


Article

Opportunities for Machine Learning in District Heating

Gideon Mbiydzenyuy ^{1,*}, Sławomir Nowaczyk ², Håkan Knutsson ³, Dirk Vanhoudt ^{4,5}, Jens Brage ⁶
and Ece Calikus ²¹ Department of Information Technology, University of Borås, SE-501 90 Borås, Sweden² CAISR, University of Halmstad, SE-301 18 Halmstad, Sweden; slawomir.nowaczyk@hh.se (S.N.); ece.calikus@hh.se (E.C.)³ The School of Business, Engineering and Science, University of Halmstad, SE-301 18 Halmstad, Sweden; hakan.knutsson@hh.se⁴ VITO, Boeretang 200, 2400 Mol, Belgium; dirk.vanhoudt@vito.be⁵ EnergyVille, Thor Park 8310, 3600 Genk, Belgium⁶ NODA Intelligent Systems, SE-374 35 Karlshamn, Sweden; jens.brage@noda.se

* Correspondence: gideon.mbiydzenyuy@hb.se

Abstract: The district heating (DH) industry is facing an important transformation towards more efficient networks that utilise significantly lower water temperatures to distribute the heat. This change requires taking advantage of new technologies, and Machine Learning (ML) is a popular direction. In the last decade, we have witnessed an extreme growth in the number of published research papers that focus on applying ML techniques to the DH domain. However, based on our experience in the field, and an extensive review of the state-of-the-art, we perceive a mismatch between the most popular research directions, such as forecasting, and the challenges faced by the DH industry. In this work, we present our findings, explain and demonstrate the key gaps between the two communities and suggest a road-map ahead towards increasing the impact of ML research in the DH industry.



Citation: Mbiydzenyuy, G.; Nowaczyk, S.; Knutsson, H.; Vanhoudt, D.; Brage, J.; Calikus, E. Opportunities for Machine Learning in District Heating. *Appl. Sci.* **2021**, *11*, 6112. <https://doi.org/10.3390/app11136112>

Academic Editor: Francesco Melino

Received: 16 May 2021

Accepted: 17 June 2021

Published: 30 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Machine Learning; district heating; review; road-map; research opportunities

1. Introduction

As the urban space continues to expand, it has become increasingly challenging to adapt existing solutions for the supply of electricity, heating and cooling, water and waste management, without major and often prohibitively costly changes to the underlying infrastructure [1]. The ongoing digitalisation of the energy sector, an otherwise traditional industry, can be used to address some of these challenges.

The digitalisation of the district heating (DH) sector offers opportunities such as new incentives for lowering peak energy consumption [1,2]; optimisation of indoor climate (e.g., temperature levels in homes); optimisation of energy production; and around-the-clock monitoring of substations and the underlying network [3]. Visions such as smart cities, smart grids and 5th Generation District Heating (5GDH) [1,4] are largely driven by the opportunities created by digitalisation under the umbrella term of Industry 4.0 [5,6]. A recurring theme with Industry 4.0 is the use of data intensive methods, often under the label of Machine Learning (ML) [7,8], to gain useful insights from data. It comes in the form of incentives, better indoor climate, optimisation of energy supply and so on, out of otherwise unwieldy data.

The approaches towards applying ML in the DH sector are, however, often quite unstructured and experimental, particularly due to the lack of a structured way of collecting and sharing relevant data or benchmarking ML models. In this article, we attempt to report on the current landscape of ML as applied to DH networks and to discuss possibilities and challenges from the perspectives of the district heating industry and research. Thus, the article is designed as a position paper at the intersection of ML and DH stakeholders.

ML focuses on computational techniques for learning intrinsic relationships, from structured and unstructured data, resulting in models that generalise these relationships for the purpose of analysing new but previously unseen data. The value created by data-intensive methods such as ML inherently depends on the quality of data employed and the data collection cost. While data collection has traditionally been a costly endeavour, developments in sensor technology and networking have lowered such costs with the help of technology such as Internet of Things (IoT) [9]. Data about utilities, such as water, electricity and heating, although often collected for billing, can and are being employed to create additional value, e.g., detecting deviating and sub-optimal behaviours of DH substations [10].

District heating is a technology for distributing heat generated in a centralised location, through a system of insulated pipes, for residential and commercial heating requirements, such as space heating and domestic hot water production. DH has been used in a large scale for more than 100 years [11]. The oldest generation of district heating (1st Generation District Heating or 1GDH) made use of steam distribution systems. The second generation, 2GDH, distributes pressurised hot water (between 80 °C and 150 °C) in hollow concrete ducts and is common in many cities in Europe. In the third generation, 3GDH, pre-insulated pipes, buried in the ground, were developed. Pre-assembled substations were introduced, gradually with more advanced control and automation. 4GDH [12] is distributing hot water at minimum temperature levels, usually around 60 °C (minimum level to prevent spread of legionella bacteria). Currently, ambient loops are being investigated, with neutral distribution temperatures (15–60 °C) in combination with decentralised heat pumps to boost the temperature in buildings. This concept is sometimes referred to as 5GDH [4]. Although today all generations of DH are still in operation, most existing DH systems have undergone transformations in the form of generational adaptations throughout their lifetimes. New advances in technology have characterised the transition between these generations. For example, the 3G to 4G transition was enabled by an information and communication technology layer and associated smart energy infrastructure (<https://www.energyville.be/en/press/expert-talk-digital-district-heating-networks-why-digital>) (accessed on 15 May 2021). In addition, each transformation stage has consistently lowered temperature levels in the networks. As a result, the heat loss is consistently reduced, and new opportunities are created for integrating many low-temperature sustainable energy heat sources. Solar energy, urban waste heat from subways and similar energy sources are often used together in combination with heat pumps to reach desired temperatures. However, with lower temperatures, the system becomes increasingly sensitive to the performance of each part, such as heat exchangers, which then have to operate across correspondingly lower temperature differences.

Densely populated environments are accelerating urban sprawl, causing a major constrain on older DH networks, initially dimensioned to operate in a limited geographic area. For DH networks to remain effective, they need to adapt to such expansions while, at the same time, meeting new requirements, e.g., higher expected standards of service, decarbonisation, etc. Such adaptations can be supported through numerous innovative technologies, including early detection of faults or increased efficiency of operations. Digitalisation and data driven approaches, such as ML, offer opportunities for improving many aspects of DH operations. With newer generations (4GDH and 5GDH), it is increasingly possible to collect data that can be used to increase heat energy utilisation, with better indoor climate, while optimising production and monitoring capabilities. This article sets an ambitious goal to contribute towards advancing our understanding of how ML algorithms can be employed to address the most relevant issues limiting the adaptation of DH to increasingly stricter requirements in society.

Motivation and Objective

Our objective was to understand and discuss to what extent the current state-of-the-art ML has addressed the key problems in the DH industry. Knowledge about how advances

in ML have addressed relevant problems in DH is essential to determine any (mis)matches between these two fields. Such knowledge will help establish an anchor for future ML and DH research. By employing a literature review and expert insights, the article aims to serve as a reference to the current state-of-the-art, while identifying gaps, challenges and future opportunities.

With a market share of 50% in Nordic countries (with limited use of fossil fuel sources), 2–10% in Western European countries (with still a high reliance on fossil fuel sources) or 55% in northern China or Korea and stream systems in the US, it can be expected that adapting old generations of DH systems alongside newer generations will remain a challenge well into the future. The difficulty of adapting generations of DH to meet new demands requires both policy instruments, as well as innovative solutions made available through research. For example, the European Union (EU) is promoting policies related to decarbonisation of heating and thermal processes [13,14], which forms a key incentive towards utilising alternative fuel sources. ML is mainly driven by the ability to collect and store a vast amount of data and the increasingly available low cost cloud-based computational resources. It has rapidly been adopted in many process industries, but also other areas with easy access to structured data, such as road traffic and health care sectors, and increasingly for processing unstructured datasets, such as images and documents. ML offers several opportunities for problems related to monitoring of operations and optimisation. However, it has become increasingly clear that advances in DH-related ML research have a disproportionate focus on some problem areas, specifically forecasting [15–18], while missing out on several interesting problems, e.g., substation heat transfer, occupancy comfort, etc.

