Energy 335 (2025) 138016

Contents lists available at ScienceDirect

Energy

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

From bearings to substations: Transfer Learning for fault detection in district heating

Jonne van Dreven [a,d,e,*], Abbas Cheddad [a,c], Sadi Alawadi [a], Ahmad Nauman Ghazi [b], Jad Al Koussa [d,e], Dirk Vanhoudt [d,e]

[a] Department of Computer Science, Blekinge Institute of Technology, Karlskrona, Sweden
[b] Department of Software Engineering, Blekinge Institute of Technology, Karlskrona, Sweden
[c] Institute of Computer Science, University of Tartu, Tartu, 51009, Estonia
[d] Water and Energy Transition Unit, Flemish Institute for Technological Research (VITO), Mol, Belgium
[e] EnergyVille, Genk, Belgium

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Transfer learning
Fault detection
District heating

ABSTRACT

Fault Detection and Diagnosis (FDD) in District Heating (DH) systems is vital for improving operational efficiency. As DH networks evolve towards Fourth-Generation District Heating (4GDH), their reliance on lower operational temperatures intensifies the need for robust FDD. However, implementing effective FDD faces challenges due to the lack of labelled fault data and the complexity of DH substations. This paper introduces a novel FDD methodology using Transfer Learning (TL) to bridge the gap between faulty bearings, controlled DH substation experiments and real-world operational data. We propose a fault signature method that aligns the data to reduce the domain gap, revealing similarities between faulty bearings' vibration patterns and ΔT readings in DH substations. The TL-enhanced models demonstrated robust performance, achieving F1 scores up to 98% on lab data and 91% on real-world operational data, respectively. These results mark a notable advancement in FDD of DH substations, as our method offers accurate fault detection and valuable insights across diverse operational contexts, ranging from valve issues, faulty sensors, wrong control strategies and normal behaviour. Notably, temperature dynamics resemble behaviour akin to faulty bearing vibrations, highlighting their potential as a critical indicator of a faulty substation, enabling more effective FDD in DH systems.

1. Introduction

As global cities confront the pressing need to reduce greenhouse gas emissions and transition towards sustainable energy systems, district heating (DH) networks have emerged as a key technology for delivering low-carbon thermal energy on a large scale. DH systems are transforming into smart, demand-adaptive distributed energy grids, and operating temperatures have fallen from roughly 180–200 °C (1GDH) to 130–150 °C (2GDH) and 80–100 °C (3GDH) to a maximum of 60–70 °C (4GDH) for newer networks to minimise losses and integrate low-grade renewables [1]. Therefore, DH systems are considered among the most efficient energy systems for delivering thermal heat in densely populated areas [2]. DH systems transport thermal energy from centralised production units through a network of insulated pipelines, often incorporating renewable or recovered sources to enhance energy efficiency and mitigate environmental impacts [3]. At each building, a compact substation interface efficiently transfers and regulates heat between the main network and the building heating circuit. Typically, a substation consists of a combination of plate heat exchangers, circulation pumps, valves, an array of temperature, flow, and pressure sensors, and vendor-specific control systems. However, communication hardware often restricts data to hourly intervals, and control logic can vary by network or from substation to substation. Moreover, when the equipment ages unevenly, it leads to undocumented, ad-hoc fixes over time. The coarse hourly sampling, undocumented hardware heterogeneity and site-specific control idiosyncrasies make automatic, data-driven substation monitoring extremely challenging.

As the complexity and scope of these networks grow, ensuring their operational reliability becomes a critical concern [4]. Faults [5], such as poorly tuned valves, faulty sensors, suboptimal control strategies, or suboptimally designed or operated systems, can lead to elevated return temperatures, wasting valuable energy, increased financial costs, diminished reliability, and reduced integration potential with advanced

* Corresponding author at: Department of Computer Science, Blekinge Institute of Technology, Karlskrona, Sweden.
E-mail address: jonne.van.dreven@bth.se (J. van Dreven).

https://doi.org/10.1016/j.energy.2025.138016
Received 17 January 2025; Received in revised form 28 June 2025; Accepted 11 August 2025
Available online 18 August 2025

Abbreviations

4GDH	Fourth-Generation District Heating
AE	Autoencoder
ARIMA	Autoregressive Integrated Moving Average
CNN	Convolutional Neural Network
DBA-DTW	Dynamic Time Warping Barycentre Averaging
DH	District Heating
DL	Deep Learning
DTS	Deviating Temperature Sensor
DTW	Dynamic Time Warping
FDD	Fault Detection and Diagnosis
HHC	High Heat Curve
IF	Isolation Forests
LSL	Large Secondary Leakage
LSTM	Long Short-Term Memory
ML	Machine Learning
MVL	Minor Valve Leak
OCSVM	One Class Support Vector Machines
OCSVM++	Our enhanced OCSVM model with domain adaptation
OCV	Oversized Control Valve
PdM	Predictive Maintenance
SV	Stuck Valve
TL	Transfer Learning
VAE	Variational Autoencoder
VL	Valve Leak
WSP	Wrong Sensor Placement

low-temperature networks [6,7]. For instance, a large sub-optimal substation can incur up to 100,000 euros per year in additional pumping power [8]. Previous works [6,9] indicate that between 43%–75% of the tested substations and secondary systems perform sub-optimally due to faults, while in Copenhagen it is estimated to be over 50% [10]. A survey [5] has shown that the topic of Fault Detection and Diagnosis (FDD) is gaining popularity. Out of 56 participating utilities, 50 stated that they performed some form of analysis on customer data to detect deviating installation behaviours. Most utilities currently identify faults during service visits to their customers' installations. While some utilities employ methods primarily based on the analysis of individual customer installations, e.g., using threshold methods [11], most utilities perform this analysis manually. Therefore, a significant portion of faults remains undetected and persists for an extended period. This challenge becomes more apparent as the DH sector transitions to 4GDH systems. These systems operate at lower temperatures, facilitating better integration with renewable energy sources and enhanced flexibility to meet future energy grid demands [12,13]. While these DH generations promise greener and more adaptive infrastructures, they also narrow the margin for error, intensifying the need for effective FDD techniques.

Nevertheless, recent developments in smart heat metering offer opportunities for data-driven FDD [14] and predictive maintenance (PdM) [15]. Automatic FDD can reduce customer energy use by as much as 14% [16]. Furthermore, FDD and PdM strategies are more cost-effective and less disruptive than traditional (annual) maintenance audits [6,17]. Smart heat meters in DH substations typically log supply- and return temperatures, accumulated volume, flow rates, and accumulated energy at hourly intervals via commercial metering solutions (e.g., Kamstrup Multical, Landis+Gyr E450) and SCADA controllers (e.g., Siemens QVD). While this leads to more open datasets [18], the granularity, which is sufficient for billing and general monitoring, complicates accurate fault detection, which requires higher resolution data [19]. Most FDD methods in DH rely on hourly trained models, such as using Autoregressive Integrated Moving Average (ARIMA) for forecasting and anomaly detection [14]. However, such models may suffer from baseline contamination, whereby undiagnosed faults in the training data become integrated into the learned pattern, reducing sensitivity to actual anomalies. Additionally, traditional Gaussian-based fault detection models often fail in real-world scenarios, e.g., heavy tails and non-linearities, motivating the development of more advanced and robust methods [20]. Moreover, lacking labelled and high-frequency data hinders the advancement of robust FDD and PdM solutions [21]. Fortunately, studies aim to address the issue of data collection practices in DH systems [22,23]. Nevertheless, without consistent, high-quality labelled datasets, it is difficult to train and validate algorithms capable of generalising the diversity of equipment, network configurations, or end-user consumption patterns for accurate FDD. Moreover, the complexity of the impact of faults, which may vary widely across different substations and contexts, further undermines the direct transfer of knowledge from one scenario to another.

In response, we propose a novel approach that uses Transfer Learning (TL) to bridge the gap between controlled lab experiments and the complexities of real-world DH substations. TL has proven effective in numerous fields for mitigating data scarcity and domain adaptation challenges [24]. We use lab-generated data from an experimental setup built at the VITO/EnergyVille[1] Thermo-Technical Laboratory in Genk, Belgium [25]. to identify and learn ground-truth fault patterns of a DH substation. Interestingly, we observe trends in vibration patterns from faulty bearings similar to the return or ΔT readings of substations. To capitalise on this similarity, we enhance our models with TL to perform fault detection and link these patterns, from faulty bearings, fine-tuned by controlled lab experiments, to real-world operational substations. This study also proposes a novel fault signature generation method designed to encapsulate daily substation operational patterns over specified time frames, thereby mitigating the domain discrepancy between fault characteristics in bearings and DH substations.

The main contributions of this work are:

- We propose a TL framework for fault detection in DH substations, leveraging labelled vibration data from faulty bearings and fine-tuning on lab-emulated DH faults.
- We propose a novel daily-signature generation pipeline to minimise the domain gap between mechanical vibrations and highlighting temperature dynamics.
- We demonstrate that the differenced temperature differentials (ΔT) and return temperatures (T_r) of faulty substations exhibit vibration-like, high-variance fault signatures akin to faulty bearing vibrations, justifying cross-domain feature learning.
- Our TL-enhanced models achieve F1 scores up to 98% on unseen lab emulation data and 91% on independent real-world substation data, showcasing excellent performance and generalisation.

The remainder of this paper is organised as follows: Section 2 reviews relevant advances in DH FDD and TL, Section 3 describes the data acquisition process and provides a brief overview of the laboratory emulations, Section 4 details our methodology for implementing TL-driven FDD solutions, Section 5 presents our experimental setup including training protocol and performance evaluation metrics, Section 6 presents our results and discusses their implications, and we conclude with insights from our experiments in Section 7.

[1] A research facility focused on sustainable energy system development (https://energyville.be).

2. Related work

In recent years, Deep learning (DL) has significantly advanced FDD [26–30]. These models excel in capturing the complex, non-linear patterns inherent in real-world engineering systems, effectively modelling sophisticated relationships that traditional methods often fail to capture. Of particular relevance to fault detection is the Autoencoder (AE), a type of neural network that is particularly adept at semi-supervised learning, allowing it to detect anomalies by learning compressed representations of input data [31].

