PAPER • OPEN ACCESS

Assessing impact of borehole field data's input parameters on hybrid deep learning models for heating and cooling forecasting: A local and global explainable AI analysis

To cite this article: N Ahmed *et al* 2023 *IOP Conf. Ser.: Mater. Sci. Eng.* **1294** 012056

Assessing impact of borehole field data's input parameters on hybrid deep learning models for heating and cooling forecasting: A local and global explainable AI analysis

N Ahmed[1,2,*], M Assadi[1], Q Zhang[1] and A A Ahmed[1]

[1] Department of Energy and Petroleum Engineering, University of Stavanger, 4021 Stavanger Norway
[2] US Pakistan Centre for Advanced Studies in Energy (USPCASE), National University of Sciences & Technology (NUST), H-12 Sector, Islamabad, Pakistan

[*] Correspondence: Naveed.ahmed@uis.no

Abstract. Achieving accurate performance forecasting of borehole heat exchanger is essential for optimizing ground source heat pump systems, enabling optimal control, and facilitating energy-efficient operations with enhanced sustainability of the built environment. This study aims to investigate and quantify the impact of model architecture, the number of input data sensors, and their accurate identification on multivariate hybrid deep learning models. Moreover, the significance of incorporating a recent development in deep learning to pay selective attention to the input data i.e., attention-based mechanisms in LSTM-CNN and CNN-LSTM architectures is also investigated. The significance of input parameters for the data-driven AI models is assessed through a significance interpretability analysis utilizing Explainable-AI local-method, namely Shapley Additive Explanations and global-explanation methods i.e., permutation feature importance method and Friedman statistical test. The findings highlight the efficacy of attention mechanisms in capturing temporal dependencies in LSTM-CNN-At and spatial patterns in CNN-LSTM-At, may not necessarily enhance their multistep forecasting capabilities for the borehole field data in comparison to LSTM-CNN architecture. The 24 hours ahead forecasting results show that the order of accuracy is LSTM-CNN> LSTM-CNN-At> CNN-LSTM> CNN-LSTM-At. The findings emphasize that by carefully designing the model layers, it is feasible to remove redundant borehole field sensors for data measurement while maintaining the forecasting accuracy of the hybrid data-driven models.

1. Introduction

The building sector's heating demands contribute significantly to global CO2 emissions and consume a substantial amount of energy. To foster the transition towards safe and sustainable heating and cooling solutions on a global scale, borehole coupled heat pumps (BCHPs) play a vital role [1]. Ensuring accurate performance forecasting of borehole heat exchangers (BHEs) becomes crucial for optimizing ground source heat pump (GSHP) systems. By achieving precise forecasts, system design can be optimized, energy usage can be reduced, and overall system reliability can be enhanced, thus fostering the sustainability of the built environment [2]. The performance of GSHP systems is influenced by several factors, including mass flow rates, heat transfer fluid temperatures and variable weather conditions. Consequently, effective control measures encompassing both demand-side management and the geothermal heat source are necessary to minimize overall energy consumption [3]. Accurate forecasting of GSHP energy output becomes imperative to curtail energy waste and facilitate the development of reliable models for assessing the thermal characteristics of BHEs.

In current work, deep learning models are investigated due to their adaptability in handling complex, multivariable borehole heat exchanger systems. In such scenarios with dynamic data conditions, deep neural networks excel by learning intricate relationships directly from data, making them well-suited for real-time monitoring and forecasting of non-linear systems. Achieving accurate forecasting with data-driven machine learning models is influenced by various factors, including input variables, training algorithms, exogenous variables, and model architecture [4]. Hybrid deep neural networks like LSTM-CNN and CNN-LSTM are known for their complexity and tend to outperform simpler models (e.g., Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN) random forest, linear regression, decision trees) in delivering accurate predictions. However, the interpretability of these multi-layered networks poses a challenge as they often behave as "black-box" models, making it difficult to comprehend the rationale behind specific predictions [5]. In real-world heating and cooling applications, the interpretability of simpler models instills greater trust among users. Therefore, achieving a balance between performance and interpretability is crucial for developers when deploying deep learning models [6]. This research confronts the significant challenge of identifying key input variables that reliably forecast heating and cooling temperatures in borehole heat exchangers based on seasonal building loads.