By comparing and contrasting knowledge synthesised from the above two fields, we propose a road-map for research in ML and DH, anchored on identified gaps in the current state-of-the-art. Similar recent efforts identified vital underlying factors influencing thermal load pattern forecasting in DH networks, alongside the most efficient ML algorithms applied for District Heating and Cooling (DHC) load/demand forecasting [19]. In the current article, we further look at problems of interest from the perspective of the industry and we propose a road-map for increasing the impact of ML applications in DH.

2. District Heating Networks

DH as a technology is designed to operate for decades, although the environment in which those systems operate often evolves faster than expected with changes in population. All major DH systems in Europe were built in the period of 1950–1980, and designed for a 20–30 year economic lifetime, although today many of these remain in operation. With continued urbanisation accelerating the urban sprawl, these networks need to be adapted to meet additional and growing demands. This primarily concerns higher expected standards of service, generally achieved through improved monitoring, planning (lowered need for overcapacity) and other forms of optimisation.

The process of delivering heat involves several sub-processes that together constitute the DH value chain:

1. Supply (production) of heat from, e.g., Combined Heat and Power (CHP) plants, boilers or recovery of surplus heat;
2. Distribution of heat;
3. Transfer of heat to customer (substation, mostly owned by customer);
4. Consumption of heat at residential homes, offices, public buildings and industries/firms.

For each of the processes in the value chain, the heating networks need to adapt to overcome several challenges, originating from a number of trends, such as:

A shift from monopolies to a competitive market. DH has been monopolised (similar to power and water distribution) in most countries but is gradually deregulated and exposed to competition. The transition toward a competitive DH market is well underway and will only become more serious in the future. DH, as a technology, is desired by the EU, but the question is whether DH companies will be able to prevail and flourish in the rapid

shift towards electrification that will intensify competition, e.g., with heat pumps. Such a competitive market will result in vertical and horizontal competition between different actors at different levels of the value chain.

A shift from a local, centralised to distributed heat supply. A transition from monopolies to competitive markets will also result in more decentralised DH networks. Large central plants will gradually be replaced by decentralised heat sources involving multiple actors. To meet the increasing demand created by the urban sprawl, distributed supply networks will be a better alternative compared to centralised supply.

Cross-sector city collaboration, between city planning, construction and the infrastructure utilities (heating, cooling, water, sewage water, fibre optic and electric power), will need to be improved to share investment costs, thus adding new actors and further decentralisation of DH systems. Such distributed and decentralised supply networks will create new opportunities for incentives as a result of competition.

A shift from infrastructure centric to a process and service centric industry. A decentralised and competitive DH network will empower customers by offering them the possibility to select their heat provider. As such, the DH network will have to shift focus from infrastructure centric to a processes and service centric industry. The traditional front-end of DH has been the border of the customer property, which means that the substation is owned by the customer. From efficiency and operation points of view, this division of ownership has caused excessive pumping of hot water (and more heat losses), due to poor performance of the substations and the secondary heat system of the customers. In traditional DH operations, the inefficiency cost has partially been neglected. In the adoption of 4GDH, with lower temperatures, this inefficiency is intolerable. In parallel, the DH companies must come closer to their customers and offer them new services (generating new revenues and reducing costs). The challenge of Capital Expenditures (CAPEX) is often neglected in the extremely infrastructure intensive DH industry [20] and by the academic community. Large production plants are expensive but will gradually be phased out when the 4GDH and 5GDH are phased in.

A shift from high temperature infrastructure to low temperature infrastructure. Reduction of temperatures in DH supply and distribution is important to meet one of the major challenges to DH decarbonisation. Lower temperatures allow for capturing more surplus heat, making DH more circular. Surplus heat can substitute the use of primary energy, fossil fuels and renewable energy. Lower network temperatures will also result in less heat losses. Furthermore, decentralised systems will generally reduce the distance between the energy production and consumption, thus providing an opportunity to reduce heat losses even further.

2.1. Operational Challenges in Adapting District Heating Networks

From the Machine Learning perspective, there are two broad directions in which data-driven techniques can support the DH management and lead to more efficient operations:

1. Control—as a material flow process, a large portion of DH systems rely on improving control mechanisms at various points in the value chain. Optimal control comes down to the determination of the best control signals for the four levels of control in DH networks [11], as shown in Figure 1. Inadequate incentives and control strategies often result in wasteful energy supply margins.
2. Analytics—insufficient use of analytic tools for evaluating the result from a combination of control strategies leads to lack of information about the true performance of the DH system. By improving the use of analytics, more comprehensive situation assessment can be used in planning and feedback to the control units, akin to model predictive control systems.

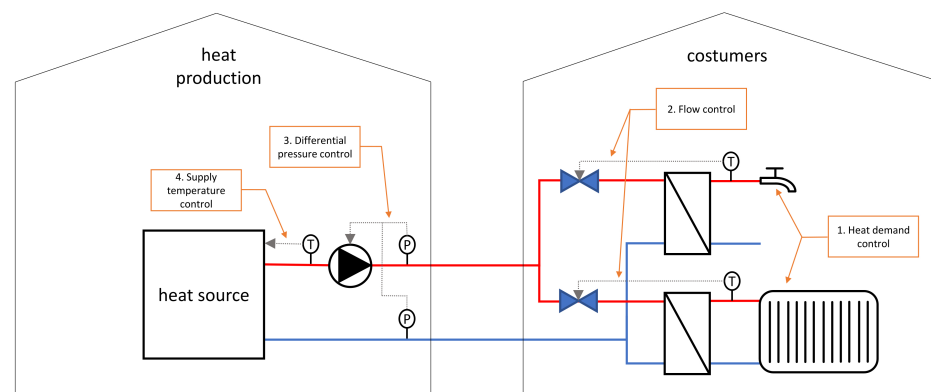


Figure 1. Typically, the control of heat consumption and heat production in DH networks is achieved by means of 4 independent, non-interacting control actions. Loop #1 controls the heat consumed by the individual buildings, #2 controls the individual flow rate of the DH water being consumed. By aggregating individual building consumption, numbers #1 and #2) also determine the heat demand and flow rate in the entire network. At the production side, #3 controls the pressure levels in the network and #4 controls the supply temperature of the network to fulfill the heat demand and flow rates from the buildings [21].

Control strategies involve active interaction in the operation of the network in real-time, such as modifying the flow rates, the temperature levels or the power flows, depicted in Figure 1, to meet a certain objective for either a subsystem, the global network or even the energy system level. Analytics, on the other hand, involve off-line supervision of the network behaviour. There is no active real-time intervention in the network operation, but monitoring and interpreting the overall behaviour of the network provides knowledge and insights useful to optimise the efficiency, sustainability or profitability of the network. We present some specific inefficiencies that are common at various phases in the value chain as follows:

2.1.1. Heat Supply

In this phase, heat energy is produced from CHP plants, boilers, geothermal, solar or recovery of surplus heat. In the production process, a number of control challenges need to be addressed:

1. Unit commitment, e.g., choice of boilers to switch on. This relates to the sequence of production units to start or stop, where each one can also be controlled continuously within their own limits. In a CHP system, units could also include heat supply generated from solar cells whereby their output production is uncertain. Further, the amount of energy stored in the systems (thermal inertia) needs to be taken into consideration to fully understand the heat supply capacity. To be able to solve these problems in a good way, the heat load should be predicted, as well as the production for each of the supply units in a CHP system. Short-term heat load forecasting is therefore very important, since, without knowing how the heat demand of a network will evolve in the coming hours, it is impossible to select which units are the most efficient ones to switch on. In case large CHP units or heat pumps are present, which trade their power on the intra-day or reserve power markets, forecasting of the power price is also important.
2. Supply temperature levels and pressure heads. Apart from the control of the units to be switched, the control of supply temperature to the network, as well as the control of the pressure head in the network, can introduce inefficiencies into the DH system. Typically, these controls are static and rule-based. The set point of the supply temperature is, e.g., often determined based on the (moving average of the historic) outdoor temperature and a heating curve. The pressure head is often determined by measurements on critical locations in the network. These rule-based controllers guarantee the heat delivery to the customers; however, they do not optimise towards

economical or ecological parameters. Rather than adopting a rule-based approach, a more dynamic control of the supply temperature and pressure head can help optimise setting temperature levels and controlling pressure heads.