Traditional shallow ML models are presently the primary contributors to FDD in DH systems [14]. Consequently, studies are addressing the data scarcity [32]. For instance, a study in [25] details a systematic method for generating data from a lab-based DH substation emulations and evaluates the performance of common ML models, such as an One-Class Support Vector Machine (OCSVM) and Isolation Forest (IF). These models achieved F1 scores of 0.74 and 0.85 on a controlled laboratory-generated dataset, respectively. Although these models provide a baseline for fault detection, they often face challenges in handling complex non-linear patterns and adapting to the diverse operational contexts typically encountered in real-world DH scenarios.

To address these limitations and the challenge of limited labelled data, Transfer Learning (TL) presents a compelling solution by leveraging knowledge transfer from related domains to improve model performance [24,33]. Comprehensive reviews by Chen et al. [34] and Li et al. [35] emphasise the potential of TL to enhance FDD by improving model adaptability across different domains. These works explore key TL frameworks, including knowledge calibration, joint distribution adaptation, and deep adaptation networks, which are particularly relevant for addressing the challenges of generalisation and domain variability.

The application of TL and AE techniques for FDD tasks, however, remains an underexplored area within the domain of DH [14]. Chavan et al. [36] proposed a TL methodology for FDD applied to building services and DH systems, while Wang et al. [37] introduced layer transfer and merged transfer models for thermal load prediction in DH substations using TL techniques. Vallee et al. [38] developed a synthetic dataset to enhance FDD in DH systems, evaluating multiple models and highlighting the potential for model transfer via TL. The AE method has recently been employed, as seen in [39], where the authors proposed a hybrid anomaly detection method that combines a physical model with a Long Short-Term Memory (LSTM)-based Variational Autoencoder (VAE), comparing it to LSTM and LSTM-AE baseline models. Similarly, Choi and Yoon [40] applied AE-driven FDD methods to building automation systems, including DH systems. Hong et al. [41] proposed a system-level virtual sensing method using AE to improve sensing accuracy and address sensor absences in DH systems. Shahid et al. [42,43] employed LSTM-VAE and Convolutional Neural Network (CNN)-LSTM AE for anomaly detection in DH consumption in school buildings.

AE-driven FDD research is more prevalent in other domains such as electrical systems [44–47], telecommunications [48,49], and the automotive domain [50–56]. The latter domain, particularly faulty bearings, represents an established FDD domain with abundant labelled data, such as [57], which provides an opportunity to transfer knowledge to FDD in DH. While deep models (AEs, VAEs, LSTM-AEs) and TL approaches have each shown promise for FDD in various domains, their joint application to real-world DH systems remains underexplored. Moreover, prior DH studies had to rely on simulated data due to labelled data scarcity, which can fail to capture field-level noise and dynamics. Therefore, it remains an open question as to how these proposed solutions would hold up in real-world operational scenarios. Additionally, while faulty-bearing datasets offer rich anomaly signatures, as far as we are aware, no work has yet transferred that knowledge to the DH FDD.

It is important to recognise that moving from synthetic data generated by simulations to real-world applications involves significant (domain-mismatch) challenges observed in [38,58]. While beneficial, as large amounts of data can be generated in a short period, simulated signals often omit the complex measurement noise, non-linear interactions and unmodelled dynamics present in real DH networks. Moreover, there may be discrepancies between optimal simulated sensors and field-deployed equipment, which can degrade over time, introducing additional bias. Simulated data can also have incomplete fault coverage, as it only captures a limited set. At the same time, operational networks experience a diverse, overlapping, or simultaneously occurring spectrum of faults, which complicates model generalisation. In general, depending on the simulation, data can become too synthetic, resulting in a domain shift. To address this, our work seeks to exploit faulty bearing datasets to mitigate the scarcity of labelled data and to leverage controlled lab emulations, a physical simulation and step closer to reality, of DH substations to ease the generalisation process to real-world DH systems. Specifically, we propose a novel TL framework that leverages fault signatures from DH substations, integrating knowledge from faulty bearing datasets and DH lab emulations to improve fault detection in real-world DH systems, even when labelled data is scarce.

3. Data acquisition process

To validate our proposed approach, we have used three different datasets, namely:

1. **Faulty Bearing dataset**: This dataset is provided by the Bearing Data Center at the Case Western Reserve University [57], which serves as a key source for our proposed TL approach. The dataset offers vibration data captured under various fault conditions, including normal operations and single-point defects in drive and fan-end bearings. Data collection was performed at a standard sampling rate of 12,000 samples per second across all experiments, with an additional dataset recorded at 48,000 samples per second for drive-end bearing tests, providing a comprehensive basis for fault diagnosis analysis.
2. **Lab DH substation dataset**: We utilise experimental data from DH substation lab emulations to train and validate our ML models. We provide a brief overview of the lab setup and data below, for full details, we refer the reader to [25]. The experimental setup, illustrated in Fig. 1, mimics a residential DH substation, enabling the generation of realistic consumption data. The laboratory skid represents the DH network, supplying the necessary temperature approximately 80 °C and flow via a circulation pump. Additionally, a bypass system ensures continuous water circulation on the primary side during periods of no heat demand.
   The substation includes a heat exchanger, controllable valve, and external controller for inducing faults. On the secondary side, a pump circulates water to radiators in a climate chamber, where a power profile controls heat demand. When the chamber temperature drops below (21 °C), the pump activates, and the primary valve opens to maintain the required secondary supply temperature based on a heating curve aligned with outdoor temperatures. We monitor temperatures, flow rates, and pressure drops using calibrated PT100 sensors and electromagnetic flow meters. To simplify control, radiators lack thermostatic valves. Table 1 summarises our collected features.
   We conducted six test scenarios: normal operation, Minor Valve Leak (MVL), Valve Leak (VL), Stuck Valve (SV), High Heat Curve (HHC), and Deviating Temperature Sensor (DTS). These six fault scenarios were chosen according to three criteria: (i) their frequency in field-survey studies; (ii) reproducibility in our laboratory emulation; and (iii) their varying severity in

Table 1
Time series data characteristics from the lab-simulated and real-world DH substations.

Feature	Abbr.	Type	Description
Timestamp	TS	Datetime	Recorded time of measurement
Primary supply temp.	T_s	Continuous	Temperature of primary supply line (°C)
Primary return temp.	T_r	Continuous	Temperature of primary return line (°C)
Primary flow	Q	Continuous	Flow rate of primary circuit (L/min)
Outdoor temp.	T_{out}	Continuous	Ambient temperature outside the system (°C)
Secondary supply temp.	T_{ss}	Continuous	Temperature of secondary supply line (°C)
Secondary return temp.	T_{sr}	Continuous	Temperature of secondary return line (°C)
Secondary flow	Q_s	Continuous	Flow rate of secondary circuit (L/min)
Temp. difference	ΔT	Continuous	Difference between T_s and T_r (°C)
Energy consumption	E	Continuous	Total energy used by a substation (kWh)
Volume	V	Continuous	Total volume of water circulated (m^3)

thermal-signal impact from easily detectable (SV) to more subtle deviations (MVL, HHC, and DTS). In particular, Månsson et al. [5] report that these faults collectively account for many real-world substation failures. These faults were induced under identical outdoor temperature and heating demand profiles, based on a typical Belgian winter. Outdoor temperatures (see Fig. 2 for the temperature and heat demand profile) during a typical Belgian winter can range from cold days (approximately −5 °C) up to milder days (up to approximately 10 °C), which provides a broad spectrum of thermal boundary conditions for analysis. The fault dynamics were observed over a two-week period, with data sampled every 10 s (resulting in 120.960 samples for each scenario per sensor measurement), allowing for resampling to coarser intervals. Each scenario began with an initialisation phase to reach a steady state, after which data collection started. Upon completion of a fault test, the substation was returned to its normal baseline operation and verified before the same protocol was applied for the next fault scenario. All experiments ran under identical boundary conditions throughout the two-week test.

3. **Real-world DH substation dataset:** This dataset was collected at 5 min intervals during January 2024, resulting in 8928 samples from 248 substations in a DH network (operating temperatures of approximately 85 °C) in Shandong Province, China. We selected this dataset through our collaboration with the local DH utility, as DH datasets with fault annotations are exceptionally scarce and difficult to obtain, yet vital for progressing data-driven FDD in DH. The dataset includes both primary and secondary side data, along with geographic coordinates of the substations. It is partially labelled, identifying fault types such as HHC, Wrong Sensor Placement (WSP), Oversized Control Valve (OCV), and Large Secondary Leakage (LSL), as well as normal operational behaviour. Unique labels are assigned to distinguish scenarios across different substations. For instance, we labelled substations under normal operation as 'normal', where 'XX' serves as a unique identifier for each substation within the class. Similarly, fault scenarios such as 'WSP01' and 'WSP02' correspond to distinct substations with the WSP fault. Notably, not all faults are observed in every substation, resulting in an unequal distribution of class types. For instance, only three substations exhibited the LSL fault, named 'LSL01', 'LSL02' and 'LSL03', respectively, leaving 'LSL04' absent. This labelling structure ensures clarity and granularity for analysing and modelling substation behaviour.

Return temperatures T_r are an important indicator of operational efficiency and potential faults in DH substations [59]. T_r (patterns) can signal heat transfer inefficiencies or faulty components, making them valuable for fault detection. Therefore, in this work, we analyse both the return temperature signature (T_r) and the temperature differential (ΔT) signature at a 15 min sampling frequency to improve fault detection. (ΔT) is defined as:

$$\Delta T = T_s - T_r, \tag{1}$$

where T_s is the primary supply temperature and T_r is the primary return temperature. Higher ΔT values indicate a more efficient system with better heat transfer performance. Furthermore, because our TL pipeline is pre-trained on bearing vibration data, which only provides temperature-analogous signals, we restrict our analysis to T_r and ΔT to maintain a consistent feature space across source and target domains.

3.1. Pre-training rationale

Fig. 3 shows that bearing and DH fault signatures visually share high-variance spikes in their stationary signals. To quantify this resemblance, we extracted the per-day variance in each domain and computed summary metrics, such as mean and standard deviation, to capture central tendency and spread, as well as kurtosis and skewness, to characterise tail behaviour and asymmetry. We then applied three non-parametric hypothesis tests: (i) the Kolmogorov–Smirnov test for distributional shape, (ii) the Fligner–Killeen test for variance homogeneity, and (iii) the Mann–Whitney U test for differences in central tendency. The detailed results are reported in Section 6.1 (Table 4).

