hu L. 2019.** Deep learning for sensor-based activity recognition: a survey. *Pattern Recognition Letters* **119(4)**:3–11 DOI [10.1016/j.patrec.2018.02.010](https://doi.org/10.1016/j.patrec.2018.02.010).
- Wang WK, Chen I, Hershkovich L, Yang J, Shetty A, Singh G, Jiang Y, Kotla A, Shang JZ, Yerrabelli R, Roghanizad AR, Shandhi MMH, Dunn J. 2022.** A systematic review of time series classification techniques used in biomedical applications. *Sensors* **22(20)**:8016 DOI [10.3390/s22208016](https://doi.org/10.3390/s22208016).
- Wang R, Li C, Fu W, Tang G. 2020a.** Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction. *IEEE Transactions on Neural Networks and Learning Systems* **31(10)**:3814–3827 DOI [10.1109/tnnls.2019.2946414](https://doi.org/10.1109/tnnls.2019.2946414).
- Wang X, Ma Y, Wang Y, Jin W, Wang X, Tang J, Jia C, Yu J. 2020b.** Traffic flow prediction via spatial temporal graph neural network. In: *Proceedings of the Web Conference 2020, WWW '20*. New York, NY, USA: Association for Computing Machinery, 1082–1092.
- Wang J, Ma Y, Zhang L, Gao RX, Wu D. 2018.** Deep learning for smart manufacturing: methods and applications. *Journal of Manufacturing Systems* **48(2)**:144–156 DOI [10.1016/j.jmsy.2018.01.003](https://doi.org/10.1016/j.jmsy.2018.01.003).
- Wang H, Mao L, Zhang H, Wu Q. 2024.** Multi-prediction of electric load and photovoltaic solar power in grid-connected photovoltaic system using state transition method. *Applied Energy* **353(2)**:122138 DOI [10.1016/j.apenergy.2023.122138](https://doi.org/10.1016/j.apenergy.2023.122138).
- Wang Q, Wang S, Wei B, Chen W, Zhang Y. 2021b.** Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features. *IEEE Access* **9**:1 DOI [10.1109/access.2021.3066489](https://doi.org/10.1109/access.2021.3066489).
- Wang Y, Yang K, Li H. 2020.** Industrial time-series modeling via adapted receptive field temporal convolution networks integrating regularly updated multi-region operations based on PCA. *Chemical Engineering Science* **228**:115956 DOI [10.1016/j.ces.2020.115956](https://doi.org/10.1016/j.ces.2020.115956).
- Weichert D, Link P, Stoll A, Rüping S, Ihlenfeldt S, Wrobel S. 2019.** A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology* **104(5)**:1889–1902 DOI [10.1007/s00170-019-03988-5](https://doi.org/10.1007/s00170-019-03988-5).
- Wu Y, Gu M, Wang L, Lin Y, Wang F, Yang H. 2021.** Event2Graph: event-driven bipartite graph for multivariate time-series anomaly detection. ArXiv DOI [10.48550/arXiv.2108.06783](https://doi.org/10.48550/arXiv.2108.06783).
- Wu Z, Pan S, Long G, Jiang J, Zhang C. 2019.** Graph WaveNet for deep spatial-temporal graph modeling. ArXiv DOI [10.48550/arXiv.1906.00121](https://doi.org/10.48550/arXiv.1906.00121).

- Xu D, Zhang Q, Ding Y, Zhang D. 2022.** Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. *Environmental Science and Pollution Research* **29**(3):4128–4144 DOI [10.1007/s11356-021-15325-z](https://doi.org/10.1007/s11356-021-15325-z).
- Yan W, Shao H, Wang X. 2004.** Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers & Chemical Engineering* **28**(8):1489–1498 DOI [10.1016/j.compchemeng.2003.11.004](https://doi.org/10.1016/j.compchemeng.2003.11.004).
- Yang Z, Keung J, Kabir MA, Yu X, Tang Y, Zhang M, Feng S. 2021.** AComNN: attention enhanced compound neural network for financial time-series forecasting with cross-regional features. *Applied Soft Computing* **111**(3):107649 DOI [10.1016/j.asoc.2021.107649](https://doi.org/10.1016/j.asoc.2021.107649).
- Yao P, Yang S, Li P. 2021.** Fault diagnosis based on ResNet-LSTM for industrial process. In: *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Vol. 5, 728–732 DOI [10.1016/j.compchemeng.2003.11.004](https://doi.org/10.1016/j.compchemeng.2003.11.004).
- Yassine A, Singh S, Hossain MS, Muhammad G. 2019.** IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems* **91**(4):563–573 DOI [10.1016/j.future.2018.08.040](https://doi.org/10.1016/j.future.2018.08.040).
- Yu C, Li Y, Chen Q, Lai X, Zhao L. 2022a.** Matrix-based wavelet transformation embedded in recurrent neural networks for wind speed prediction. *Applied Energy* **324**:119692 DOI [10.1016/j.apenergy.2022.119692](https://doi.org/10.1016/j.apenergy.2022.119692).
- Yu H, Li T, Yu W, Li J, Huang Y, Wang L, Liu A. 2022b.** Regularized graph structure learning with semantic knowledge for multi-variables time-series forecasting. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 2362–2368.
- Yu C, Ma X, Ren J, Zhao H, Yi S. 2020.** Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In: Vedaldi A, Bischof H, Brox T, Frahm J-M, eds. *Computer Vision—ECCV 2020, Lecture Notes in Computer Science*. Cham: Springer International Publishing, 507–523.