2.1.2. Heat Distribution

The distribution side consists of the piping network, booster pump installations, thermal storage systems and heat transfer stations between primary and secondary networks. All of these can create inefficiencies in a DH network. The distribution side is often a black spot for DH network operators, since only limited data are collected. Furthermore, as long as there are no heat transfer stations, thermal storage or booster pump stations present, as often is the case for small and medium-sized networks, no control points are present in the distribution networks. In that case, optimal control of the distribution side is not relevant. In case one or more of the mentioned installations are present, the control can be optimised in the same way as described in the production side section. Furthermore, as an advanced form of thermal energy storage, the temperature in the supply pipes of the network could be temporarily increased at times, resulting in the use of the network pipes themselves as thermal energy storage.

Analytics is seen as more important in the distribution side. Detection of leakages or insulation faults in the piping system remain a major source of inefficiencies and will gain importance over time. Simulation-based analysis can be used to detect malfunctioning components in the network, whereby measured behaviour is compared to simulated behaviour, and an alarm is raised once the difference exceeds a certain threshold. Since each network has its own topology, it is very complicated to build a simulation model for each individual network, especially for large networks. ML could have great potential in this field, by building automated, black- or grey-box data-based simulation models as an alternative.

2.1.3. Heat Transfer

In the building substations, the flow rate from the DH network is controlled both to power the building's space heating circuit and the domestic hot water production circuit. This is typically done by means of a traditional PI-controller which makes sure that the supply temperatures in those circuits meet their set points. The set points, in turn, are often calculated based on a heating curve together with an outside temperature measurement (space heating circuit) or is a fixed value (domestic hot water circuit). This heat transfer process often results in inefficiencies if not properly managed.

Optimal control in building substations is often referred to as Demand Side Management (DSM). By switching from the rule-based heating curve to a more advanced control, the flow rate (and as a result, also the heating power) to the building can be temporarily reduced or increased. Since buildings contain a lot of thermal mass (concrete, furniture, etc.), this can be done up to a few hours, without thermostats or tenants noticing. This is a powerful tool, since it actually uses the buildings as short-term thermal energy storage. By doing this in a coordinated way between the buildings connected, it is possible to shift the demand profile of a DH network to a desired shape. Since DH networks are demand driven, the production will follow this demand profile.

When it comes to analytics, it should be emphasised that a substation consists of a large number of components (heat exchangers, valves, pumps, control units, etc.), and that those components can always fail, often without the consumer or the DH network operator noticing it. A study [21] claims that three quarter of the substations work sub-optimally. The result of these errors always is a bad cooling of the DH water in the substation, and therefore a higher return temperature back to the network than necessary. Such faults in substations constitute an important source of inefficiencies and this has led to increased interest in fault detection.

2.1.4. Heat Consumption

Behind the substation in the building are the installations for space heating and domestic hot water provision. These are (often like the substations) typically not owned by the DH network operator. However, for the optimal operation of the DH network, they are very important, since they determine the water flow rate, return temperature and power consumption necessary to minimise DH inefficiencies. The ability to influence secondary-side behaviour, e.g., thermostatic radiator valves, thermostat clock schemes and night setback modes, balancing of heating supply circuits, control of Domestic Hot Water (DHW) storage vessel, etc., could result in strategies for reducing inefficiencies.

Additionally, at the consumption side, analytics is important. Indeed, malfunctioning installations on the secondary side of the substations cause high return temperatures to the substations, which are then transferred to the heating network. Therefore, supervision of the building installation itself should be considered to detect unwanted bypasses, malfunctioning thermostatic radiator valves, unbalanced hydraulic circuits, etc. User behaviour is another major source of uncertainties. The ability to make user-friendly visualisation applications, or advanced applications like natural language generation can offer a strategy for minimising such inefficiencies.

3. Method

With the aim of understanding to what extent current state-of-the-art ML has addressed relevant problems in DH, this research article was developed as follows: (1) workshops to create domain insights, (2) surveys and literature review to refine ideas from those workshops and to establish an inventory of existing knowledge, (3) analysis of information acquired and (4) a road-map proposal building on identified knowledge together with expert opinions. Figure 2 provides a summary of the iterative process and how each iteration was validated.

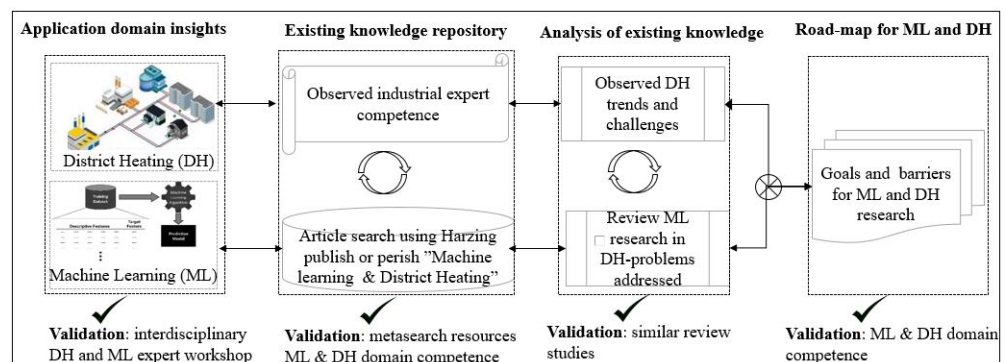


Figure 2. Method outline indicating iterations through which the article was developed and how each iteration was validated. Each iteration involved closed collaboration between DH and ML researchers.

1. Brainstorming in workshops to better understand the problem area. This article employed brainstorming, a technique that is common in the idea generation phase of several scientific endeavours [22]. Through a series of workshops under the umbrella theme of data analytics for DH, discussions between experts from different R&D projects, as well as DH industrial experts, it became clear that even with a common aim, there were discrepancies in the approaches adopted by most ML experts from the expectations of the DH community, e.g., different expectations on the type and quality of data to be collected from DH networks. The interdisciplinary nature of the workshop participants with different competences exposed the different expectations related to data analytics and DH.
2. Inventory of existing knowledge. Problem areas identified in step 1, above, were further expanded through a review of relevant literature in the area, but also through consultative discussions with DH experts. For instance, we draw inspiration from work done in the *Digital road-map for district heating and cooling* (<https://www.euroheat.org>).

[org/wp-content/uploads/2018/05/Digital-Roadmap_final.pdf](https://www.mdpi.com/wp-content/uploads/2018/05/Digital-Roadmap_final.pdf) accessed on 15 May 2021) to identify interesting problem areas in DH. To understand how ML research has developed in the area of DH, a state-of-the-art literature review was completed [23]. To identify the most relevant research, a search was conducted using the phrase “Machine Learning and District Heating” with the help of Harzing’s Publish or Perish software [24].

In line with the clear trend of recent growth in the field, the search was limited to the last 10 years (2010 to 2021) during which most of the research in the area has been reported (see Figure 3).

Initially, 1000 articles were selected. These were pre-screened by reading through the title and in some cases the abstract of each article when the title was unclear. We cross-validated against each others perception to minimise the selection bias. The result was a selection of 179 relevant articles. By relying on a meta-search resource such as Harzing’s Publish or Perish, the most relevant research articles could be indexed from a variety of databases. The search was limited to “Machine Learning and District Heating” because the authors came to the conclusion that no relevant article will exclude such keywords even if some may also include Artificial Intelligence.

3. An analysis of existing knowledge to determine the extent to which various DH relevant problems have been addressed by ML research.

Other literature reviews were analysed to determine whether they provided enough of an overview of how ML has been applied in the field of DH. However, we perceived that the identified review studies left room for focusing on the discrepancies between relevant industrial challenges in DH and the state-of-the-art applications of ML. Still, they were invaluable in shaping our understanding of the current state-of-the-art in ML applications and district heating. Relevant information was extracted from the selected articles in a way that allowed for identification of major themes in terms of the problem area addressed. We make use of relevant review articles together with our own experiences, when determining how different DH problems have been addressed by existing ML research.

4. Proposed road-map for DH and ML. Finally, we have identified several mismatches by comparing the trends in scientific articles, culminating from the above steps, against knowledge from DH experts and our own experiences (see Section 5). Building on those identified (mis)matches, we propose a road-map for ML in DH (see Section 6).

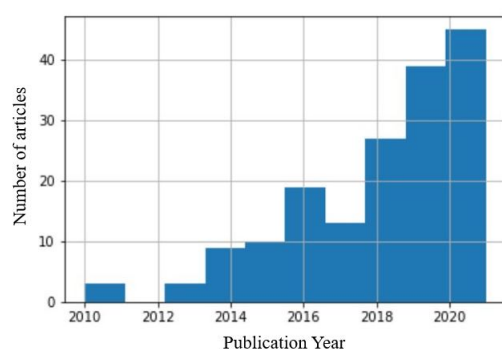


Figure 3. Number of published articles addressing “Machine Learning and District Heating” over the last 10 years (2010–2021).