4. Methods and model construction

This section outlines our methodology for fault detection in DH substations, focusing on capturing operational signatures to differentiate between normal and faulty behaviour while minimising the domain gap with faulty bearings (see Section 4.1). We incorporate TL to leverage knowledge from the related faulty bearing domain and fine-tune using DH lab-generated data. This combined approach aims to produce a robust model that enhances detection accuracy while improving its generalisation and reliability for real-world DH systems (see Section 4.2) in data-scarce scenarios.

4.1. Data preprocessing—fault signature generation-

The fault signatures serve two purposes. First, they aim to capture the daily operational dynamics of substations over a defined period. Importantly, our approach focuses on the relative behavioural changes and temporal patterns in those dynamics, i.e., how signals evolve, rather than on their absolute magnitudes, which makes the detection robust to scale differences. Second, they help reduce the domain gap between fault characteristics in faulty bearings and those in faulty substations. Fig. 4 illustrates the process of creating a fault signature, which comprises the following steps:

(a) *Segmentation*: The time series data from a substation is divided into individual days over a specific period. Dataset 1 consisted of 14 consecutive days, and Dataset 2 of 31 consecutive days, respectively. This step enables the isolation of daily operational patterns for analysis.

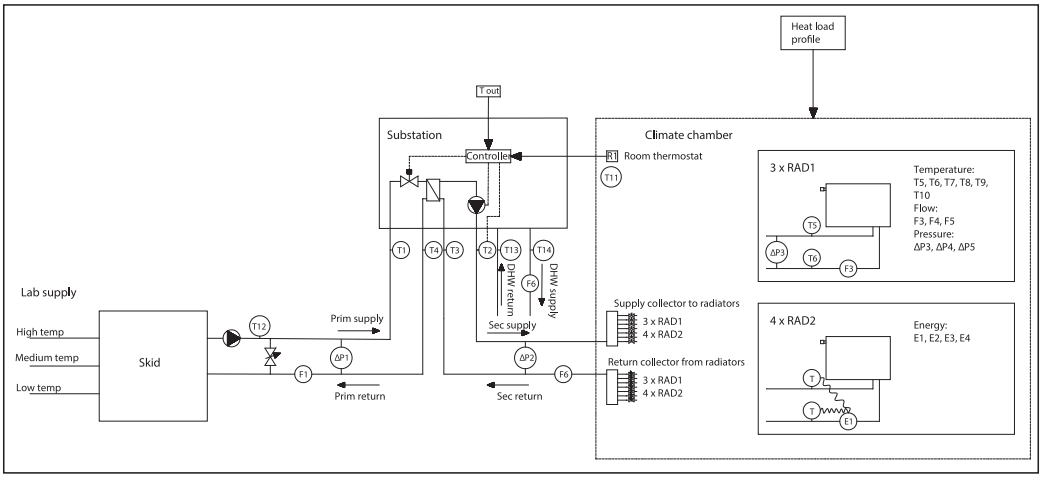


Fig. 1. A schematic of the lab setup used for the DH substation emulation [25]. This setup was built at the VITO/EnergyVille Thermo-technical Laboratory in Genk, Belgium. The figure is licensed under CC BY 4.0 License http://creativecommons.org/licenses/by/4.0/.

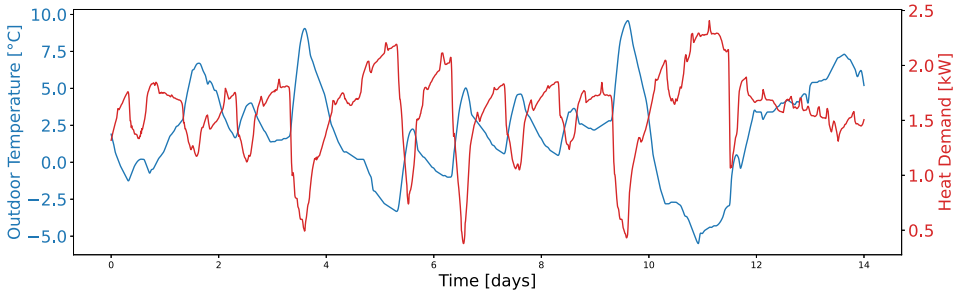


Fig. 2. The temperature and heating demand profile used in the lab setup [25]. The figure is licensed under CC BY 4.0 License http://creativecommons.org/licenses/by/4.0/.

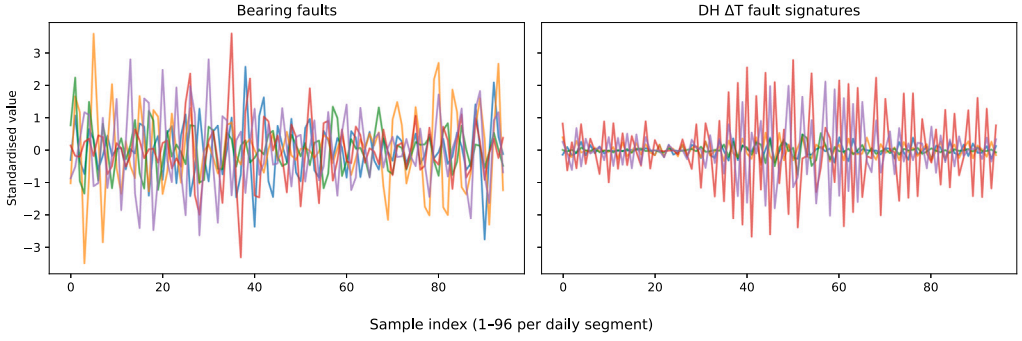


Fig. 3. Comparison of processed fault-signature time series in the bearing and DH domains. (a) The first five faulty bearing vibration segments. (b) The first five return temperature differential (ΔT) segments from the lab DH substation emulation. Both series exhibit intermittent, high-variance excursions characteristic of fault events, illustrating the cross-domain resemblance that motivates our TL approach.

(b) *Standardisation*: The segmented data is standardised using z-score normalisation. This operation transforms the series to zero mean and unit variance to remove absolute scale differences, enabling direct overlay of operational patterns across days, whereas min–max (0–1) scaling would preserve relative amplitudes and thus retain differences in temperature magnitude. Since our goal is to characterise dynamic behaviour (step-to-step variations) rather than absolute temperature levels, z-score is preferred.

(c) *Alignment*: Dynamic Time Warping Barycentre Averaging (DBA-DTW) [60], with a warp constraint (Sakoe-Chiba [61]), aligns the daily patterns. This step prevents over-extension of the time series and maintains meaningful temporal relationships.

(d) *Average Pattern*: DBA-DTW generates a robust average pattern that represents consistent daily behaviour for the specified period. This step effectively filters out noise, resulting in a more reliable fault signature.

(e) *Differencing*: We apply first-order differencing, $\Delta p_t = p_t - p_{t-1}$, to the daily signature p_t. This operation transforms the series into its dynamic component, i.e., the incremental changes, thereby emphasising behavioural patterns driven by faults while suppressing baseline offsets and slow trends [62].

The fault signature generation process builds upon our previous work [25] but with a distinct objective. While the earlier method aimed to capture and demonstrate the impact of faults, our new approach targets reducing the domain gap between faulty bearings and DH

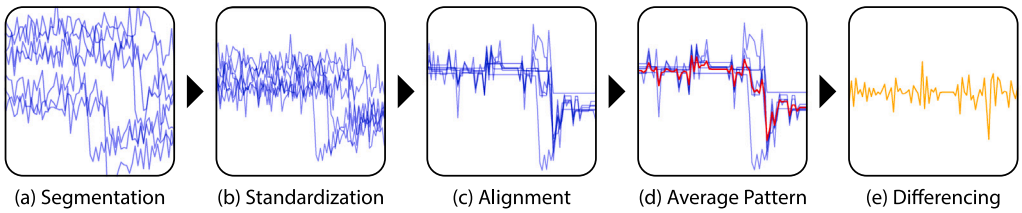


Fig. 4. Fusion process of the daily DH substation signatures.

substations. Moreover, we introduce several key enhancements. Notably, we have replaced the traditional Dynamic Time Warping (DTW) alignment method with the DBA-DTW approach. This change enables the simultaneous alignment of multiple time-series data, resulting in a more robust average pattern. By doing so, we avoid the biases that arise from selecting a single representative sample, such as the reference day used in our previous work [25]. Furthermore, we incorporate an additional *differencing* step to highlight and capture dynamic changes more effectively. This new approach transforms DH data to align more closely with the source domain of faulty bearings, thereby reducing the domain gap and enhancing the model's capacity to generalise from bearing faults to faults in DH systems.

4.2. Model training for fault detection

We employ an OCSVM [63] model enhanced with a domain adaptation technique (which we designate as OCSVM++ in this study), an AE [64,65], and a VAE [66] to detect faults in DH systems. In typical fault detection scenarios, models are commonly trained on normal operational data due to its abundance, as faulty data is often scarce and challenging to obtain. While we still have data scarcity challenges for both class types, we have more faulty labelled data.

Given the scarcity of labelled data in DH systems, we employ a TL approach. The source domain for TL consisted of faulty bearing data, which exhibit characteristics similar to faults in DH substations, such as irregular vibration patterns and large disturbances. The models were initially trained on the source domain to learn generalisable fault features. Subsequently, we fine-tune the models using lab-generated data from controlled substation fault scenarios, optimising the model parameters to suit the specific operational context of DH substations.

The OCSVM++ model is trained to delineate a decision boundary encompassing faulty behaviour patterns. This approach enables the identification of anomalies that deviate from the encapsulated class, effectively detecting instances of normal operation through a process of exclusion. In contrast, the AE, an artificial neural network, learns the underlying data distribution of a class by minimising the reconstruction error [67]. Any significant deviations from this learned representation signal an anomaly, effectively distinguishing between faults and normal operations. While OCSVMs are not traditionally designed for TL in the same way DL models (like AEs or VAEs) are. DL models often excel at TL because they can learn hierarchical, generalisable features that can be fine-tuned across domains [68]. OCSVMs, in contrast, are inherently kernel-based models that rely on explicit feature mappings, making their performance heavily dependent on the quality and compatibility of input features across domains [69]. Therefore, the feature distributions between the source (faulty bearings) and the target domain (DH data) should be similar or made similar by careful feature engineering as explained in Section 4.1. While this aligns with TL principles, where knowledge from a source domain aids in performing tasks in a target domain [70], we should be extra aware of challenges such as negative transfer if the source and target domains are not well-aligned [71].