Individual algorithms, such as LSTM and 1-D CNN, have demonstrated effectiveness in various cases [7–9]. However, the rising popularity of hybrid deep learning models, such as LSTM-CNN and CNN-LSTM, can be credited to their capability of capturing both temporal and spatial dependencies within input datasets, leading to enhanced accuracy in heating and cooling load forecasting. In the study by Zhang et al. [10], a CNN-LSTM model was developed to predict the outlet temperature of borehole heat exchangers by considering spatial-temporal features. The model was trained using soil temperature field, inlet and outlet water temperatures. The proposed model exhibited higher prediction accuracy, with an average coefficient of determination (R2) value that surpassed the ANN, LSTM, Simple-RNN, and CNN algorithms by 3.54%, 2.35%, 5.75%, and 4.60%, respectively. In another investigation, Yao et al. [11] introduced a combined approach that involved a differentiated CNN-LSTM for short-term heat demand prediction in district-level heating systems. The objective of this approach was to forecast heat demand half an hour ahead, utilizing solely historical heat data. In the realm of building load prediction, Explainable Artificial Intelligence (EX-AI) plays a crucial role in enhancing user trust and providing transparency for hybrid deep neural networks. Research conducted by Tsoka and fellow researchers [12] has highlighted the effectiveness of EX-AI methods in reducing the required number of input features for accurate load forecasting while preserving accuracy. Chung et al. [13] investigated the input parameters for EX-AI-based load forecasting in office buildings across different climate regions. By utilizing local feature importance methods of SHAP and LIME, the study revealed improved identification of essential input variables compared to global explainable methods.

Determining the optimal selection of sensors from the borehole field for deep machine learning forecasting models can be a challenging task due to the diverse influences of building characteristics, weather variables, and internal heat-gain profiles on building load variability. The extensive investigation of deep machine learning models has shed light on their performance; however, the influence of model architecture, the number and accurate identification of input sensors on the performance of hybrid algorithms for heating and cooling multistep forecasting remains challenging. This study aims to examine and quantify this impact while exploring the significance of incorporating attention-based mechanisms in LSTM-CNN and CNN-LSTM architectures. The effectiveness of attention mechanisms in capturing temporal dependencies in LSTM-CNN-At and spatial patterns in CNN-LSTM-At is analyzed, with a specific focus on identifying essential input parameters for the deep AI models using Explainable-AI methods at both local and global levels. The implications of this research are significant, as the proposed multistep prediction hybrid model (LSTM-CNN) can contribute to industry applications such as 24-hour ahead prediction of BHE output, demand-side management of heating and cooling, and building operations.

2. Methodology

This study aims to investigate the impact of model architecture, the number of input sensors, and their accurate identification on the performance of hybrid algorithms for multistep forecasting in ground source heat pump applications. The methodology, illustrated in Figure 1, involves a series of steps starting with data extraction from real borehole heat exchanger installations. The collected data undergoes cleaning and pre-processing before training and validating hybrid deep learning models to achieve the desired forecasting accuracy. The constructed hybrid deep models, including LSTM-CNN, CNN-LSTM, and attention-based models LSTM-CNN-At and CNN-LSTM-At, are trained, validated, and tested using varying numbers of independent input parameters in three different variations of dataset (Variation-1, Variation-2, Variation-3). The trained deep learning models are then utilized to predict the performance of borehole heat exchangers 24 hours in advance, ensuring efficient management of thermal loads. Furthermore, the importance of input parameters for the hybrid deep models is determined through an interpretability analysis using Local and Global Ex-AI techniques.