4. Machine Learning Applied to District Heating

Understanding how advances in ML have addressed the key problems in DH can help to determine any (mis)matches between these two fields, i.e., whether ML research efforts are directed towards the most relevant problem areas in the DH industry. As ML technology continues to advance and the cost of data collection continues to decrease, it is no surprise that the previous decade has witnessed a surge in research efforts aimed at

applying ML in DH. As a consequence, there has been several review studies aimed at understanding the overall landscape of this new research focus.

Slow improvement in DH technology through newer generations promises to result in better data collection opportunities, e.g., high resolution data instead of today's—at best—1 h or 15 min intervals. These opportunities can enable DH systems to satisfy the demands of today's society, e.g., supplying DH at low temperatures while optimising the indoor climate. Improving building energy performance, reducing costs, satisfying regulatory constraints, meeting sustainability goals, etc., are some of the challenges addressed by published ML research [15,25]. A synthesis of how ML has been applied in DH begins in this section with an overview of previous review studies to highlight current knowledge, knowledge gaps and possible future directions. This is followed by an analysis of articles focusing on ML and DH published in the 2010–2021 period.

4.1. Highlights of Existing Literature Reviews

To understand existing literature reviews in this area, we provide an overview of the knowledge gaps identified in various studies, the types of ML models reported, datasets employed and building characteristics (features) studied. Most research articles on the application of ML in DH have occurred in the last 10 years [7,15,18]. Each of the review studies addressed different research gaps, considered research articles over different time periods and considered various building types.

Some choose to cover the full breadth of ML methods, while others limit the scope of the ML models included [7,26]. Knowledge gaps addressed by previous review studies include: (1) lack of an overview of building performance models, in general [7,26], and the lack of data driven approaches for building performance, in particular [15,25], (2) lack of comprehensive studies of ML in urban building energy performance [7], (3) lack of ML performance and accuracy assessment on DH problems [16] and (4) lack of open datasets for benchmarking algorithms [15]. Other review studies highlight the need to compare the forecasting performance of common ML algorithms such as Artificial Neural Networks (ANN), Support Vector Machine (SVM) or Gaussian-Based Regression (GBR) [7,18]. Another gap is the lack of discussion of the key strengths and weaknesses of “white-box” models, “black-box” models, but also hybrid models, with focus on building modelling and energy performance prediction [26].

Many review studies have synthesised and reported on the different types of ML models applied in the energy sector as a whole. Examples include a study of the application of ANN, SVM, GBR and clustering methods when analysing building energy performance as a whole [8]. A similar review [7] came up with the conclusion that ANN, Support Vector Regression (SVR), Genetic Algorithm (GA) and Random Forest (RF) were common methods applied to individual buildings. Reviews have also focused on specific methods, such as deep learning [18] and ANN [17].

To improve prediction accuracy, there is increasing interest in using hybrid methods, e.g., ANN and SVM [8,15]. None of the reviews found a single overall best model although SVM was reported to perform relatively better compared to ANN and regression-based methods, yet the most used method for building energy prediction remains ANN [16].

Review studies have found no evidence to suggest that a given ML model is superior for a specific problem domain, but rather, the suitability of every model is determined by the choice of features and size of data [8]. Feature selection is difficult because the desired features rarely align with available data. For example, most ML forecasting models do not incorporate building space functionality, occupancy and humidity, due to lack of data even though this could be a useful feature [7,16]. As a result of the type of data and models used, predictions are often limited to short term [16,25]. In general, reviews have found it difficult to establish an overview of the dataset employed due to a lack of detailed information about the dataset from various studies. In addition, there is lack of testbeds and automated and streamlined data-processing for building energy prediction [15].

Model accuracy has been a difficult topic in the reviews due to the lack of a common set of criteria through which ML models can be compared in the domain of building energy forecasting [7]. Still, many articles make attempts to validate their findings by comparing performance with other algorithms on similar tasks [15,17].

Following the analysis of the scope of various review articles, we can now state that ML research for DH is dominated by energy demand forecasting [7,8,15,19,25]. Following the analysis of the scope of various articles, we do not know of any review that has looked into the relevance of ML research problems from the perspective of the DH industry. Thus, the current study does not only aim to understand what problem areas have been addressed as in Ntakolia et al. [19], but dares to further make an assessment of the relevance of such problems.

4.2. Review of Recent Research Articles

Among the papers we have reviewed, the vast majority focused on forecasting energy demand. More specifically, among the 179 articles we studied, 137 addressed a concrete DH challenge, most commonly proposing new ML solutions (the rest consisted primarily of surveys and position papers, or papers outside the scope of this discussion). A stunning 98 of them considered forecasting to be this challenge, making it 72% of the field. The second most popular category was anomaly detection, with only 9 papers. In the spirit of focusing on gaps rather than the obvious, we will not discuss forecasting further here, especially given that—as indicated above—several reviews of DH forecasting have already been published. Instead, we will focus on more unique papers addressing the less popular areas, since we believe they deserve increased visibility. Thus, Figure 4 offers a clearer picture that would have been difficult to observe if forecasting studies were included. In Figure 4, the studies have been categorised according to the DH value chain, with the size of each bubble suggesting the extent to which the problem has been addressed for the given category.

The second most popular emerging research topic in DH is anomaly and fault detection. Many of the studies, as shown in Figure 4, utilise primary-side billing data from substations. Different from other tasks, methods for anomaly detection show significant variability depending on the specific cases that they are applied to [27]. While some methods are developed in unsupervised fashion, there are also techniques using supervised Machine Learning that require ground-truth knowledge of the faults. Evaluation of different methods also vary based on their characteristics and the nature of the datasets used.

Abghari et al. [10] proposed a higher order mining approach to identify faulty substations. Their method consists of sequential pattern mining on raw data, clustering analysis and Minimum Spanning Tree (MST) construction on the extracted patterns. The dataset used in this study consists of hourly average measurements from 82 buildings located in Southern Sweden. Another method in [28] utilises clustering and association analysis to identify substations with abnormal operations. Clustering is used to find distinct operating patterns, while association analysis is used to learn substation operating rules. They applied their method on consumption data from the primary side of a DH system located in China. On the other hand, the work in [29] proposes a two-level fault detection and isolation scheme with Convolutional Neural Networks (CNNs). They use simulated data labelled with different fault scenarios to evaluate the performance of their model. Bode et al. [30] also leveraged supervised ML and trained different classifiers with real-world and synthetically generated data containing ground-truth faults. In their analysis, they applied and compared Logistic Regression (LR), k-Nearest-Neighbour (kNN), Classification and Regression Tree (CART), Random Forest (RF), Naive Bayes Classifier (NB), Support Vector Machine (SVM) and Multi-layer Perceptron.

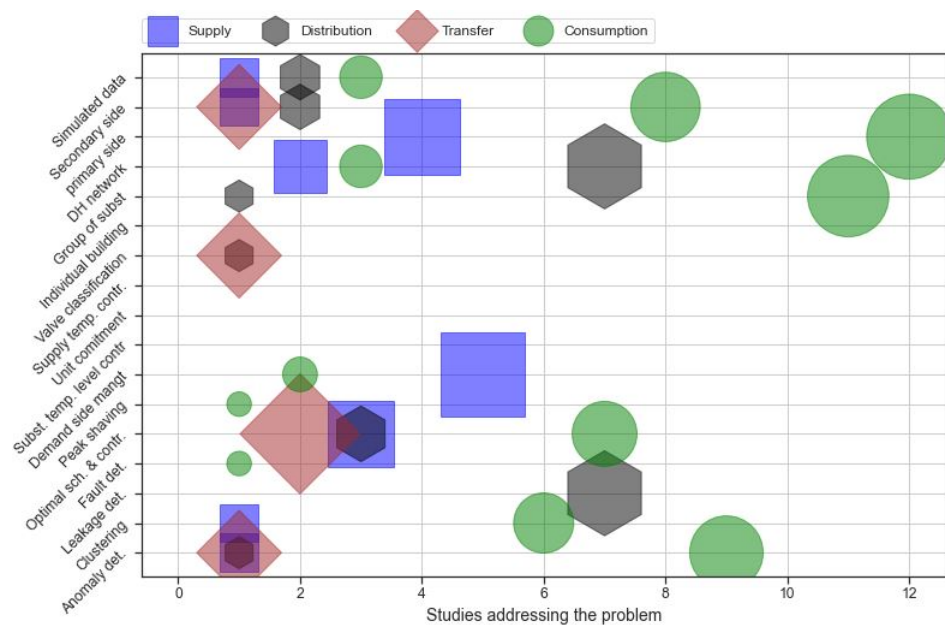


Figure 4. Bubble plot of different problems and ML studies after removing forecasting. White spots in this diagram indicate that we could not find any relevant study for the problem stated on the vertical axis. The size of the bubble (normalised for each category) is proportionate to the number of studies identified in each category of the DH value chain.