Our AE architecture, illustrated in Fig. 5, is designed to encode and decode the variability in the provided signature of 15 min time series data for fault detection in DH systems. The model comprises an encoder, bottleneck, and decoder, using fully connected dense layers to capture patterns in the data. The encoder reduces the input dimensionality through three dense layers. At the model's core is a bottleneck layer (latent space), capturing the most salient features of the input data. The decoder mirrors the encoder with dense layers, progressively reconstructing the input data. Dense layers were chosen over convolutional layers, as we aim to capture a signature's statistical structure and variability rather than localised spatial patterns [72].

The VAE follows a similar architecture to the AE but introduces a probabilistic latent space. Instead of learning fixed latent variables, the VAE learns a probability distribution with a mean (μ) and a logarithmic variance ($\log \sigma^2$). This enables the model to sample from a Gaussian distribution during training and inference. The loss function for the VAE combines two components: the reconstruction loss and the Kullback–Leibler divergence [73]. This hybrid loss function ensures that the VAE not only reconstructs the input data accurately but also learns a structured latent space that facilitates anomaly detection and allows the model to generalise better.

5. Experimental settings

We partition the bearing dataset into training and evaluation subsets using an 80%/20% ratio. For the laboratory dataset, we use the first week for fine-tuning and validation, while the second week forms an unseen test set (dataset 2). The real-world dataset served as a completely independent unseen test set (dataset 3). This split ensures sufficient data for training while maintaining a robust evaluation set to assess the generalisation capability of the models. We evaluate the proposed approach over these two unseen test sets. Dataset 2 consists of unseen lab data, which serves as known ground-truth information. Dataset 3 contains real-world DH data to assess the model's practical applicability in operational environments. To ensure a robust evaluation, we employed Monte Carlo cross-validation [74], taking into account the time-dependent nature of the time series data. We segment the data into daily intervals, with each day treated as an independent unit. This approach assumes that fault signatures, reflecting the first data cycle, are minimally influenced by previous days. While some dependencies between consecutive days may exist, they are considered negligible in detecting the fault patterns driven by the faults themselves. To quantify whether differences in performance or distributions were systematic rather than random, we applied a suite of non-parametric hypothesis tests and effect-size measures. Specifically, we used the Kolmogorov–Smirnov test to compare the overall score distributions, the Fligner–Killeen test to assess the homogeneity of variances, and the Mann–Whitney U test to evaluate two independent distributions. Wherever we found statistically significant differences (at the $\alpha = 0.05$ level), we additionally reported Cohen's d to quantify the practical magnitude of those effects. Moreover, we have used different evaluation metrics to examine the proposed approach performance, including:

- **Accuracy**: Proportion of correctly identified instances (normal and faulty) among the total number of instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2}$$

where TP are True Positives, TN are True Negatives, FP are False Positives, and FN are False Negatives, respectively.

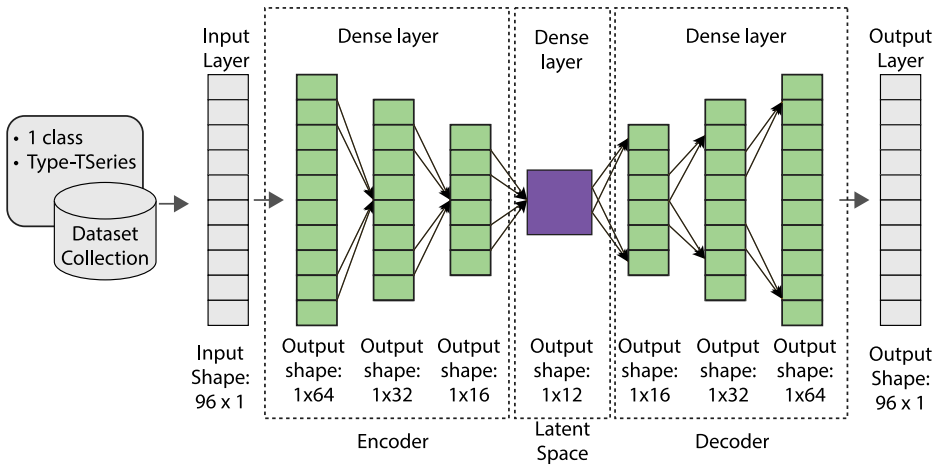


Fig. 5. The AE model architecture used in this work.

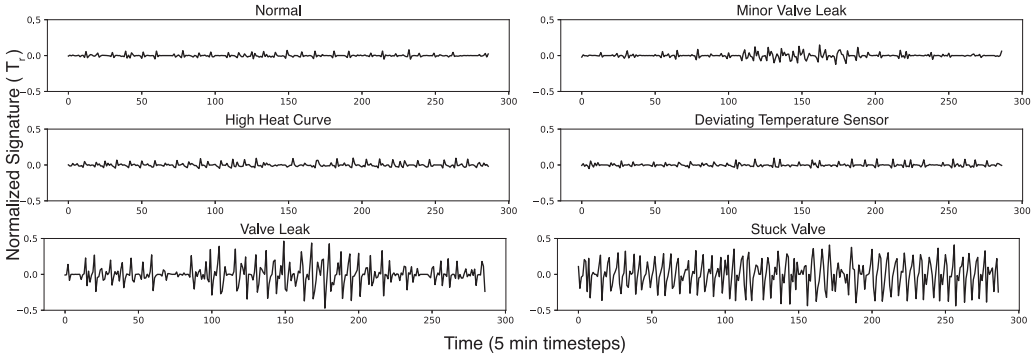


Fig. 6. Examples of daily return temperature (T_r) signatures for each lab-emulated fault scenario with five-minute timesteps. Each panel shows the daily T_r dynamics of a scenario: (1) Normal operation, (2) Minor Valve Leak, (3) High Heat Curve, (4) Deviating Temperature Sensor, (5) Valve Leak, and (6) Stuck Valve.

- **Precision**: Proportion of correctly identified positive instances among all instances identified as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

- **Recall (Sensitivity)**: Proportion of correctly identified positive instances among all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

- **F1 Score**: Harmonic mean of precision and recall, balancing false positives and false negatives:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

6. Results and discussion

To investigate the impact of various faults on DH substations, we conducted controlled experiments in a laboratory setting [25], emulating both normal operation and specific faulty conditions. Fig. 6 presents the daily signatures for each lab scenario, depicting the daily behaviour of a substation over a given period. Using first-order differencing in its last step, these signatures highlight temporal changes between consecutive time steps, effectively removing trends and rendering the time series stationary. This allows us to focus the analysis on the system's short-term adjustments and the magnitude of its responses.

In the signatures, higher peaks correspond to larger deviations between time steps, indicating abrupt adjustments or fluctuations in the substation's behaviour. These peaks reflect the system's response to dynamic operational conditions or faults, with more significant peaks suggesting severe or sudden changes in thermal dynamics. Conversely, lower peaks indicate smoother and more stable adjustments, which are characteristic of normal operation. The signatures facilitate the comparison of system stability across different scenarios, particularly under faulty conditions where larger peaks signify increased instability.

This distinction is apparent in Table 2, where we rank the scenarios based on the variability of their ΔT and return temperature signatures over two weeks. The rankings demonstrate that normal operation consistently ranks the highest in stability, exhibiting the lowest variance in both signatures. In contrast, faulty scenarios such as SV and VL display significantly higher variances, ranking lowest in stability.

To statistically validate the observed differences in variability between normal and faulty conditions, we performed the Fligner–Killeen test [75], a robust, non-parametric method for assessing homogeneity of variances across multiple groups. This test determines whether the observed variability patterns, where faulty conditions exhibit higher variances than normal operations, are statistically significant rather than random fluctuations, thereby supporting our hypothesis that faults introduce more significant variability. Table 3 summarises the p-values obtained when comparing normal operation to each faulty scenario for the ΔT and T_r signatures over two weeks. The results reveal a statistical significance in variances between normal operation and all faulty scenarios, with p-values far below the significance (α) threshold of 0.05 (largest p-value observed for $T_r = 0.002$ and $\Delta T = 0.006$ with the exception for week 1 MVL and DTS where p-value is 0.082 and 0.128, respectively). Notably, normal operation consistently exhibits the lowest variance, reflecting stable system behaviour. In contrast, major faults such as stuck or leaking control valves display significantly higher variances, indicating substantial variability in the daily signatures.

Table 2
Average ranking of scenarios for Week 1 and Week 2 of the signature using ΔT and T_r Stability. Fault scenarios include MVL, VL, SV, HHC, and DTS. We highlight the lowest variance in bold.

Scenario	ΔT signature			T_r signature		
	Average rank	W1	W2	Average rank	W1	W2
Normal	**1.5**	**0.00424**	**0.00223**	**1.0**	**0.00018**	**0.00009**
DTS	3.0	0.00557	0.00380	2.0	0.00083	0.00046
HHC	4.0	0.00746	0.00617	3.0	0.00140	0.00082
MVL	1.5	0.00443	0.00178	4.0	0.00334	0.00148
OCV	5.0	0.10767	0.05239	5.0	0.04187	0.02305
SV	6.0	0.25919	0.17003	6.0	0.06524	0.04288

Table 3
Fligner–Killeen test of variance p-values for normal operation compared to faulty scenarios (MVL, VL, SV, HHC, DTS). Values are shown in three decimals; those below 0.001 are reported as <0.001. Statistical significance at $\alpha = 0.05$ is in bold.

Scenario		Week 1		Week 2	
		ΔT	T_r	ΔT	T_r
Normal	MVL	0.082	**<0.001**	**0.006**	**0.002**
	VL	**<0.001**	**<0.001**	**<0.001**	**<0.001**
	SV	**<0.001**	**<0.001**	**<0.001**	**<0.001**
	HHC	**0.001**	**<0.001**	**<0.001**	**<0.001**
	DTS	0.128	**<0.001**	**0.003**	**<0.001**

This statistical evidence supports the conclusion that faulty conditions inherently introduce greater instability to the signature, as captured by the increased variability in the stationary ΔT and T_r signatures.