2.1 Pre-processing and cleaning of dataset

In the current investigation three variations of datasets with different numbers of input parameters were utilized to train and test the developed deep learning algorithms, as outlined in Table 1. These datasets consisted of essential parameters derived from real BHE installations, ensuring their relevance in training the AI models. Specifically, the dataset used in this study was sourced from a geothermal installation situated at the Gale Bullman Multi-purpose building in Missouri, USA [14]. It encompassed 144 boreholes, each with a depth of 122m, which were interconnected to a large-scale GSHP system. The collected data spanned approximately one year, recorded at 15-minute intervals. To ensure data quality, a thorough cleaning process was implemented, resulting in a dataset devoid of significant outliers. Any identified outliers were removed and replaced with interpolated values. Furthermore, supplementary weather data including wind speed (V_{wind}), solar radiation (R_{sol}) and ambient temperature (T_{amb}) were obtained from a reputable US weather service provider [15].

Table 1. Different variations of data with different number of input parameters

Datasets	Names of input-parameters	Input-parameters
Variation-1	T_{out}, T_{in}, T_{amb} , m^0	4
Variation-2	T_{out}, T_{in}, T_{amb}, m^0, R_{sol}	5
Variation-3	T_{out}, T_{in}, T_{amb}, m^0, R_{sol},V_{wind}	6

The identification and handling of outliers in these datasets were performed using three standard deviation method[16]. The effectiveness of the data cleaning procedure enhanced data quality with significantly reduced noise levels over time. To assess the accuracy of the hybrid models developed in this study, two commonly used evaluation metrics, Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE), were employed and defined by the following equations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{1}$$

$$MAPE = \frac{100}{n}\sum_{1=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{2}$$