Leakage detection can be seen as a specific case of anomaly detection which aims to detect water leakages in DH pipes. The majority of the works addressing this task try to locate leakage faults using infrared images of the areas where the underground pipes are located. These approaches mainly use supervised ML methods trained with these images to classify leakages. For example, ref. [31] applies and compares traditional ML algorithms such as LR, RF and SVM and a deep learning-based CNN model to automatically classify leakage faults on infrared images, captured by an Unmanned Aerial Vehicle (UAV). Similarly, [32] uses RF and [33] uses different linear and non-linear classifiers such as Linear Discriminant Analysis (LDA), linear SVM, RF and AdaBoost. Different from other methods, the work in [34] uses the measurements from flow meters and pressure sensors instead of images to detect leakages. They implement a hydraulic simulation model to simulate all possible leakage faults in DH networks, and train an XGBoost classifier with data including flow and pressure readings and simulated faults.

Having in-depth knowledge of the customers and a better understanding of their heat use is important for effective district heating operation and management. Especially for the consumption side, analysing how customers consume heat helps to implement new control strategies, personalise demand management for specific customer groups and empower consumers to optimise their behaviours. However, properly segmenting customer groups and discovering their typical and atypical consumption patterns is a complex task, especially for DH systems involving many customers with different characteristics. Therefore, data-driven methods using different Machine Learning techniques to automatically cluster various customers and their consumption behaviours attracted many efforts recently.

Calikus et al. [35] proposed an approach to automatically cluster customers and extract heat load patterns. They have represented each customer's consumption behaviour as a weekly aggregation over four different seasons and used k-shape clustering to cluster these weekly consumption profiles. They extract the most common shape in each cluster as well as atypical patterns that do not show similarity with any groups. They have applied their work on data collected from two district heating networks in southern Sweden and included 2200 substations in total. Another work in [36] has done clustering analysis on residential buildings in Denmark. They have used daily heat consumption to represent the behaviour in each building and applied k-means clustering. Their analysis involves

primary side consumption data of 8000 single family houses. Although majority of the works have used the classical and widely used approach k-means, there are some works that have chosen alternative methods such as Partitioning Around Medoids (PAM) [37], Gaussian Mixture Models (GMM) [38] and Nearest-Neighbour clustering [39].

Since clustering is an unsupervised Machine Learning task, it is often difficult to properly evaluate whether the clusters found are actually meaningful. Most of works use internal clustering evaluation measures such as Silhouette Coefficient [35,40], Bayesian Information Criterion (BIC) [36,38], Davies–Bouldin Index [40], Dunn Index [37] and so on to evaluate their approach and also to determine the appropriate number of clusters. Other works involve visual analysis of computed clusters to demonstrate the cluster quality and interpret the characteristics of the discovered consumption patterns. For example, [35] incorporates a visual inspection step, in which a domain expert assigns corresponding control strategy based on the characteristics of each cluster.

There are also a few even more unique papers, addressing other tasks in DH using Machine Learning. For example, De Somer et al. [41] proposed a reinforcement learning-based approach for optimal scheduling of the heating cycles of the domestic hot water. In another work [42], a prediction model of secondary supply temperature is built using indoor temperature and building thermal inertia to achieve more refined control. On the other hand, Potočnik et al. [43] focuses on assessing the quality and condition of valves installed in district heating systems. A method for classification of valve sounds is proposed, based on acoustic features and Machine Learning models.

5. Gaps between Industrial District Heating Needs and Current Machine Learning Research

This section describes the (mis)match between current ML research and key challenges faced by the DH industry. It builds to a large extent on the academic and industrial experience of the authors, and is consequently somewhat subjective. Regardless, we believe that presenting this perspective is going to be valuable in making ML applications more effective in DH.

By far, the most common application of ML to DH is energy consumption forecasting [7,8,15,19,25]. This can be used for planning production, but without additional capabilities such as demand side control, the usefulness of detailed consumption forecasting is limited. Other, arguably more important, DH challenges are largely ignored. A possible explanation for this discrepancy is the availability of primary side energy, volume and sometimes temperature data, already collected by the DH operator for billing and monitoring purposes. Consequently, research has focused on what can be done with the readily available data. The situation is likely to remain until standardised IoT solutions become more prevalent.

5.1. Benefits of Demand Forecasting

Accurate forecasting of the demand, however, only brings value to the extent that it can be used to improve (or optimise) network operations. The most commonly cited idea is reducing resource consumption through modulated production, based on predicted demand. Given that most networks tend to use a combination of different energy sources and fuels, from intermittent energy sources over CHPs, heat pumps to coal and oil boilers, forecasting has the potential to allow greater production flexibility. Renewable energy sources can be used whenever available, and fossil fuels only as the last resort. However, there still needs to be sufficient safety margins to guarantee that the heat is available on demand. In this context, there are clear diminishing returns related to precision of the forecasting, while imposing critical constraints on the robustness (i.e., significant under-estimations of the demand can render any given solution completely infeasible). These practical considerations are seldom addressed in the literature of ML as applied to the field of DH.

The long lead times from production to the heat being available at a given point in the heat network, which can sometimes be several hours, pose another constraint for planning. District heating is, in this regard, quite different from many other domains, even

ones superficially very similar, like electric grids. It has been shown that such long-term effects can be difficult to estimate purely from the primary side data. Instead, a promising direction is to investigate and perform peak shaving by modulating demand and reduce the use of more expensive or non-renewable fuels. The customer side demand could be explicitly lowered in periods where no cheap and sustainable energy is available, and increased when there is an abundance of it. This, in turn, requires integrating with the corresponding heating systems as well as indoor temperature sensors to guard against control actions jeopardising the indoor climate. While there are a number of commercial solutions available, the heterogeneous nature of buildings, and the range of possible objective functions, makes this difficult.

Finally, on the even more long-term perspective, accurate forecasting can provide a significant benefit for infrastructure planning. It amounts to using better-understood, and hence “narrower” safety margins. However, in the papers we have reviewed, there is no indication of any ML-driven analysis targeting these aspects.

5.2. Cost–Benefit Analysis

The disconnection between the ML demand forecasting research and the business of running a heat network is clearly seen in the choice of metrics used by forecasting papers. Almost all used metrics (the most common being Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE)) are symmetric, i.e., they penalise over-estimations and under-estimations equally. However, from a DH perspective, there is a considerable difference between the two errors. Over-estimation of the demand leads to some amount of energy waste, while under-estimation makes the district heating network fail at its primary function, i.e., to provide sufficient heat to the customers. The latter situation is unacceptable and the forecasting must take this into account. To be able to trim the margins of production, one would need to understand the statistical distribution of demand and how it evolves over time, with some degree of confidence. Hence, there is room for research on facilitating a cost–benefit analysis of provisioning the network, the suitable choice of metrics and forecast methods taking the refined metrics into account. In the review, we have found no study comprehensively addressing this issue. Ultimately, the goal is to minimise the operation costs of the network as a whole, while ensuring heat delivery to all the customers according to the contracts. Clearly, while accurate demand forecasting is an important piece in this puzzle, these potential benefits will not be realised without using the forecasts to *control* the network. In our literature review, few papers have made this connection. From that perspective, the forecasting itself is severely over-researched, often without the exact purpose mentioned, or with it being expressed only vaguely. At the same time, many other important aspects are ignored. At a high level, a DH network control algorithm is supposed to, for a specified time horizon, decide on the optimal values for all relevant parameters. The most important of those are the operation for different heat sources (which ones to start, and which to shutdown). However, it also includes supply water temperature and pressure, as well as possible differential pressures for pumping stations. This guarantees the timely arrival of a sufficient amount of heat to every customer, with reasonable safety margins, both in terms of energy and time. Automatically controlling the complete network would require dedicated hardware on-site, or excellent footing with the company supplying the Building Management Systems (BMS) or substation controllers.