We can explain the distinct behaviour patterns observed in the daily signatures by considering how each specific fault impacts the thermal dynamics of the DH system:

- **Normal Operation**: The system maintains a stable thermal exchange with gradual and consistent adjustments. The signatures show small, regular peaks, indicating smooth management of temperature changes. The low variance reflects steady-state operation with minimal and predictable shifts.
- **MVL**: A minor leak in the control valve slightly disrupts the flow, causing minor deviations in thermal balance. The system makes subtle adjustments to compensate, resulting in small but noticeable peaks in the signature. Variance is higher than normal but remains moderate.
- **VL**: A significant control valve leak exacerbates the disruption between the primary and secondary sides. When heat demand is low or absent, the leaking valve allows unintended circulation on the primary side without corresponding demand on the secondary side. This mismatch leads to unstable heat transfer and irregular temperature fluctuations, reflected in higher variance.
- **SV**: This represents a severe fault where the control valve is stuck, causing constant flow on the primary side regardless of demand on the secondary side. This imbalance results in severe and unpredictable temperature fluctuations, with larger and more erratic peaks in the signature and significantly higher variance due to the system's inability to regulate flow properly.
- **HHC**: Adjusting the heat curve to a higher setting forces the system to overcompensate, pushing more heat into the system than required. This leads to rapid and aggressive adjustments. The signatures exhibit higher peaks, indicating the system's overreaction to the imposed demand and increased variability.
- **DTS**: A faulty sensor provides inaccurate temperature readings, causing the system to misinterpret the secondary side temperatures. This results in incorrect adjustments to the heat supply, with the system either overcompensating or undercompensating. The signatures show higher peaks and increased variability due to the instability introduced by sensor errors.

Table 4
Per-day variances summary for bearing and DH fault signatures.

Domain	Mean	Std dev	Kurtosis	Skewness
Bearing faults	0.711	0.703	4.85	1.95
DH substation faults	0.632	0.624	0.59	1.19

In summary, faulty conditions manifest as larger and more erratic peaks in the signatures than normal operation, indicating abrupt or irregular adjustments in system behaviour. Faults directly impacting the heat exchange process, such as stuck valves and control valve leaks, cause more severe instability. Deviations in sensor accuracy introduce artificial variability by causing incorrect control responses. These findings confirm that faults inherently lead to greater shifts in system dynamics, as captured by the ΔT and T_r signatures.

6.1. Pre-training validation

To quantify the resemblance between faulty bearing and DH substation data, we computed the per-day variances of the signals in each domain. Summary statistics (Table 4) show comparable heavy-tailed, skewed behaviour. Non-parametric tests, the Kolmogorov–Smirnov test ($D = 0.367, p = 0.412$) for distributional shape, the Fligner–Killeen test ($V = 2.682, p = 0.101$) for variance homogeneity, and the Mann–Whitney U test ($U = 3947, p = 0.345$) for median differences. Cohen's $d = 0.52$ indicates a moderate mean shift.

The early layers of our DL model learn generic features [76]. By pre-training on the larger, label-rich bearing dataset, we acquire reusable low- and mid-level representations of abrupt temporal changes. Fine-tuning then adjusts higher-level weights to the DH domain, bridging any residual scale or distributional gap, to better fit the nuances of the target data, thereby improving model performance without requiring extensive retraining from scratch. Theory shows that TL delivers its greatest gains when source and target domains are related but not identical, with moderate differences that fine-tuning can effectively bridge [77]. Tan et al. [78] state that TL relaxes the hypothesis that the training data must be independent and identically distributed (i.i.d.) with the test data. The authors review deep TL techniques, including fine-tuning, which help address distributional shifts between source and target domains. Although bearing-vibration variances exhibited substantially higher kurtosis, suggesting some domain-specific characteristics, the moderate mean shift and non-significant results from all three tests ($\alpha = 0.05$) suggest that the distributions are similar enough to justify the cross-domain pre-training assumption.

6.2. Real-world substation analysis

To assess the applicability of our findings to real-world conditions, we analysed data from operational DH substations (details in Section 3). Fig. 7 presents examples of a daily ΔT signature for eight real-world substations under normal and faulty conditions, and Table 5 ranks scenarios based on stability.

While some faults and normal behaviours can be visually distinguished in real-world scenarios, the signatures exhibit less visible distinction compared to the lab-generated data. This reduced clarity makes

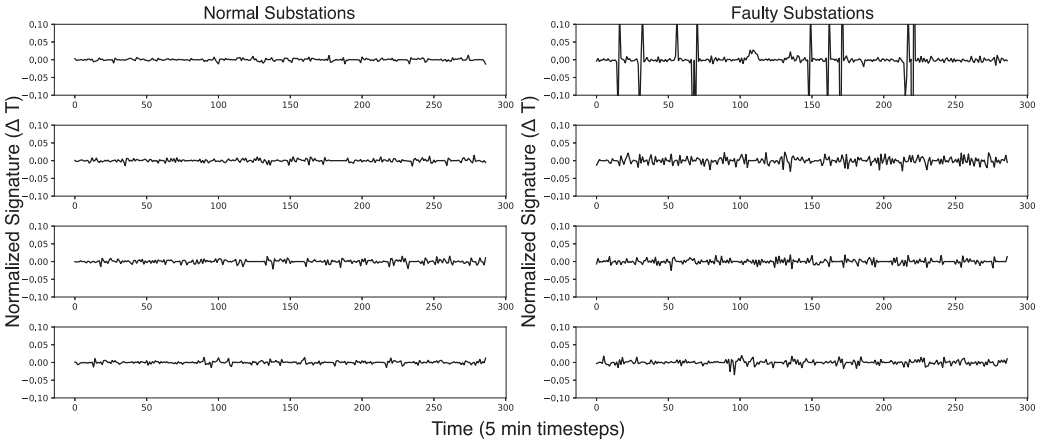


Fig. 7. Examples of daily temperature differential (ΔT) signatures from eight substations of the real-world DH network (5-min timesteps). Left: Normal operation, showing tighter fluctuations around zero mean. Right: Faulty operation, with markedly higher variance around the mean, highlighting the pronounced step-to-step excursions induced by faults.

Table 5
Average ranking of scenarios for Week 1 to Week 4 using ΔT and T_r Stability. We denote the top 5 best-performing scenarios per signature type in bold. Fault scenarios include HHC, WSP, OCV, and LSL.

Scenario	ΔT signature					T_r signature				
	Average rank	W1	W2	W3	W4	Average rank	W1	W2	W3	W4
Normal01	**1.00**	**1**	**1**	**1**	**1**	**1.00**	**1**	**1**	**1**	**1**
Normal02	**3.50**	**4**	**3**	**3**	**4**	**5.25**	**4**	**6**	**5**	**6**
Normal03	**3.50**	**3**	**4**	**4**	**3**	6.25	5	7	6	7
Normal04	**2.00**	**2**	**2**	**2**	**2**	**2.50**	**2**	**3**	**3**	**2**
Normal05	**5.75**	**6**	**6**	**6**	**5**	10.50	10	10	11	11
HHC01	9.75	8	13	9	9	10.50	9	13	10	10
HHC02	6.50	5	7	7	7	11.50	11	11	12	12
WSP01	9.75	7	12	10	10	**4.75**	**6**	**4**	**4**	**5**
WSP02	13.50	10	15	15	14	14.75	14	15	15	15
WSP03	7.50	14	5	5	6	9.25	13	8	8	8
WSP04	14.50	15	14	14	15	14.25	15	14	14	14
OCV01	11.25	12	11	11	11	12.50	12	12	13	13
LSL01	8.25	9	8	8	8	**2.50**	**3**	**2**	**2**	**3**
LSL02	12.00	13	10	13	12	5.75	7	5	7	4
LSL03	11.25	11	9	12	13	8.75	8	9	9	9
Normal	**3.15**	–	–	–	–	**5.10**	–	–	–	–
Faulty	10.43	–	–	–	–	9.45	–	–	–	–

manual interpretation challenging, highlighting the need for advanced methods like DL models. Such models are adept at identifying subtle, salient features in complex datasets, enabling reliable FDD even when the human eye does not discern the patterns.

We employed the Mann–Whitney U test [79], a non-parametric rank-sum test for comparing two independent distributions. A p-value below our threshold of $\alpha = 0.05$ indicates the observed difference is unlikely due to chance. We found that both ΔT ($U = 12.0; p < 0.001$) and T_r ($U = 168.0; p < 0.001$) differ significantly between normal and faulty substations. These results demonstrate that the distributional patterns in the faulty group are systematically distinct from the normal group over the four weeks.

However, as observed in Table 5, it is essential to note that not all faults impact the T_r signature substantially. For instance, the LSL shows low variation in their T_r signatures, which may suggest that this feature is not optimal for detecting their impact on the DH system. These issues are mitigated when analysing the ΔT signature, where all normal scenarios consistently rank higher in stability than faulty conditions. Despite potential uncertainties in the labelling accuracy of real-world data, these observations suggest that the phenomena observed in the lab are also present in operational systems and that faults cause instability in system dynamics.

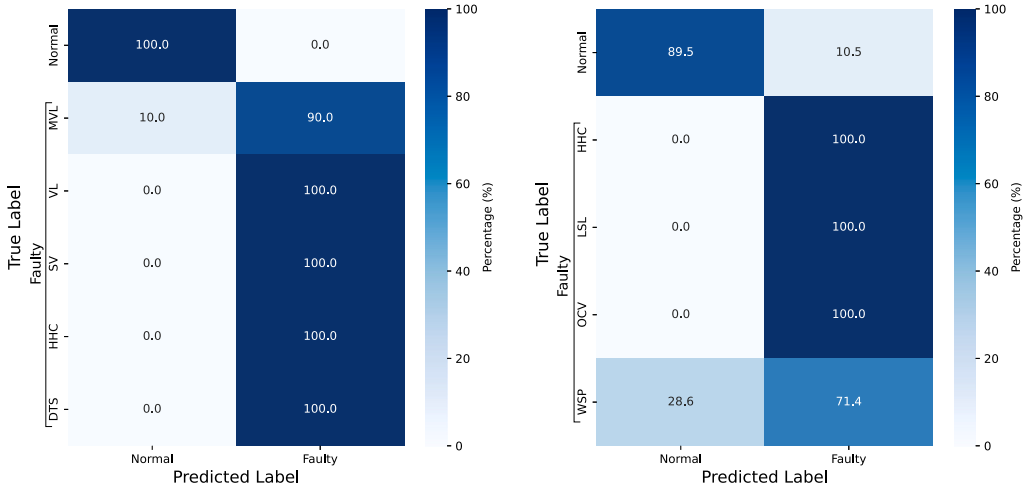
6.3. Model evaluation

We tested several ML models on both lab and real-world datasets to evaluate the effectiveness of our proposed TL fault detection approach. We summarise the performance metrics based on the unseen test data in Table 6. The OCSVM++ model used for fault detection was configured with a radial basis function kernel, a γ value of 0.89 to control the influence of individual data points, and a ν value of 0.7 to specify the upper bound on the fraction of training errors and the lower bound on the fraction of support vectors.