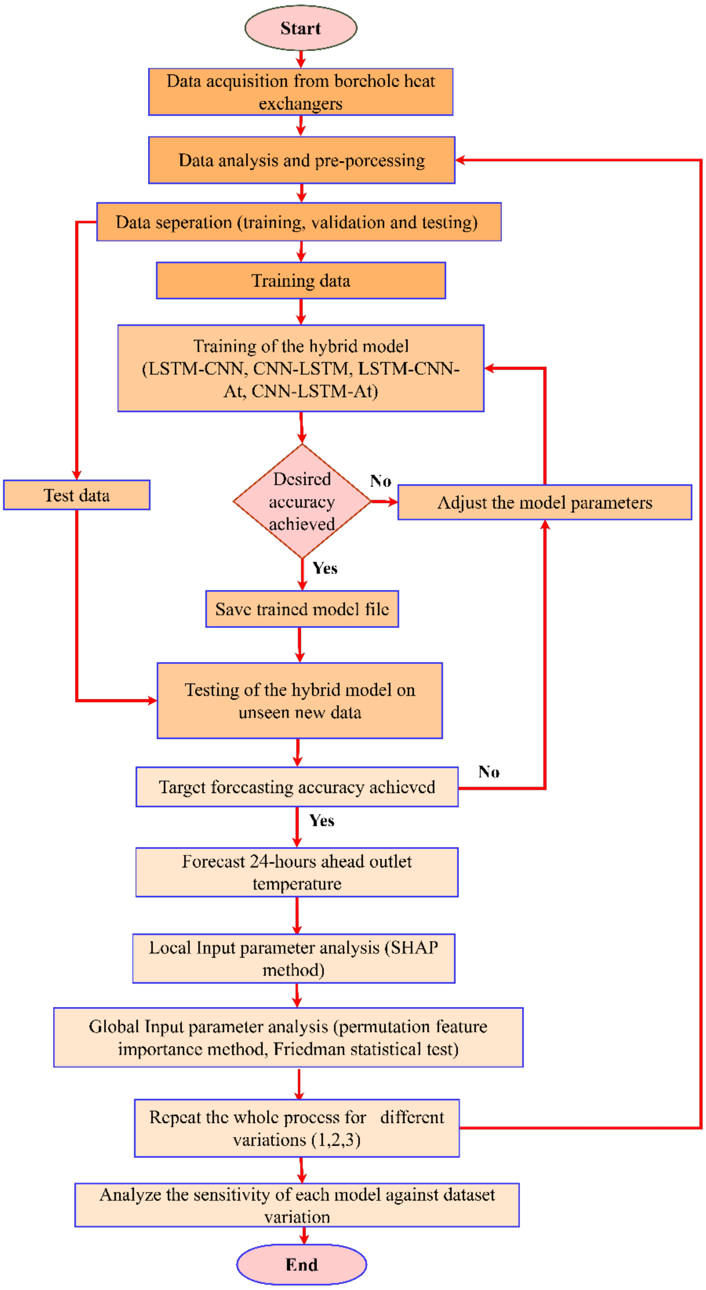


Figure 1. Overview of the implemented methodological approach for the current work

In the equations, n represents the total number of samples present in the dataset, y_i signifies the actual observed value at the ith time instance, and $\hat{y}_i$ denotes the corresponding predicted value

2.2 Feature importance analysis using local and global level methods

To determine the most crucial input parameters from BHE installations for accurate multistep forecasting of hybrid AI models and enhance their interpretability, three distinct methods are employed: a local-level method called Shapley Additive Explanations and two global-level methods, namely the permutation feature importance method and the Friedman statistical test. The SHAP method emplyed in the current work is a local interpretability approach providing insights into the contribution of individual input features (T_{out}, T_{in}, T_{amb}, m^0, R_{sol},V_{wind}) in generating predictions. In addition to the local-level analysis, the permutation feature importance method is used to assess the importance of input

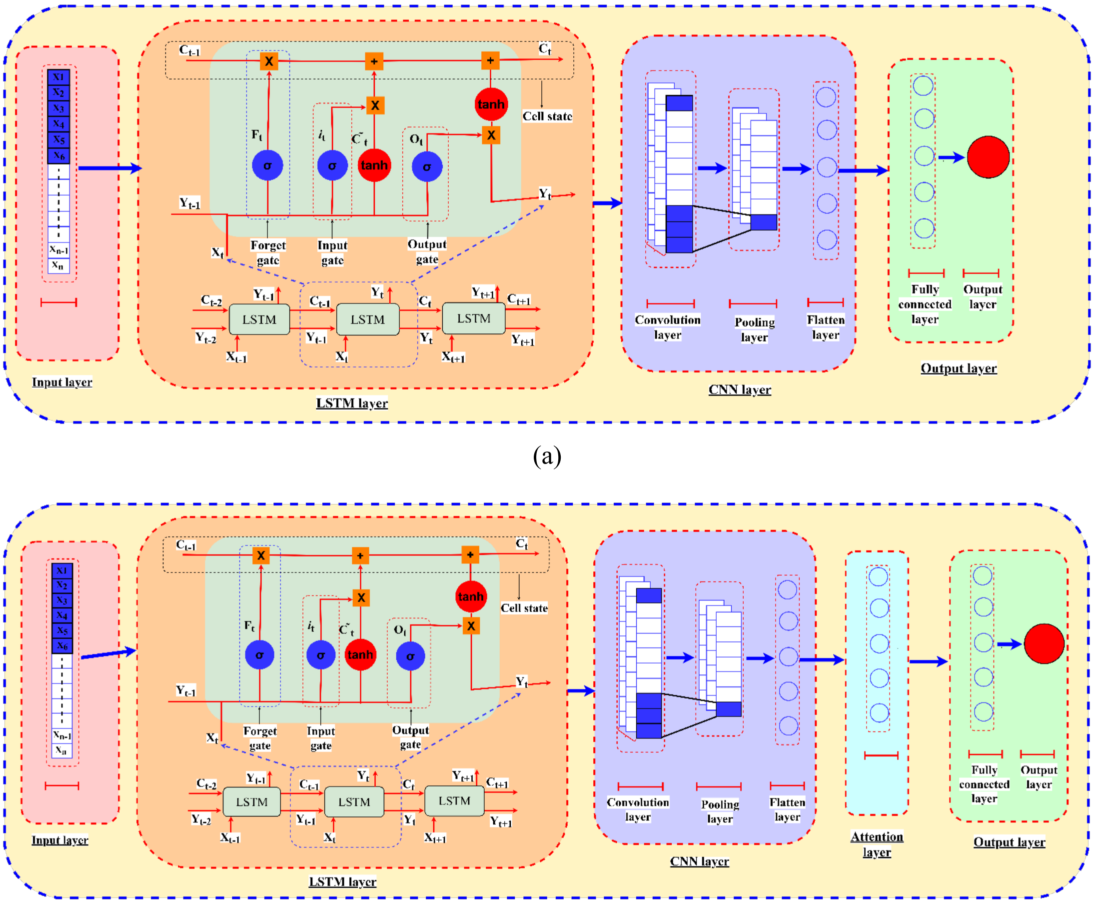
IOP Conf. Series: Materials Science and Engineering 1294 (2023) 012056 doi:10.1088/1757-899X/1294/1/012056

