EuroSun2026

BENCHMARKING MACHINE LEARNING METHODS FOR THERMAL LOAD FORECASTING IN DISTRICT HEATING NETWORKS: REGIME-DEPENDENT RELEVANCE OF AMBIENT TEMPERATURE FEATURES

Summary

District heating networks are becoming more complex, increasing the need for advanced control strategies and short-term thermal load forecasting. This study benchmarks decision-tree ensembles (XGBoost, LightGBM) against deep-learning models (LSTM-encoder-decoder, CNN) on two European networks, evaluating performance with and without ambient temperature forecast as a feature across heating and summer regimes. Results show that all models benefit from temperature information in winter, achieving lower RMSE and MAPE. However, their accuracy deteriorates in summer, when hot water demand is largely independent of ambient temperature. These findings highlight the importance of season-aware feature selection and suggest the use of adaptive modeling approaches to ensure robust performance throughout the year.

Keywords: Load prediction, District heating, Machine Learning

Field of application for the findings: Advanced control of district heating networks

Addressed audience: Utilities, Operators, Planners

1. Motivation

District heating networks are becoming increasingly complex to operate as they incorporate renewable and waste-heat sources as well as sector-coupling technologies such as CHP units and heat pumps. These dynamics make the use of advanced control strategies increasingly important. Most strategies like model-predictive control, demand-side management and fault detection require accurate short-term thermal load forecasts. Many machine learning algorithms use historical consumption and ambient temperature data for this purpose. (Geysen 2018, Pressa 2023, Runge 2023) While the heating demand in winter strongly correlates with ambient temperature, the hot water demand in summer is mostly independent of it. For this reason, the dependency strongly correlates by season. However, in district heating networks with various buildings and control regimes it is impossible to indicate an exact moment in time, when the heating season starts. It is unclear, how well different types of machine learning models handle this discrepancy between ambient temperature dependent heating demand and independent summer demand.

2. Objective

Benchmarking of models across different datasets and with focus on inclusion of ambient temperature features and seasonal evaluations supports faster and more goal-oriented model selection. This paper evaluates whether decision tree-based models or deep learning approaches more effectively predict thermal loads in district heating networks characterized by both space heating and domestic hot water demand for residential buildings. The prediction quality is assessed in dependency of the ambient temperature as exogenous variable as well as for different seasons. The general validity of the results shall be achieved by evaluating multiple data sets with varying weather conditions. The detailed feature engineering varies from method to method and is introduced in the next chapter.

3. Method

In this study, several machine learning and deep learning models for thermal load forecasting in a district heating network are benchmarked. The models considered include Long Short-Term Memory (LSTM) encoder-decoder networks with and without attention mechanisms, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM) and Convolutional Neural Networks (CNNs).

XGBoost is a gradient boosting decision tree algorithm commonly used for tabular data prediction. It constructs ensembles of regression trees sequentially, where each new tree reduces the residual error of the previous trees. The model uses historical load and temperature features, as well as lagged values, to predict future thermal load. XGBoost implements a level-wise tree growth strategy, while LightGBM employs a leaf-wise approach, often achieving higher accuracy and computational efficiency on large datasets. LightGBM uses explicitly engineered tabular features, including lagged load values, rolling statistical aggregates, weather history and forecasts, and calendar indicators. Feature selection is performed via TreeSHAP importance analysis combined with seasonal evaluation, identifying the most relevant features separately for heating and summer regimes.

The LSTM encoder-decoder models are designed to handle sequential time series data, capturing temporal dependencies in the thermal load. An attention mechanism is additionally implemented in one variant, allowing the decoder to weigh past time steps differently, potentially improving performance on longer sequences with varying dynamics, different time scales and different exogeneous corvariates.

In addition, CNNs are evaluated for their ability to capture local temporal patterns in the load sequences. By applying convolutional filters along the time dimension, the network identifies recurring short-term trends and seasonality effects. CNN takes the same physical variables as multi-channel sequential input tensors over a lookback window. Since the LightGBM feature selection already identifies which physical variables carry predictive value across regimes, the CNN channel set can be restricted to those pre-validated variables to avoid irrelevant channels that would otherwise cause overfitting without adding a separate neural-network-specific selection step.

All models are trained and evaluated on the same datasets, which include historical thermal load measurements and ambient temperature. Performance is assessed using metrics such as Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) over a forecasting horizon of 24 hours. Given the strong seasonal asymmetry between heating-driven winter demand and domestic-hot-water-dominated summer demand, metrics additionally are reported separately for heating and summer regimes. Hyperparameters for all models are optimized to further improve the results. Additionally, the computational time needed for training and validation is contrasted.

This comparative approach allows the investigation of how classical tree-based models, lightweight boosted ensembles, and deep learning architectures perform in predicting thermal loads in a district heating network characterized by both heating and domestic hot water demand for various seasons.

4. Case Studies

The evaluation considers two case studies. The first is the district heating network in Malmö, Sweden, focusing on the aggregated thermal load of approximately 80 residential consumers within the broader network. The second case study focuses on the district heating network in Weil am Rhein, Germany. For this analysis, around 60 consumers are considered. For both datasets industrial customers and any consumers who were not part of the network at the start of the evaluation period were excluded, ensuring a more accurate comparison with a similar database. The available measurement data is preprocessed to filter for gaps and outliers and resampled to hourly values.