We compared the F1 scores of our best-performing TL models to those of OCSVM and IF reported in [25]. We assessed statistical significance with the Mann–Whitney U test ($\alpha = 0.05$) and effect sizes with Cohen's d. We found that the VAE model significantly outperforms both baselines with large effect sizes (vs. OCSVM: $p = 0.002; d = 6.49$ and vs. IF: $p = 0.005; d = 2.69$, respectively). Consequently, the AE model significantly outperforms OCSVM with a large effect while showing a moderate, non-significant effect against IF (vs. OCSVM: $p < 0.001; d = 5.38$, and vs. IF: $p = 0.533; d = 0.64$, respectively). These findings strongly suggest that our TL approach has a positive effect on model performance. However, while the lab emulation data is crucial to our research, it is essential to note that performance may be optimistic, given the noise-free nature of the lab emulation dataset compared to

Table 6
Performance metrics in percentages for OCSVM++, AE, and VAE (Highest scores in bold).

Data	Type	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Lab	T_r	OCSVM++	97.2 ± 4.30	97.2 ± 4.30	**100 ± 0.00**	98.5 ± 2.28
		AE	83.1 ± 4.82	98.0 ± 3.92	82.0 ± 3.92	89.2 ± 2.93
		VAE	**98.3 ± 3.33**	**100 ± 0.00**	98.0 ± 3.92	**98.9 ± 2.16**
	ΔT	OCSVM++	**90.1 ± 9.88**	94.0 ± 5.99	**94.0 ± 5.99**	**94.0 ± 5.99**
		AE	86.4 ± 4.44	**100 ± 0.00**	84.0 ± 5.23	91.2 ± 5.23
		VAE	84.7 ± 3.33	**100 ± 0.00**	82.0 ± 3.92	90.1 ± 2.16
Real-world	T_r	OCSVM++	73.2 ± 0.68	**82.2 ± 1.66**	75.2 ± 1.66	78.8 ± 0.28
		AE	**75.8 ± 1.02**	74.0 ± 3.39	78.2 ± 1.02	78.2 ± 1.02
		VAE	70.0 ± 1.44	71.1 ± 1.07	**90.5 ± 1.50**	**79.4 ± 1.14**
	ΔT	OCSVM++	72.2 ± 1.11	70.0 ± 0.88	**100 ± 0.00**	82.6 ± 0.55
		AE	**88.7 ± 0.47**	**94.0 ± 0.23**	88.6 ± 0.70	**90.9 ± 0.23**
		VAE	85.4 ± 1.47	92.5 ± 1.31	86.5 ± 3.07	88.5 ± 1.31



(a) Lab dataset (b) Real-world dataset

Fig. 8. Contingency tables showing the performance of the best-performing model for each dataset. (a) Results for the VAE model using the T_r signature on the lab dataset, and (b) results for the AE model using the ΔT signature on the real-world dataset.

real-world operational datasets. Therefore, we should be cautious when generalising these results to operational environments.

However, when applied to the real-world dataset, the models generalised effectively, particularly when using the ΔT signature. The AE achieved an F1 score of approximately 91%, and the VAE approximately 89%, indicating strong fault detection performance under real-world conditions. This is noteworthy, given the inherent noise, variability, and potential confounding factors in real-world operational data. For instance, our results contrast with [58], where the authors encountered significant difficulties maintaining performance when transitioning from controlled experiments to real-world scenarios.

Furthermore, while the OCSVM++ demonstrated strong performance on laboratory data, this highlights the inherent limitations of such controlled datasets. In laboratory conditions, boundaries between classes are typically well-defined, allowing the OCSVM++ to establish accurate decision boundaries. However, a significant performance drop is observed when transitioning to real-world data, where the data distributions are inherently more complex and overlapping.

The contingency table, presented in Fig. 8, shows the classification outcomes in terms of percentages for each fault type (TP, FN, TN, and FP). The VAE demonstrated near-perfect performance for the lab data, with a misclassification observed only in the MVL scenario. This suggests that the scenarios exhibit distinct behaviour patterns, which the model successfully identifies and leverages for classification. However, as mentioned earlier, this high accuracy can be attributed to the controlled, noise-free nature of the lab environment, which simplifies the modelling process compared to real-world scenarios. For the real-world data in the normal category, 89.5% of instances were correctly classified as normal (TN), while 10.5% were misclassified as faulty (FP). All instances for HHC, LSL, and OCV fault types were correctly identified, resulting in 100% TP rates with no misclassification. This suggests that these faults have distinctive signatures and substantial impact on ΔT that the model can effectively recognise. For the WSP fault type, 71.4% of faulty samples were correctly classified (TP), while 28.6% were incorrectly classified as normal (FN). These results demonstrate the model's high performance in identifying most fault types while maintaining reasonable specificity for normal behaviour. However, the misclassification rate for WSP indicates a potential area for improvement in distinguishing this fault type from normal behaviour.

The superior performance of the ΔT signature dynamics suggests that it may be a more robust and reliable feature for fault detection across various fault scenarios, including those like LSL that may not substantially affect the T_r signature. One potential explanation for this success is the smaller domain gap between the source domain (e.g., faulty bearings in mechanical systems) and the target domain (DH data) when using ΔT. Moreover, the ΔT signature may capture system dynamics that are more generalisable across different fault types. In contrast, features such as the T_r may exhibit sensitivity to only a limited subset of faults, reducing their reliability in diverse fault detection tasks. Interestingly, the results indicate that substations exhibit behaviour akin to faulty bearings, where ΔT demonstrates vibration-like patterns. This is a promising find, suggesting that ΔT can be effectively used for fault detection. Furthermore, the observed patterns in ΔT could also be crucial in fault diagnosis, as specific faults may correspond to distinct behavioural patterns. Nevertheless, the

results strongly suggest that both the ΔT and T_r relative dynamics are reliable features for detecting faulty substations (i.e., substations with operational faults have more abrupt system changes) showing unstable performance across its ΔT and T_r relative dynamics, as compared to normal operation substations. One of our key contributions is a TL pipeline that overcomes the scarcity of labelled DH fault data while training a robust model. First, we pre-train the models on abundant, well-annotated faulty-bearing signals. Second, we fine-tune these models on ground-truth lab-emulated data. Third, we deploy the resulting models on real-world DH substation data. By leveraging controlled lab experiments to bridge domain gaps, this approach delivers more reliable and confident fault detection in operational networks despite limited field annotations.

Beyond these technical advances, our approach yields clear operational and economic benefits. Early detection using ΔT and T_r signatures enables strategic interventions that maintain optimal ΔT, reduce return temperatures, and limit the impact of faults. Lowering distribution temperatures translates directly into cost savings [80]. Moreover, stable operation at reduced return temperatures decreases network heat losses and facilitates the integration of low-grade renewable and recovered heat sources [81]. Not only would our proposed method contribute to this goal, but it may also facilitate the systematic collection of annotated real-world DH fault data, e.g., through targeted maintenance, addressing the substantial challenge of labelled data scarcity, which currently slows the development of more advanced, data-driven diagnostic models [14]. Thus, these findings can inform utilities' decisions on sensor deployment, data infrastructure investment, and control strategy design, ultimately supporting cost-effective, resilient, and low-carbon DH networks.

6.4. Limitations

While this study demonstrates promising results for fault detection in DH substations, several limitations must be acknowledged:

- **Uncertain fault persistence in field data.** The real-world data used in this study introduces uncertainties, particularly regarding the persistence of faults. For instance, it is unclear whether faults were consistently present throughout the four-week data collection period. This ambiguity may influence the observed fault detection performance, as seen in Table 5, potentially skewing the results.
- **Limited fault diversity.** The models were trained on a limited range of fault types that may not fully capture the diversity of faults occurring in DH systems. This could restrict the model's ability to generalise to unseen or rare fault types in operational settings.
- **Data scarcity.** The scarce labelled dataset, which, while highly valuable and challenging to obtain in DH, introduces constraints on the evaluation of model performance and, therefore, may be subjected to variability, potentially leading to underestimation or overestimation of the model's true performance.
- **Economic and operational integration.** We have not yet evaluated how detected faults would be routed into real-time process optimisation or network-level heat-supply strategies. Likewise, quantifying the trade-off between investing in higher-resolution sensors, richer data annotation processes, and the resulting gains in energy efficiency and maintenance savings remains an open research question. Addressing these aspects will require collaboration with DH utilities to upgrade field instrumentation, establish systematic labelling workflows, and perform cost–benefit analyses tailored to specific network technologies and business models.

7. Conclusion

In this study, we proposed an emulation-driven fault detection approach for DH substations, leveraging TL to bridge the gap between controlled laboratory data and real-world operational conditions. The methodology involved pre-training models on a source domain with abundant fault data, such as faulty bearings, and fine-tuning them on lab-generated DH data. This process enables the models to transfer knowledge from lab experiments to real-world scenarios, effectively addressing the challenges of scarce and uncertain labelled fault data in operational DH environments. Moreover, this research introduces an innovative approach to fault signature generation, which captures the characteristics of substations' daily operational patterns across defined temporal intervals.

We evaluated several models, including OCSVM++, AE, and VAE, using both T_r and ΔT signatures. Results indicate the ΔT signature is more robust for fault detection across diverse conditions than the T_r signature. Notably, the AE model achieved an F1 score of approximately 91% on real-world data using the ΔT signature, demonstrating strong generalisation capabilities in complex, real-world DH substations. This level of performance is particularly significant given the inherent noise, variability, and confounding factors in operational DH environments. Our TL-based approach highlights the practical potential of integrating simulation-driven learning into real-world DH applications. By combining domain knowledge from well-labelled source domains with experimental DH data, we established a robust framework to mitigate data limitations in the DH field, offering a viable strategy for proactive maintenance, reduced downtime, and enhanced system efficiency.