parameters on a global scale. This technique involved systematically permuting the values of all input parameters and evaluating the resulting influence on the model's forecasting performance. By measuring the change in model performance metrics (RMSE, MAPE), the relative importance of each input parameter was investigated in the overall model performance. Moreover, the Friedman statistical test is employed, which is a widely-used non-parametric test, to compare the ranks of the input parameters based on their influencing effect on the model's prediction capabilities. Collectively, these three methods provide a comprehensive insights into the significance and influence of input parameters from borehole installations on the deep learning hybrid AI models.

2.3 Hybrid deep learning models

In this study four distinct deep machine learning architectures are employed to predict the fluid outlet temperature of a borehole heat exchanger. These architectures are trained and tested using borehole field data, considering various numbers of input parameters, to forecast the outlet temperature for the next 24 hours. Each of these developed algorithms possesses its unique set of strengths and limitations, which are critical to consider. The LSTM-CNN architecture as shown in Fig. 2 (a), combines the Long Short-Term Memory network and Convolutional Neural Network layers. The LSTM layers capture temporal dependencies in the input sequence, while the CNN layers extract spatial features from the data. This architecture can effectively model both temporal and spatial patterns in the data, making it suitable for capturing complex relationships in time-series and spatial data. The LSTM-CNN architecture can be computationally expensive and may require a large amount of training data for optimal performance. Interpretability of the model can be challenging due to its complex nature. Whereas CNN-LSTM construction has the oppositely arrange layers combining the CNN layers with LSTM layers. The CNN layers extract spatial features from the input data, and the LSTM layers capture temporal dependencies. It is effective in capturing spatial patterns and temporal dependencies simultaneously. Similar to the LSTM-CNN architecture, this model may require a large amount of data for training and can be computationally intensive. Interpretability remains challanging due to the complexity of the model.

As illustrated in Fig. 2 (b), LSTM-CNN-At architecture arranges systematically LSTM and CNN layers with an attention mechanism. The attention mechanism function takes the LSTM-CNN layer's output as input and computes the attention weights to give more importance to certain time steps or features in the sequence The attention mechanism allows the model to selectively focus on relevant input features, enhancing its ability to capture important information. It can capture both temporal dependencies and spatial patterns while emphasizing important information.However, this architecture may require additional computational resources compared to the basic LSTM-CNN and CNN-LSTM models. Proper tuning of the attention mechanism and hyperparameters is essential for optimal performance. While the CNN-LSTM-At architecture combines CNN and LSTM layers with an attention mechanism. It leverages the spatial extraction capabilities of CNN layers and the temporal dependency modeling of LSTM layers, while also incorporating attention to highlight important features. Similar to other attention-based models, the CNN-LSTM-At architecture may require additional computational resources and careful hyperparameter tuning. The complexity of the model can make its interpretation challenging. The choice of architecture is made by taking into account the specific characteristics of the borehole heat exchanger dataset and the requirements of the forecasting task. This selection process is crucial as the considered architecture directly influences the model's capability to capture the inherent temporal and spatial patterns present in the measured BHE data.