Today, doing it at scale is still expensive and complicated, but many different actors in the market, which include the BMS companies, are feeling the pressure to stay relevant and thus might become more approachable. For example, an optimal bottom-up control of an entire DH grid from many independent BMSs is an interesting long-term goal worth pursuing.

While it is relatively easy for a research group to negotiate with a heat provider for access to (anonymised) primary side data, gaining access to a large number of indoor temperature readings often requires a considerable effort in the form of negotiating with building owners, mounting adequate sensors and actuators throughout the buildings, connecting everything

to the internet, managing privacy consideration, etc. This will hopefully improve with more streamlined IoT solutions, but today, it still presents a serious obstacle.

Another aspect is the challenge to decide on an objective function for model predictive control. In simulations, this is often simplified into keeping the indoor temperature at, for example, 21 ± 1 °C, but the indoor temperatures can vary much more depending on time of year, time of day, weather and social behaviours. While there are standards providing guidance on thermal comfort assessment, such as EN ISO 7730 or ANSI/ASHRAE Standard 55, it is unclear how to value indoor climate changing over time, and more research is needed. At the same time, these difficulties are not a sufficient reason to limit the research progress to the degree we see today. It is worth noting that many larger building owners have already been involved in less streamlined research efforts, and can be reluctant to participate in yet another experiment. ML researchers must substantially address the business case upfront, rather than as an afterthought. In a sense, the community has used up a lot of “initial trust”, and needs to make up for that in the future.

5.3. Practical Considerations

The key practical consideration is that the primary side measurements for substations are used for billing purposes and thus tend to be available for research. Any other information, from failure histories to network measurement data and network topology, is harder to get a hold of. The access to certain data will likely improve with ongoing digitisation efforts, but the topology (at least in Sweden) is classified as part of the critical infrastructure, which effectively prevents it from being freely shared for research purposes. This opens up for ML research into reconstructing the topology to an adequate degree, at least sufficiently in determining the time lags throughout the network, without compromising the safety and security. It could be, for example, based on primary side temperatures and measurements from the heat plant, using variations in temperature to chart the network, etc.

Interesting work is recently appearing in the context of using ML for condition monitoring and predictive maintenance of different parts of network infrastructure [19,44,45]. Continuously monitoring the operation of complex systems, such as a DH network, is a suitable task for ML as models can typically be trained on acceptable system characteristics. Predicting faults and optimising productivity can be done as long as suitable methods work without extensive human supervision: without requiring the designer to think of all possible faults beforehand; able to do the best possible with limited data, without the need for dedicated new sensors; scaling up to “one more system and component” and multiple variants; and finally, adapting to changes over time and remaining relevant throughout the lifetime of the system. However, development and especially evaluation of predictive maintenance and fault detection is severely hindered by the lack of ground-truth information about the failures and repairs.

These practical obstacles explain the gaps in research quite convincingly, however, ultimately they run the risk of locking the development in a “local minimum”. The market hardly looks further than the next quarter. The competition in the DH industry is not particularly fierce, and therefore limited pressure is there for companies to make high-risk, high-benefit investment—instead, a stable and conservative approach is common. Thus, it is important that solutions provide immediate benefits in the current market, or they will likely be ignored. However, in the long run, such a conservative approach could seriously hinder achievement of environmental sustainability goals.

Finally, there is of course a number of AI and ML applications that could be used in the DH, but are in no way specific to it—for example, marketing, customer support and many others. We have not addressed them here at all, since the scope of this paper is to focus on specific needs of the DH sector.

5.4. Control and Optimisation

There are convincing indicators (following research and practice) that the single most useful adaptation of current ML research would be to consider the control and modelling

problems together. One of the main reasons is because control strategies open the doors to supervised ML and enable interpretation of results for unsupervised ML, e.g., anomaly detection. It brings certain challenges, of course, not the least of which is the lack of data related to the network control (e.g., network structure and current parameters) and also optimisation criteria. It could be interesting to consider the comparison between model predictive control systems (that can be built on top of the proposed ML forecasting methods) against existing classical control systems.

The benefits of linear versus non-linear or deterministic versus stochastic models should be evaluated, as well as the robustness of different solutions.

The understanding of a trade-off between building a simplified model and solving the optimisation exactly, versus building a more complex (but also more accurate) model and only obtaining approximate solution, is lacking.

The next step, then, is the optimising operation of the entire DH network on a global scale, where the complexity and volatility clearly necessitates employing ML-based tools. This requires, first, a reliable model of the network, including the pipes, the substations, customers, heat exchanging equipment, etc. One can differentiate between networks with constant supply temperature throughout the year, and those where supply temperature changes throughout the year. Depending on the existing customers and the equipment base, one might compare variable flow (with constant temperature) to a constant flow (with variable temperature) and find that one is better suited than the other, given specific local circumstances. Generally, there is a need to build a network model that is a simplified version of reality. Realistically, there is no benefit of modelling every pipe and every substation. One needs to identify “critical objects” spread throughout the network. Assuming an existing network and including demand-side management already leads to quite complex optimisation. One can next, of course, also imagine optimising the planning, development and building of the DH network itself.

Another aspect, seldom addressed in the literature but which constitutes a challenge in the industry, is how to configure and manage a large heterogeneous system. In the framework of model predictive control (MPC), it is certainly possible to build one model per building, however, that comes with each one having their own choice of parameters for how to solve the corresponding optimisation problem outside the actual mathematical formulation, e.g., regularisation, scale factors, error tolerance, number of iterations, etc. Making these kinds of adaptations at scale is challenging. A well-informed and interested engineer can make a reasonable choice for any given substation; however, it is not possible for the whole network consisting of thousands of substations, most with only somewhat reliable data.

As with any other industrial process, the efficiency of the heat generation process itself is of interest. However, we expect that any such improvements will have to be tailored to the particulars of the energy plant in question and are better addressed as such, rather than in the framework of heating networks, which is about the optimal use of the heat.

6. A Road-Map for Machine Learning and District Heating

Applications of ML in DH will create more value if researchers gain broad knowledge about different limitations in the DH industry. DH will become effective if the right incentives are designed to influence the behaviour of the customer, offering a better indoor climate and utilising low carbon energy sources.

We believe that to achieve the milestone with cost-effective development and application of ML in DH, there is a need for a road-map that identifies important goals necessary to increase the impact of ML solutions in DH, as DH systems become increasingly digitalised.

We identify goals that are largely dependent on developments in society at large, but specifically match the patterns of development in the DH industry. These goals emerge from the context of operations of DH (Section 2), current state-of-the-art research in ML and DH (Section 4) and gaps between these two areas (Section 5). We subsequently discuss

important barriers that hinder progress on the way towards achieving these goals. Finally, suggestions are given on how these barriers can be addressed.

6.1. Goals for Machine Learning in District Heating

The extent to which the following goals are achieved will determine how effective ML is, as an enabling technology, for adapting DH to meet the future challenges.

1. Scalable analytic solutions for large networks with tens of thousands of customers.
2. Tools for modelling building behaviour, and evaluating the quality of such models.
3. Forecasting integrated with optimisation of the DH network components.
4. Incorporating network topology into ML models.
5. Standardised protocols for secondary-side data collection.
6. Commonly accepted evaluation metrics for indoor climate comfort.
7. Providing customers with tools and incentives to control energy use for peak shaving.
8. Development of shared data portals and benchmark datasets with ground-truth data clearly distinguishing normal operation from data with faults.
9. Around the clock monitoring of the conditions of the DH network combined with predictive maintenance.
10. Utilising ML to demonstrate the potential of low temperature DH.

6.2. Barriers in the Application of Machine Learning and District Heating

With increasing digitalisation, decarbonisation, advances in sensor technology, low cost processing and storage in the cloud and advances in ML algorithms, we believe that the stated goals can be achieved. However, a number of specific barriers need to be taken into consideration, which also explain why these goals have not been achieved yet. We have divided the barriers into groups of technical, business and organisational barriers.