Beyond these technical advances, our focus on relative temperature dynamics highlights that observing the step-to-step correlations enables automated fault detection in modern DH systems, given sampling frequencies lower than hourly (e.g., 15 min). By facilitating early fault detection and targeted interventions, we can prevent unnecessary energy losses and inefficient resource allocation while, in parallel, gathering high-quality, fault-annotated operational data to support the development of more advanced diagnostic methods in DH. Moreover, the controlled lab experiments highlight which data frequencies and sensor signals are most informative for reliable detection and diagnosis. These insights can directly inform utility decisions on sampling intervals, data storage, and instrumentation investments, helping to design cost-effective, high-performance monitoring systems for next-generation DH grids.

Future work will extend our approach beyond fault detection to fault diagnosis, enabling the identification of fault types and severity. This advancement would provide deeper insights into the operational health of DH substations and support more targeted maintenance strategies. Additionally, efforts will be directed towards improving the performance of our method, for instance, by reducing the domain shift of faulty bearings, lab, and real-world DH data, and investigate integration of additional sensor measurements.

CRediT authorship contribution statement

Jonne van Dreven: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Abbas Cheddad:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis. **Sadi Alawadi:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis. **Ahmad Nauman Ghazi:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis. **Jad Al Koussa:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Formal analysis, Data curation. **Dirk Vanhoudt:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge and thank *Runa Smart Equipment* Co. Ltd. (China) for providing the district heating data essential for this research. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The authors do not have permission to share data.

References

[1] Lund H, Østergaard PA, Nielsen TB, Werner S, Thorsen JE, Gudmundsson O, Arabkoohsar A, Mathiesen BV. Perspectives on fourth and fifth generation district heating. Energy 2021;227:120520. http://dx.doi.org/10.1016/j.energy.2021.120520.

[2] Marszal-Pomianowska A, Motoasca E, Pothof I, Felsmann C, Heiselberg P, Cadenbach A, Leusbrock I, O'Donovan K, Petersen S, Schaffer M. Strengths, weaknesses, opportunities and threats of demand response in district heating and cooling systems. from passive customers to valuable assets. Smart Energy 2024;14:100135. http://dx.doi.org/10.1016/j.segy.2024.100135.

[3] Leoni P, Geyer R, Schmidt R-R. Developing innovative business models for reducing return temperatures in district heating systems: Approach and first results. Energy 2020;195:116963. http://dx.doi.org/10.1016/j.energy.2020.116963.

[4] Rafati A, Tahavori M, Shaker HR. Data-driven reliability analysis of district heating systems for asset management applications: A review. Sustain Cities Soc 2024;106052. http://dx.doi.org/10.1016/j.scs.2024.106052.

[5] Månsson S, Johansson Kallioniemi P-O, Thern M, Van Oevelen T, Sernhed K. Faults in district heating customer installations and ways to approach them: Experiences from swedish utilities. Energy 2019;180:163–74. http://dx.doi.org/10.1016/j.energy.2019.04.220.

[6] Gadd H, Werner S. Fault detection in district heating substations. Appl Energy 2015;157:51–9. http://dx.doi.org/10.1016/j.apenergy.2015.07.061.

[7] David A, Mathiesen BV, Averfalk H, Werner S, Lund H. Heat roadmap europe: Large-scale electric heat pumps in district heating systems. Energies 2017;10(4). http://dx.doi.org/10.3390/en10040578.

[8] Schmidt D, et al., editors. Guidebook for the digitalisation of district heating: transforming heat networks for a sustainable future. Germany: AGFW Project Company, Frankfurt am Main; 2023, p. 67, Final Report of DHC Annex TS4.

[9] Månsson S, Davidsson K, Lauenburg P, Thern M. Automated statistical methods for fault detection in district heating customer installations. Energies 2019;12(11):113. http://dx.doi.org/10.3390/en12010113.

[10] Honoré K. The age of digitalization and flexibility - from consumer to flexumer in the district heating system. In: 9th international conference on smart energy systems. Copenhagen; 2023.

[11] Gadd H, Werner S. Achieving low return temperatures from district heating substations. Appl Energy 2014;136:59–67. http://dx.doi.org/10.1016/j.apenergy.2014.09.022.

[12] Lund H, Østergaard PA, Chang M, Werner S, Svendsen S. The status of 4th generation district heating: Research and results. Energy 2018;164:147–59. http://dx.doi.org/10.1016/j.energy.2018.08.206.

[13] Lund H, Thorsen JE, Jensen SS, Madsen FP. Fourth-generation district heating and motivation tariffs. ASME Open J Eng 2022;1(011002). http://dx.doi.org/10.1115/1.4053420.

[14] van Dreven J, Boeva V, Abghari S, Grahn H, Al Koussa J, Motoasca E. Intelligent approaches to fault detection and diagnosis in district heating: Current trends, challenges, and opportunities. Electronics 2023;12(66):1448. http://dx.doi.org/10.3390/electronics12061448.

[15] Rafati A, Shaker HR. Predictive maintenance of district heating networks: A comprehensive review of methods and challenges. Therm Sci Eng Prog 2024;53:102722. http://dx.doi.org/10.1016/j.tsep.2024.102722.

[16] Leiria D, Andersen KH, Melgaard SP, Johra H, Marszal-Pomianowska A, Piscitelli MS, Capozzoli A, Pomianowski MZ. Towards automated fault detection and diagnosis in district heating customers: generation and analysis of a labeled dataset with ground truth. In: Proceedings of building simulation 2023: 18th conference of IBPSA, BS 2023. IBPSA, 2023, http://dx.doi.org/10.26868/25222708.2023.1576.

[17] Frederiksen S, Werner S. District heating and cooling. Lund, Sweden: Studentlitteratur; 2013.

[18] Schaffer M, Veit M, Marszal-Pomianowska A, Frandsen M, Pomianowski MZ, Dichmann E, Sørensen CG, Kragh J. Dataset of smart heat and water meter data with accompanying building characteristics. Data Brief 2024;52:109964. http://dx.doi.org/10.1016/j.dib.2023.109964.

[19] van Dreven J, Cheddad A, Alawadi S, Ghazi AN, Al Koussa J, Vanhoudt D. From data scarcity to diagnostic precision: A novel data augmentation and fault diagnosis framework for district heating substations. Eng Appl Artif Intell 2025;151:110662. http://dx.doi.org/10.1016/j.engappai.2025.110662.

[20] Hermans C, Al Koussa J, Van Oevelen T, Vanhoudt D. Fault detection for district heating substations: Beyond three-sigma approaches. Smart Energy 2024;16:100159. http://dx.doi.org/10.1016/j.segy.2024.100159.

[21] Neumayer M, Stecher D, Grimm S, Maier A, Bücker D, Schmidt J. Fault and anomaly detection in district heating substations: A survey on methodology and data sets. Energy 2023;276:127569. http://dx.doi.org/10.1016/j.energy.2023.127569.

[22] Månsson S, Lundholm Benzi I, Thern M, Salenbien R, Sernhed K, Johansson Kallioniemi P-O. A taxonomy for labeling deviations in district heating customer data. Smart Energy 2021;2:100020. http://dx.doi.org/10.1016/j.segy.2021.100020.

[23] Stecher D, Neumayer M, Ramachandran A, Hort A, Maier A, Bücker D, Schmidt J. Creating a labeled district heating data set: From anomaly detection towards fault detection. Energy 2024;313:134016. http://dx.doi.org/10.1016/j.energy.2024.134016.

[24] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. Proc IEEE 2021;109(1):43–76. http://dx.doi.org/10.1109/JPROC.2020.3004555.

[25] van Dreven J, Boeva V, Abghari S, Grahn H, Al Koussa J. A systematic approach for data generation for intelligent fault detection and diagnosis in district heating. Energy 2024;307:132711. http://dx.doi.org/10.1016/j.energy.2024.132711.

[26] Helbing G, Ritter M. Deep learning for fault detection in wind turbines. Renew Sustain Energy Rev 2018;98:189–98. http://dx.doi.org/10.1016/j.rser.2018.09.012.

[27] Saufi SR, Ahmad ZAB, Leong MS, Lim MH. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. IEEE Access 2019;7:122644–62. http://dx.doi.org/10.1109/ACCESS.2019.2938227.

[28] Zhang F, Saeed N, Sadeghian P. Deep learning in fault detection and diagnosis of building hvac systems: A systematic review with meta analysis. Energy AI 2023;12:100235. http://dx.doi.org/10.1016/j.egyai.2023.100235.

[29] Yu J, Zhang Y. Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review. Neural Comput Appl 2023;35(1):211–52. http://dx.doi.org/10.1007/s00521-022-08017-3.

[30] Chen Z, O'Neill Z, Wen J, Pradhan O, Yang T, Lu X, Lin G, Miyata S, Lee S, Shen C, Chiosa R, Piscitelli MS, Capozzoli A, Hengel F, Kührer A, Pritoni M, Liu W, Clauß J, Chen Y, Herr T. A review of data-driven fault detection and diagnostics for building hvac systems. Appl Energy 2023;339:121030. http://dx.doi.org/10.1016/j.apenergy.2023.121030.

[31] Qian J, Song Z, Yao Y, Zhu Z, Zhang X. A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes. Chemometr Intell Lab Syst 2022;231:104711. http://dx.doi.org/10.1016/j.chemolab.2022.104711.

[32] Schaffer M, Tvedebrink T, Marszal-Pomianowska A. Three years of hourly data from 3021 smart heat meters installed in danish residential buildings. Sci Data 2022;9(1):420. http://dx.doi.org/10.1038/s41597-022-01502-3.

[33] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59. http://dx.doi.org/10.1109/TKDE.2009.191.

[34] Chen H, Luo H, Huang B, Jiang B, Kaynak O. Transfer learning-motivated intelligent fault diagnosis designs: A survey, insights, and perspectives. IEEE Trans Neural Netw Learn Syst 2024;35(3):2969–83. http://dx.doi.org/10.1109/TNNLS.2023.3290974.

[35] Li C, Zhang S, Qin Y, Estupinan E. A systematic review of deep transfer learning for machinery fault diagnosis. Neurocomputing 2020;407:121–35. http://dx.doi.org/10.1016/j.neucom.2020.04.045.