(a)

(b)

Figure 2. Architecture of hybrid model combining (a) LSTM and CNN (b) LSTM-CNN with attention mechanism.

3. Results and discussion

The analysis presented in this section provides a quantitative assessment of how the input parameters influence the multistep forecasting capabilities of four hybrid algorithms.

3.1 Effect on multistep forecasting accuracy

In this study, three different variations are considered, each utilizing a distinct number of input parameters during the developed model training. In variation-1, data from four input sensors (T_{out}, T_{in}, T_{amb}, m^0) is used. The hybrid deep learning algorithms are employed to predict the outlet temperature for the next 24 hours, and the predicted values were compared with the actual measured values, as depicted in Figure 3. The accuracy results as shown in figure 6 revealed that the LSTM-CNN model achieved the highest accuracy, followed by the LSTM-CNN-At, CNN-LSTM, and CNN-LSTM-At models, respectively. Both the LSTM-CNN and LSTM-CNN-At models performed comparably well, exhibiting MAPE values of 1.29% and 1.59%, respectively. These hybrid algorithms combine the strengths of LSTM layers in capturing temporal dependencies and long-term memory, along with the effectiveness of CNN layers in capturing spatial patterns within the data. However, the introduction of the attention mechanism did not lead to improved accuracy. Instead, it resulted in reduced ability to capture temporal and spatial patterns within the data, accompanied by a higher number of trainable parameters. The CNN-LSTM and CNN-LSTM-At models exhibited MAPE values of 1.62% and 2.53%, respectively. The reverse order of layers in these models prioritizes capturing spatial features first and then extracting temporal dependencies, which led to a decrease in forecasting accuracy.

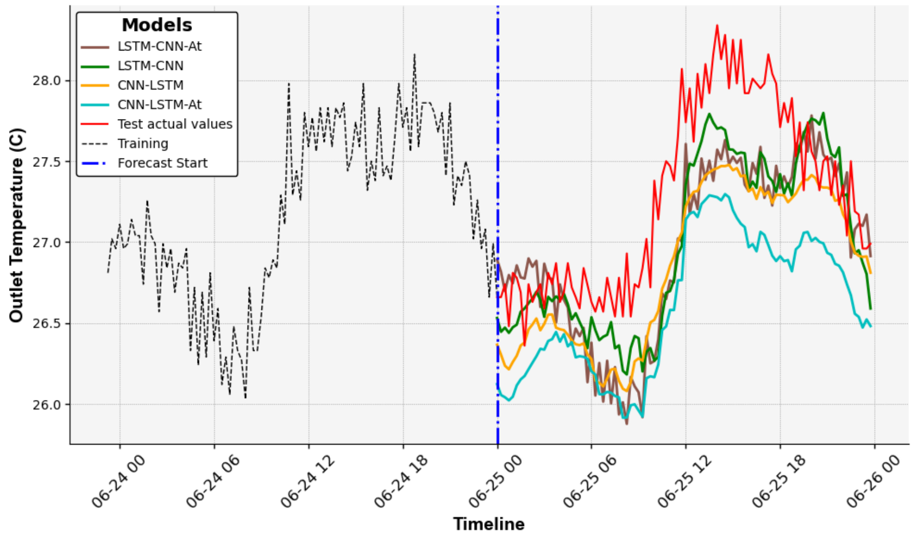


Figure 3. Performance of hybrid deep learning models on variation-1 for next 24-hour