- Yuan X, Gu Y, Wang Y, Yang C, Gui W. 2020.** A deep supervised learning framework for data-driven soft sensor modeling of industrial processes. *IEEE Transactions on Neural Networks and Learning Systems* **31**(11):4737–4746 DOI [10.1109/tnnls.2019.2957366](https://doi.org/10.1109/tnnls.2019.2957366).
- Yuan X, Li L, Shardt YAW, Wang Y, Yang C. 2021.** Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics* **68**(5):4404–4414 DOI [10.1109/tie.2020.2984443](https://doi.org/10.1109/tie.2020.2984443).
- Zambon D, Grattarola D, Livi L, Alippi C. 2019.** Autoregressive models for sequences of graphs. In: *2019 International Joint Conference on Neural Networks (IJCNN)*.
- Zambon D, Livi L, Alippi C. 2022.** Graph iForest: isolation of anomalous and outlier graphs. In: *2022 International Joint Conference on Neural Networks (IJCNN)*. Piscataway: IEEE.
- Zang D, Liu J, Wang H. 2018.** Markov chain-based feature extraction for anomaly detection in time series and its industrial application. In: *2018 Chinese Control and Decision Conference (CCDC)*. Piscataway: IEEE, 1059–1063.
- Zareen S, Akram A, Ahmad Khan S. 2020.** Security requirements engineering framework with BPMN 2.0.2 extension model for development of information systems. *Applied Sciences* **10**(14):4981 DOI [10.3390/app10144981](https://doi.org/10.3390/app10144981).
- Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. 2021.** A transformer-based framework for multivariate time series representation learning. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21*. New York, NY, USA: Association for Computing Machinery, 2114–2124.

- Zhang Q, Chang J, Meng G, Xiang S, Pan C. 2020a.** Spatio-temporal graph structure learning for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* **34(1)**:1177–1185 DOI [10.1609/aaai.v34i01.5470](https://doi.org/10.1609/aaai.v34i01.5470).
- Zhang X, Gao Y, Lin J, Lu C-T. 2020b.** TapNet: multivariate time series classification with attentional prototypical network. *Proceedings of the AAAI Conference on Artificial Intelligence* **34(4)**:6845–6852 DOI [10.1609/aaai.v34i04.6165](https://doi.org/10.1609/aaai.v34i04.6165).
- Zhang X, Lei Y, Chen H, Zhang L, Zhou Y. 2021.** Multivariate time-series modeling for forecasting sintering temperature in rotary kilns using DCGNet. *IEEE Transactions on Industrial Informatics* **17(7)**:4635–4645 DOI [10.1109/tii.2020.3022019](https://doi.org/10.1109/tii.2020.3022019).
- Zhang S, Luo J, Wang S, Liu F. 2023.** Oil price forecasting: a hybrid GRU neural network based on decomposition—reconstruction methods. *Expert Systems with Applications* **218(10)**:119617 DOI [10.1016/j.eswa.2023.119617](https://doi.org/10.1016/j.eswa.2023.119617).
- Zhang R, Lv J, Li J, Bao J, Zheng P, Peng T. 2022a.** A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations. *Journal of Manufacturing Systems* **63**:491–503 DOI [10.1016/j.jmsy.2022.05.006](https://doi.org/10.1016/j.jmsy.2022.05.006).
- Zhang Z, Wu C, Qu S, Chen X. 2022b.** An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management* **59(4)**:102988 DOI [10.1016/j.ipm.2022.102988](https://doi.org/10.1016/j.ipm.2022.102988).
- Zhang M, Zhang Y, Zhang L, Liu C, Khurshid S. 2018.** DeepRoad: GAN-based metamorphic testing and input validation framework for autonomous driving systems. In: *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE '18*. New York, NY, USA: Association for Computing Machinery, 132–142.
- Zhao H, Wang Y, Duan J, Huang C, Cao D, Tong Y, Xu B, Bai J, Tong J, Zhang Q. 2020.** Multivariate time-series anomaly detection via graph attention network. In: *2020 IEEE International Conference on Data Mining (ICDM)*. Piscataway: IEEE, 841–850.
- Zheng G, Chai WK, Duanmu J-L, Katos V. 2023.** Hybrid deep learning models for traffic prediction in large-scale road networks. *Information Fusion* **92(2)**:93–114 DOI [10.1016/j.inffus.2022.11.019](https://doi.org/10.1016/j.inffus.2022.11.019).
- Zheng C, Fan X, Wang C, Qi J. 2020.** GMAN: a graph multi-attention network for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* **34(1)**:1234–1241 DOI [10.1609/aaai.v34i01.5477](https://doi.org/10.1609/aaai.v34i01.5477).
- Zhou L, Zeng Q, Li B. 2022.** Hybrid anomaly detection via multihead dynamic graph attention networks for multivariate time series. *IEEE Access* **10**:40967–40978 DOI [10.1109/access.2022.3167640](https://doi.org/10.1109/access.2022.3167640).
- Zhu F, Jia X, Miller M, Li X, Li F, Wang Y, Lee J. 2021.** Methodology for important sensor screening for fault detection and classification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing* **34(1)**:65–73 DOI [10.1109/tsm.2020.3037085](https://doi.org/10.1109/tsm.2020.3037085).
- Zintgraf LM, Cohen TS, Adel T, Welling M. 2017.** Visualizing deep neural network decisions: prediction difference analysis. ArXiv DOI [10.48550/arXiv.1702.04595](https://doi.org/10.48550/arXiv.1702.04595).