Technical barriers. The primary technical challenge is the data availability. Successful research in ML depends on having sufficient “good” data, i.e., relevant, high quality and at a sufficient resolution of the purpose it is used for. Access to such good data, however, depends to a large degree on a value stream, from: sensor and measurements, data collection and communication, data storage and structure, data analytics, presentation of the analysis to the user and finally the value of the new knowledge to the user. At present day, the ideal value streams hardly exist. It will take years of systematic iterations of improvements to reach “good” data. However, ML researchers need to work on demonstrating the benefits that these efforts will ultimately provide—even if this is a long-term strategy. Until this initial work is done, no serious efforts on collecting more comprehensive data will begin in the industry. ML research needs to provide input on the specific needs, and influence the type of data collected by the DH industry.

In addition, it is not enough that there are data, but this data must be well-understood and categorised comprehensively. For example, knowledge about what data can be considered to be fault free, versus data that contain operations with faults, is highly necessary.

Another example is the ownership and access to the data. This can be exemplified with the “closed substation” barrier. We need substations with open communication protocols, so that they can be controlled from the outside with the help of third party solutions. This is necessary to facilitate the district level ML solutions that can be incorporated to each substation.

Business barriers. The key in this category are barriers for realising customer incentives. Business models need to create opportunities for tools to incentivise customers to positively impact their energy use behaviour. In general, purely economical and monetary factors are not enough to achieve that, and other solutions are needed, e.g., through visualisation. In addition, energy tariffs rewarding consumers for reducing their return temperature or to offer the flexibility of their building mass for better control can be implemented.

The second barrier concerns advancements in the DH industry. There are many obstacles to innovation and change of DH companies and property owners. These indus-

tries traditionally focused on exploitation (stable conditions and weak competition) and thus there is little incentive for exploration, to radically develop and improve.

Organisational barriers. The most important barriers arise from the DH ownership structure. The distributed aspect makes it difficult to create solutions based on “big picture” optimisations. For example, it would be beneficial if network operators own the building substations, so that they can control them directly. Similarly, insufficient customer awareness leads to sub-optimal decisions and configurations. Increasing the awareness of consumers can give them insight into their energy use, for example through gamification or similar tools.

Other barriers are due to changing the relation between regulations and fair competition. As previously described, we do see external influences that will change the stable conditions and the mode of exploitation; decarbonisation (change of fuels), electrification (more competition), deregulation (no monopolies). This can drive changed ownership, most likely leading to less public ownership across different parts of the DH industry.

The final barrier is lack of cross-disciplinary cooperation. The new paradigm of AI/ML will require new organisational interactions. Successful research will engage teams of data scientists, domain specific scientists (material, energy, control, etc.) and maybe even organisational experts (innovation, business process improvements). The research groups must team up with representatives from tech suppliers, from DH companies and from the building property industry.

6.3. Suggestions for Different Stakeholders

The current situation affects different stakeholders in different ways, and gives them different opportunities for advancement. In this section, we present selected suggestions for each of the five stakeholder groups: DH companies, DH customers (primarily building owners), the research community, technology suppliers and consultants and policy makers at local, national and international levels.

In terms of “priority of problems” from the industry, it is important to notice that there is not one “industry” to begin with. What is important to research or develop depends on whether one is selling to, or collaborating with, the energy companies (production or distribution), maintenance firms, building owners, etc. Additionally, DH networks are not operating in a vacuum. Other solutions, such as heat pumps and ventilation systems, must be taken into account. Much of the DH industry is still somewhat catching up with the introduction of heat pumps, which has, in some cases, benefited from using the heat pump within its optimal range of operation, and only tapping into district heating during peak demand periods when the heat pump capacity does not suffice any more. That mode of operation is very clearly economically infeasible for DH operators. How to regulate these different solutions is an open question, and it is handled very differently in different countries. Condition monitoring and predictive maintenance will be gaining prominence, especially techniques that do not look at one substation in isolation, but can analyse the complete network [46]. It is crucially important in particular for low temperature networks, which require correctly behaving substations to achieve the desired effect. To a large extent, it is still an open area for research.

Innovations and improvements would also benefit from a more openness among DH companies to communicate their challenges and problems. The DH companies and their customers (property owners) need to better understand the possibilities that come with digitalisation and machine learning, the opportunities for increasing efficiency and generating new revenues (digital service design). If DH companies and the building industry communicate their challenges better, and the research community understands this context better, the impact of the research will increase.

On the other hand, both the DH companies and their primary customers must put some effort into better understanding the possibilities that come with new technologies (digitalisation and machine learning). The opportunities for increasing efficiency and generating new revenues (digital service design) are vast, but ultimately, nobody else

can do this work. To survive long-term, they must become aware of the problems and challenges that will increase in importance in the future. An example direction is working on better alignment of usefulness and value to the user—there is a lot of talk in DH companies about customer engagement and dialog; however, very few concrete ideas for useful services have been proposed in the literature or implemented in practice.

To understand the real problems and to develop solutions, the researchers must form cross-disciplinary teams. There is also a paradigm shift in research methods. The traditional researcher (model development) in energy systems as well as in data science could perform good research from a desktop, on his own. The new research in the era of digitalisation and ML must be aligned with real data which consequently requires deep interaction with many actors, and with technology suppliers with the required knowledge and capabilities.

Policies, such as taxes, incentives and regulations on local, national and EU levels also have a potential to be smarter, focusing more on effectiveness, rather than blunt focus on energy volumes without considering time and place.

7. Discussion and Conclusions

To remain competitive and relevant, heat network technologies need to progress towards lower temperatures and greater use of renewable energy sources. The latter encompasses the reuse of low temperature waste heat and Combined Heat and Power as well as heat networks serving as a buffer for offloading electric networks integrating intermittent renewable energy sources, such as wind and solar electric systems. The overall system is in need of solutions for planning, monitoring, optimisation and control, and ML is likely part of the answer.

Efforts are being made by academia, leading niche tech suppliers, leading European heat network operators and IT giants, the latter three by their very nature with better access to data. This poses a challenge for academic ML, which needs to keep up to date with industrial advances while competing with publications backed by IT giants.

The aim of this article was to chart and discuss to what extent state-of-the-art ML has addressed key challenges in the DH industry, to identify discrepancies between the two fields and to propose a road-map with suitable goals for how to increase the impact of ML in DH. The shift towards more efficient low temperature networks stands to transform the industry, but the large costs of infrastructure, to be paid upfront but discounted over the several decades, constrain the possible solutions. Not only must the system account for continued urbanisation, but it must also remain cost-effective when subjected to increasingly stringent requirements with respect to decarbonisation and electrification. This opens up for rethinking the way DH systems are designed and operated, for example, the use of distributed heat supplies to address urban growth.

The ongoing digitisation of the DH industry, in combination with today's increasingly affordable and capable IoT solutions and cloud-based computational resources, seems ready for disruption: to use AI, in general, and ML, in particular, to address the challenges faced by DH, and for the creation of new businesses structured around the same. This explains part of the extreme growth in the number of research publications focusing on applying ML techniques to DH. However, based on our experience in the field, and review of the state-of-the-art, we perceived a mismatch between the most popular research directions, such as different forms of forecasting, and the challenges faced by DH. Consequently, we suggested a number of goals, ranging from open source benchmark datasets to methods for empowering consumers, to increase the impact of ML in DH. The goals involve both technical and non-technical barriers to be overcome. We hope they will serve as a source of inspiration.