5. Results

The models have been developed in the course of the last years and have been applied to data of one of the networks. Feature engineering and optimization have been performed individually to reach the best results. For this paper all the models will be applied for both networks with the same sets of data for training, validation and testing. The results for RSME, MAPE and seasonal evaluation of the metrics as well as computation time for training and validation will be compared for both models for a full year. Figure 1 and 2 show the effect of a LSTM model that includes the ambient temperature. While in October there is a significant improvement, the impact is less visible in July. This observation is in line with other evaluations that show that all models show a significant improvement when including the ambient temperature in winter, while they even tend to get worse in summer. Initial results indicate that decision tree-based models and deep learning architectures may respond differently to temperature features in summer. The question what measures need to be taken to reach good results in all seasons including the transitional seasons will be answered.

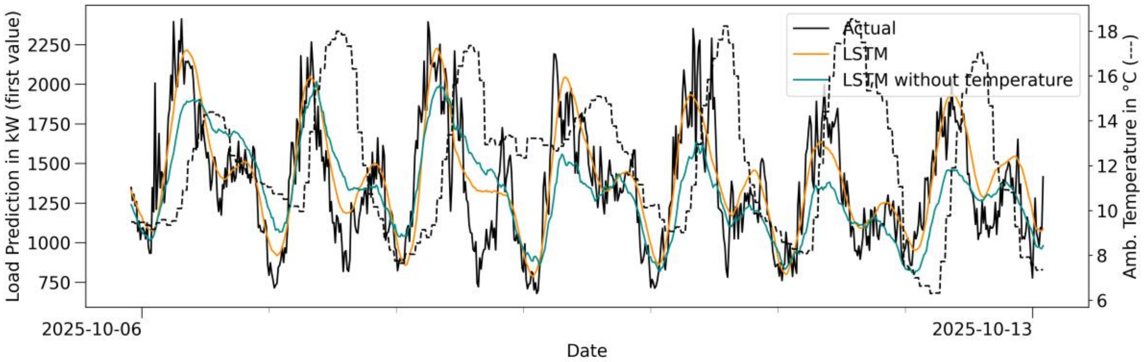

Fig. 1: First timestep of LSTM thermal load prediction with and without consideration of temperature in a week in October

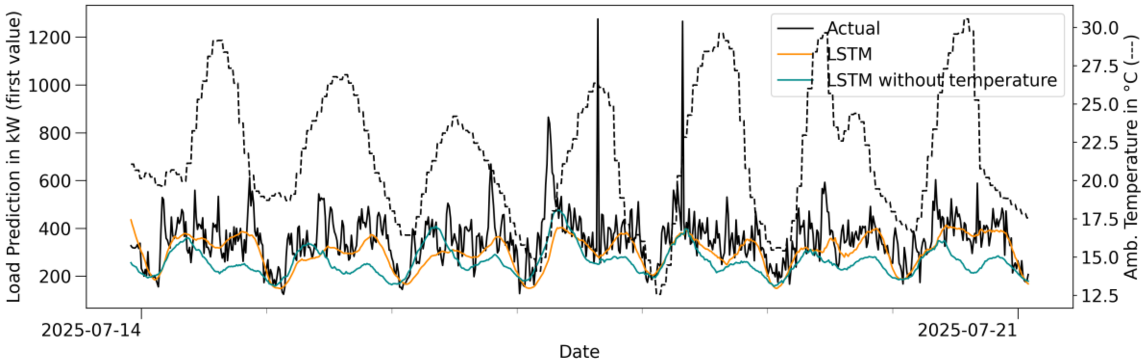

Fig. 2: First timestep of LSTM thermal load prediction with and without consideration of temperature in a week in July

6. Conclusions

In summary, our benchmark demonstrates that both decision-tree ensembles (XGBoost, LightGBM) and deep-learning architectures (LSTM-encoder-decoder with attention, CNN) can accurately forecast thermal loads in district-heating networks when ambient temperature is incorporated, particularly during the heating season. They consistently achieve lower RMSE and MAPE in winter, while in summer, prediction errors increase, highlighting the limited explanatory power of temperature-driven features when domestic hot-water demand dominates. Seasonal-specific feature selection improves model robustness during transitional periods, suggesting that adaptive feature sets could further close the performance gap. These results provide practical guidance for utilities in need of fast, season-aware model selection and enable the prediction to maintain high accuracy throughout the entire year.

7. References

Geysen, Davy; et.al. (2018): Operational thermal load forecasting in district heating networks using machine learning and expert advice. In: Energy and Buildings 162, S. 144–153. DOI: 10.1016/j.enbuild.2017.12.042.

Pressa, Christian; et.al. (2023): Benchmarking of State-Of-The-Art Machine Learning Methods for Highly Accurate Thermal Load Forecasting in District Heating Networks, S. 3154–3162. DOI: 10.52202/069564-0283.

Runge, Jason; Saloux, Etienne (2023): A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system. In: Energy 269, S. 126661. DOI: 10.1016/j.energy.2023.126661.