[36] Chavan K, Réhault N, Rist T. Transfer learning methodology for machine learning based fault detection and diagnostics applied to building services. J Phys: Conf Ser 2023;2600(8):082038. http://dx.doi.org/10.1088/1742-6596/2600/8/082038.

[37] Wang C, Yuan J, Huang K, Zhang J, Zheng L, Zhou Z, Zhang Y. Research on thermal load prediction of district heating station based on transfer learning. Energy 2022;239:122309. http://dx.doi.org/10.1016/j.energy.2021.122309.

[38] Vallee M, Wissocq T, Gaoua Y, Lamaison N. Generation and evaluation of a synthetic dataset to improve fault detection in district heating and cooling systems. Energy 2023;283:128387. http://dx.doi.org/10.1016/j.energy.2023.128387.

[39] Zhang F, Fleyeh H. Anomaly detection of heat energy usage in district heating substations using lstm based variational autoencoder combined with physical model. In: 2020 15th IEEE conference on industrial electronics and applications. ICIEA, 2020, p. 153–8. http://dx.doi.org/10.1109/ICIEA48937.2020.9248108.

[40] Choi Y, Yoon S. Autoencoder-driven fault detection and diagnosis in building automation systems: Residual-based and latent space-based approaches. Build Environ 2021;203:108066. http://dx.doi.org/10.1016/j.buildenv.2021.108066.

[41] Hong Y, Yoon S, Kim Y-S, Jang H. System-level virtual sensing method in building energy systems using autoencoder: Under the limited sensors and operational datasets. Appl Energy 2021;301:117458. http://dx.doi.org/10.1016/j.apenergy.2021.117458.

[42] Shahid ZK, Saguna S, Åhlund C. Autoencoders for anomaly detection in electricity and district heating consumption: A case study in school buildings in sweden. In: 2023 IEEE international conference on environment and electrical engineering and 2023 IEEE industrial and commercial power systems europe (EEEIC / i & CPS europe). 2023, p. 1–8. http://dx.doi.org/10.1109/EEEIC/ICPSEurope57605.2023.10194605.

[43] Shahid ZK, Saguna S, Åhlund C. Variational autoencoders for anomaly detection and transfer knowledge in electricity and district heating consumption. IEEE Trans Ind Appl 2024;60(5):7437–50. http://dx.doi.org/10.1109/TIA.2024.3425805.

[44] Pereira J, Silveira M. Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In: 2018 17th IEEE international conference on machine learning and applications. ICMLA, 2018, p. 1275–82. http://dx.doi.org/10.1109/ICMLA.2018.00207.

[45] Al-Abassi A, Sakhnini J, Karimipour H. Unsupervised stacked autoencoders for anomaly detection on smart cyber–physical grids. In: 2020 IEEE international conference on systems, man, and cybernetics. SMC, 2020, p. 3123–9. http://dx.doi.org/10.1109/SMC42975.2020.9283064.

[46] Pydi DP, Advaith S. Attention boosted autoencoder for building energy anomaly detection. Energy AI 2023;14:100292. http://dx.doi.org/10.1016/j.egyai.2023.100292.

[47] El Mokhtari K, McArthur JJ. Autoencoder-based fault detection using building automation system data. Adv Eng Inform 2024;62:102810. http://dx.doi.org/10.1016/j.aei.2024.102810.

[48] Chen Z, Yeo CK, Lee BS, Lau CT. Autoencoder-based network anomaly detection. In: 2018 wireless telecommunications symposium. WTS, 2018, p. 1–5. http://dx.doi.org/10.1109/WTS.2018.8363930.

[49] Said Elsayed M, Le-Khac N-A, Dev S, Jurcut AD. Network anomaly detection using lstm based autoencoder. In: Proceedings of the 16th ACM symposium on qoS and security for wireless and mobile networks. Q2SWinet '20, New York, NY, USA: Association for Computing Machinery; 2020, p. 37–45. http://dx.doi.org/10.1145/3416013.3426457.

[50] Coelho G, Matos LM, Pereira PJ, Ferreira A, Pilastri A, Cortez P. Deep autoencoders for acoustic anomaly detection: experiments with working machine and in-vehicle audio. Neural Comput Appl 2022;34(22):19485–99. http://dx.doi.org/10.1007/s00521-022-07375-2.

[51] Gupta P, Pradhan MK. Fault detection analysis in rolling element bearing: A review. Mater Today: Proc 2017;4(2, Part A):2085–94. http://dx.doi.org/10.1016/j.matpr.2017.02.054.

[52] Zhang S, Ye F, Wang B, Habetler TG. Semi-supervised learning of bearing anomaly detection via deep variational autoencoders. 2019, http://dx.doi.org/10.48550/arXiv.1912.01096, (arXiv:1912.01096), arXiv:1912.01096 [cs, eess, stat].

[53] Ahmad S, Styp-Rekowski K, Nedelkoski S, Kao O. Autoencoder-based condition monitoring and anomaly detection method for rotating machines. In: 2020 IEEE international conference on big data (big data). 2020, p. 4093–102. http://dx.doi.org/10.1109/BigData50022.2020.9378015.

[54] Huang X, Wen G, Dong S, Zhou H, Lei Z, Zhang Z, Chen X. Memory residual regression autoencoder for bearing fault detection. IEEE Trans Instrum Meas 2021;70:1–12. http://dx.doi.org/10.1109/TIM.2021.3072131.

[55] Hakim M, Omran AAB, Ahmed AN, Al-Waily M, Abdellatif A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. Ain Shams Eng J 2023;14(4):101945. http://dx.doi.org/10.1016/j.asej.2022.101945.

[56] Michau G, Fink O. Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. Knowl-Based Syst 2021;216:106816. http://dx.doi.org/10.1016/j.knosys.2021.106816.

[57] CSof Engineering. Data castle bearings dataset. Western Reserve University; 2021, URL https://engineering.case.edu/bearingdatacenter.

[58] Bode G, Thul S, Baranski M, Müller D. Real-world application of machine-learning-based fault detection trained with experimental data. Energy 2020;198:117323. http://dx.doi.org/10.1016/j.energy.2020.117323.

[59] Leiria D, Johra H, Anoruo J, Praulins I, Piscitelli MS, Capozzoli A, Marszal-Pomianowska A, Pomianowski MZ. Is it returning too hot?time series segmentation and feature clustering of end-user substation faults in district heating systems. 2024, http://dx.doi.org/10.2139/ssrn.4894104, (4894104).

[60] Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit 2011;44(3):678–93. http://dx.doi.org/10.1016/j.patcog.2010.09.013.

[61] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978;26(1):43–9. http://dx.doi.org/10.1109/TASSP.1978.1163055.

[62] Hamilton JD. Time series analysis. Princeton University Press; 2020.

[63] Manevitz LM, Yousef M. One-class svms for document classification. J Mach Learn Res 2002;2:139–54.

[64] Rumelhart DE, McClelland JL, Group PR. Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations. 1986, http://dx.doi.org/10.7551/mitpress/5236.001.0001.

[65] Berahmand K, Daneshfar F, Salehi ES, Li Y, Xu Y. Autoencoders and their applications in machine learning: a survey. Artif Intell Rev 2024;57(2):28. http://dx.doi.org/10.1007/s10462-023-10662-6.

[66] Kingma DP, Welling M. Auto-encoding variational bayes. 2022, http://dx.doi.org/10.48550/arXiv.1312.6114, (arXiv:1312.6114), arXiv:1312.6114.

[67] Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis. MLSDA'14, New York, NY, USA: Association for Computing Machinery; 2014, p. 4–11. http://dx.doi.org/10.1145/2689746.2689747.

[68] Gupta V, Choudhary K, Tavazza F, Campbell C, k. Liao W, Choudhary A, Agrawal A. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. Nat Commun 2021;12(1):6595. http://dx.doi.org/10.1038/s41467-021-26921-5.

[69] Cristianini N, Scholkopf B. Support vector machines and kernel methods: The new generation of learning machines. AI Mag 2002;23(33):31. http://dx.doi.org/10.1609/aimag.v23i3.1655.

[70] Day O, Khoshgoftaar TM. A survey on heterogeneous transfer learning. J Big Data 2017;4(1):29. http://dx.doi.org/10.1186/s40537-017-0089-0.

[71] Gao P, Wu W, Li J. Multi-source fast transfer learning algorithm based on support vector machine. Appl Intell 2021;51(11):8451–65. http://dx.doi.org/10.1007/s10489-021-02194-9.

[72] Nazari F, Yan W. Convolutional versus dense neural networks: comparing the two neural networks' performance in predicting building operational energy use based on the building shape. In: Building simulation, vol. 17, IBPSA; 2021, p. 495–502. http://dx.doi.org/10.26868/25222708.2021.30735.

[73] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86.

[74] Xu Q-S, Liang Y-Z. Monte carlo cross validation. Chemometr Intell Lab Syst 2001;56(1):1–11. http://dx.doi.org/10.1016/S0169-7439(00)00122-2.

[75] Fligner MA, Killeen TJ. Distribution-free two-sample tests for scale. J Amer Statist Assoc 1976;71(353):210–3. http://dx.doi.org/10.1080/01621459.1976.10481517.

[76] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Advances in neural information processing systems. vol. 27, Curran Associates, Inc.; 2014, URL https://papers.nips.cc/paper_files/paper/2014/hash/532a2f85b6977104bc93f8580abbb330-Abstract.html.

[77] Tripuraneni N, Jordan M, Jin C. On the theory of transfer learning: The importance of task diversity. Adv Neural Inf Process Syst 2020;33:7852–62.

[78] Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Kůrková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, editors. Rtificial neural networks and machine learning – ICANN 2018. Cham: Springer International Publishing; 2018, p. 270–9. http://dx.doi.org/10.1007/978-3-030-01424-7_27.

[79] McKnight PE, Najab J. Mann–whitney u test. John Wiley & Sons, Ltd; 2010, p. 1. http://dx.doi.org/10.1002/9780470479216.corpsy0524.

[80] Geyer R, Krail J, Leitner B, Schmidt R-R, Leoni P. Energy-economic assessment of reduced district heating system temperatures. Smart Energy 2021;2:100011. http://dx.doi.org/10.1016/j.segy.2021.100011.

[81] Averfalk H, Werner S. Economic benefits of fourth generation district heating. Energy 2020;193:116727. http://dx.doi.org/10.1016/j.energy.2019.116727.