Author Contributions: Conceptualization, G.M., H.K., S.N.; methodology and validation, D.V., E.C., G.M., H.K., J.B., S.N.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: The work presented in this article was financed by the Swedish Knowledge Foundation (KK-stiftelsen <http://www.kks.se/om-oss/in-english/> accessed on 15 May 2021) under grant Dnr.20170182 within the project Data Analytics for Fault Detection in District Heating (DAD) and grant Dnr.20160301 within the project Self-Monitoring for Innovation (SeMI). The Swedish agency for innovation, VINNOVA (<https://www.vinnova.se/en> accessed on 15 May 2021), financed the work in this article through the SAM project, Dnr. 2018-03349.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bawany, N.Z.; Shamsi, J.A. Smart city architecture: Vision and challenges. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*. [CrossRef]
- Bertoldi, P. Overview of the European Union policies to promote more sustainable behaviours in energy end-users. In *Energy and Behaviour*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 451–477.
- Jones, D.A.; Eckert, C.; Garthwaite, P. Managing Margins: Overdesign in Hospital Building Services. In *Proceedings of the Design Society: DESIGN Conference*; Cambridge University Press: Cambridge, UK, 2020; Volume 1, pp. 215–224.
- Buffa, S.; Cozzini, M.; D’Antoni, M.; Barattieri, M.; Fedrizzi, R. 5th generation district heating and cooling systems: A review of existing cases in Europe. *Renew. Sustain. Energy Rev.* **2019**, *104*, 504–522. [CrossRef]
- Lasi, H.; Fettke, P.; Kemper, H.G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
- Saloux, E.; Candanedo, J.A. Forecasting district heating demand using machine learning algorithms. *Energy Procedia* **2018**, *149*, 59–68. [CrossRef]
- Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S. Machine learning applications in urban building energy performance forecasting: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *133*, 110287. [CrossRef]
- Seyedzadeh, S.; Rahimian, F.P.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 1–20. [CrossRef]
- Reynolds, J. Real-Time and Semantic Energy Management Across Buildings in a District Configuration. Ph.D. Thesis, Cardiff University, Wales, UK, 2019.
- Abghari, S.; Boeva, V.; Brage, J.; Johansson, C.; Grahn, H.; Lavesson, N. Higher order mining for monitoring district heating substations. In *Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Washington, DC, USA, 5–8 October 2019; pp. 382–391.
- Frederiksen S.; Werner, S. *District Heating and Cooling*; Student Literature; Interak: Czarnków, Poland, 2013.
- Lund, H.; Østergaard, P.A.; Chang, M.; Werner, S.; Svendsen, S.; Sorknæs, P.; Thorsen, J.E.; Hvelplund, F.; Mortensen, B.O.G.; Mathiesen, B.V.; et al. The status of 4th generation district heating: Research and results. *Energy* **2018**, *164*, 147–159. [CrossRef]
- An, E. *Strategy on Heating and Cooling*; European Commission COM: Brussels, Belgium, 2016.
- EU Commission. *Clean Energy the European Green Deal*; European Commission COM: Brussels, Belgium, 2019.
- Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *285*, 116452. [CrossRef]
- Bashir, M.B.; Alotaibi, A.A. Smart buildings Cooling and Heating Load Forecasting Models. *IJCSNS* **2020**, *20*, 79.
- Runge, J.; Zmeureanu, R. Forecasting energy use in buildings using artificial neural networks: A review. *Energies* **2019**, *12*, 3254. [CrossRef]
- Runge, J.; Zmeureanu, R. A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings. *Energies* **2021**, *14*, 608. [CrossRef]
- Ntakolia, C.; Anagnostis, A.; Moustakidis, S.; Karcanias, N. Machine learning applied on the district heating and cooling sector: A review. *Energy Syst.* **2021**. [CrossRef]
- Knutsson, H.; Holmén, M.; Lygnerud, K. Is Innovation Redesigning District Heating? A Systematic Literature Review. *Designs* **2021**, *5*, 7. [CrossRef]
- Gadd, H. To Analyse Measurements Is to Know! Ph.D. Thesis, Lund University, Lund, Sweden, 2014.
- Isaksen, S.G. *A review of Brainstorming Research: Six Critical Issues for Inquiry*; Creative Research Unit, Creative Problem Solving Group-Buffalo Buffalo: New York, NY, USA, 1998.
- Grant, M.J.; Booth, A. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* **2009**, *26*, 91–108. [CrossRef]
- Harzing, A.W. Publish or Perish. 2007. Available online: <https://harzing.com/resources/publish-or-perish> (accessed on 25 February 2021).
- Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]

26. Idowu, S.; Saguna, S.; Åhlund, C.; Schelén, O. Forecasting heat load for smart district heating systems: A machine learning approach. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014; pp. 554–559.
27. Calikus, E.; Nowaczyk, S.; Sant’Anna, A.; Dikmen, O. No free lunch but a cheaper supper: A general framework for streaming anomaly detection. *Expert Syst. Appl.* **2020**, *155*, 113453. [[CrossRef](#)]
28. Xue, P.; Zhou, Z.; Fang, X.; Chen, X.; Liu, L.; Liu, Y.; Liu, J. Fault detection and operation optimization in district heating substations based on data mining techniques. *Appl. Energy* **2017**, *205*, 926–940. [[CrossRef](#)]
29. Li, M.; Deng, W.; Xiahou, K.; Ji, T.; Wu, Q. A data-driven method for fault detection and isolation of the integrated energy-based district heating system. *IEEE Access* **2020**, *8*, 23787–23801. [[CrossRef](#)]
30. 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. [[CrossRef](#)]
31. Hossain, K.; Villebro, F.; Forchhammer, S. UAV image analysis for leakage detection in district heating systems using machine learning. *Pattern Recognit. Lett.* **2020**, *140*, 158–164. [[CrossRef](#)]
32. Berg, A.; Ahlberg, J. Classification and temporal analysis of district heating leakages in thermal images. In Proceedings of the 14th International Symposium on District Heating and Cooling, Stockholm, Sweden, 7–9 September 2014.
33. Berg, A.; Ahlberg, J. Classifying district heating network leakages in aerial thermal imagery. In Proceedings of the Swedish Symposium on Image Analysis, Luleå, Sweden, 11–12 March 2014.
34. Xue, P.; Jiang, Y.; Zhou, Z.; Chen, X.; Fang, X.; Liu, J. Machine learning-based leakage fault detection for district heating networks. *Energy Build.* **2020**, *223*, 110161. [[CrossRef](#)]
35. Calikus, E.; Nowaczyk, S.; Sant’Anna, A.; Gadd, H.; Werner, S. A data-driven approach for discovering heat load patterns in district heating. *Appl. Energy* **2019**, *252*, 113409. [[CrossRef](#)]
36. Gianniou, P.; Liu, X.; Heller, A.; Nielsen, P.S.; Rode, C. Clustering-based analysis for residential district heating data. *Energy Convers. Manag.* **2018**, *165*, 840–850. [[CrossRef](#)]
37. Ma, Z.; Yan, R.; Nord, N. A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. *Energy* **2017**, *134*, 90–102. [[CrossRef](#)]
38. Wang, C.; Du, Y.; Li, H.; Wallin, F.; Min, G. New methods for clustering district heating users based on consumption patterns. *Appl. Energy* **2019**, *251*, 113373. [[CrossRef](#)]
39. Kiluk, S. Diagnostic information system dynamics in the evaluation of machine learning algorithms for the supervision of energy efficiency of district heating-supplied buildings. *Energy Convers. Manag.* **2017**, *150*, 904–913. [[CrossRef](#)]
40. Tureczek, A.M.; Nielsen, P.S.; Madsen, H.; Brun, A. Clustering district heat exchange stations using smart meter consumption data. *Energy Build.* **2019**, *182*, 144–158. [[CrossRef](#)]
41. De Somer, O.; Soares, A.; Vanthournout, K.; Spiessens, F.; Kuijpers, T.; Vossen, K. Using reinforcement learning for demand response of domestic hot water buffers: A real-life demonstration. In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Turin, Italy, 26–29 September 2017; pp. 1–7.
42. Yuan, J.; Wang, C.; Zhou, Z. Study on refined control and prediction model of district heating station based on support vector machine. *Energy* **2019**, *189*, 116193. [[CrossRef](#)]
43. Potočník, P.; Olmos, B.; Vodopivec, L.; Susič, E.; Govekar, E. Condition classification of heating systems valves based on acoustic features and machine learning. *Appl. Acoust.* **2021**, *174*, 107736. [[CrossRef](#)]
44. Buffa, S.; Fouladfar, M.H.; Franchini, G.; Lozano Gabarre, I.; Andrés Chicote, M. Advanced Control and Fault Detection Strategies for District Heating and Cooling Systems—A Review. *Appl. Sci.* **2021**, *11*, 455. [[CrossRef](#)]
45. Nowaczyk, S.; Rögnvaldsson, T.; Fan, Y.; Calikus, E. *Handbook of Smart Cities*; Chapter Towards Autonomous Knowledge Creation from Big Data in Smart Cities; Springer: Berlin/Heidelberg, 2020; pp. 1–35. [[CrossRef](#)]
46. Rögnvaldsson, T.; Nowaczyk, S.; Byttner, S.; Prytz, R.; Svensson, M. Self-monitoring for maintenance of vehicle fleets. *Data Min. Knowl. Discov.* **2018**, *32*, 344–384. [[CrossRef](#)]