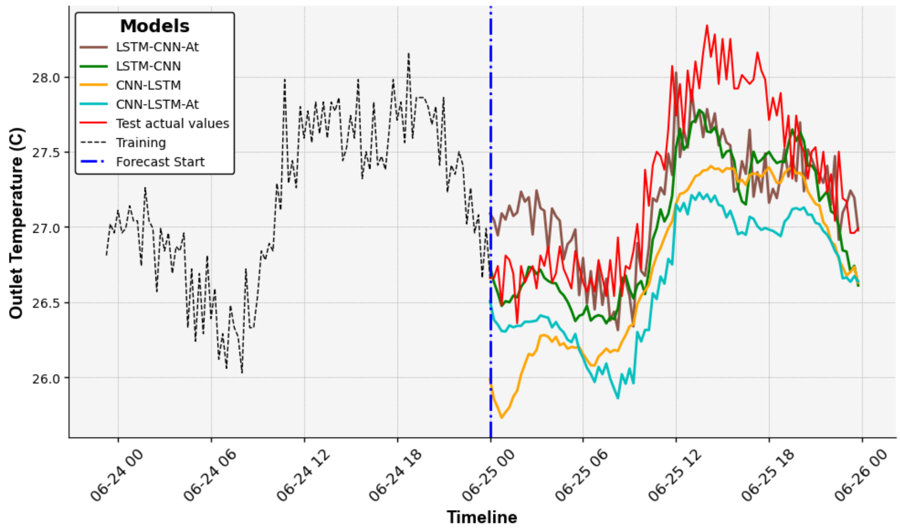


Figure 4. Performance of hybrid deep learning models on variation-2 for next 24-hour

In variation-2, an additional input parameter, solar radiation was included in the measured dataset of the borehole heat exchanger. This resulted in a total of five input parameters considered for the deep learning models (T_{out}, T_{in}, T_{amb}, m^0, R_{sol}). The results depicted in Figure 4 demonstrate that the inclusion of R_{sol} as an input feature has improved the forecasting accuracy of the hybrid algorithms, except for CNN-LSTM, for the next 24 hours. The order of prediction accuracy remains consistent with the previous variation, with LSTM-CNN achieving the highest accuracy, followed by LSTM-CNN-At, CNN-LSTM, and CNN-LSTM-At. Among these models, LSTM-CNN exhibits the lowest MAPE of 0.72%, indicating the highest level of accuracy. These findings suggest that the addition of solar radiation as an input parameter contributes to improved forecasting accuracy for most of the hybrid algorithms. However, CNN-LSTM-At performs relatively less accurately compared to other models.

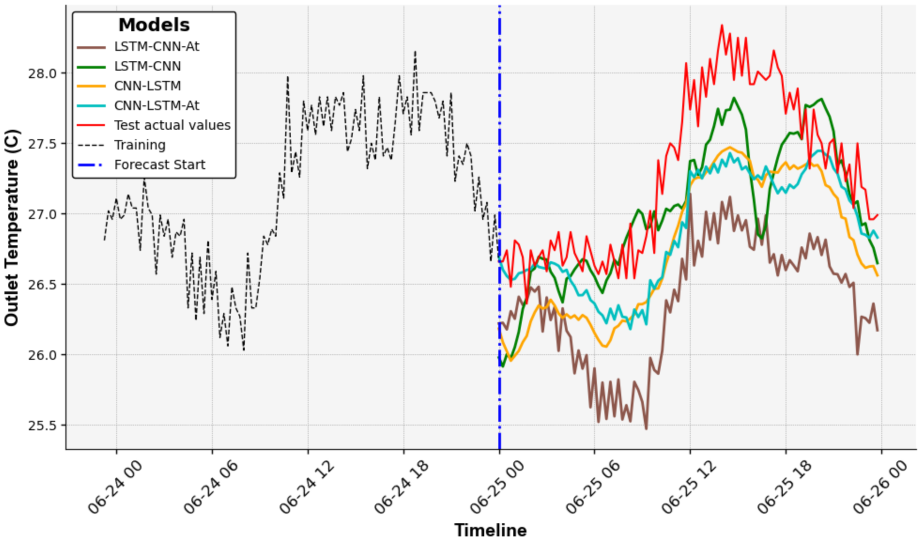


Figure 5. Performance of hybrid deep learning models on variation-3 for next 24-hour

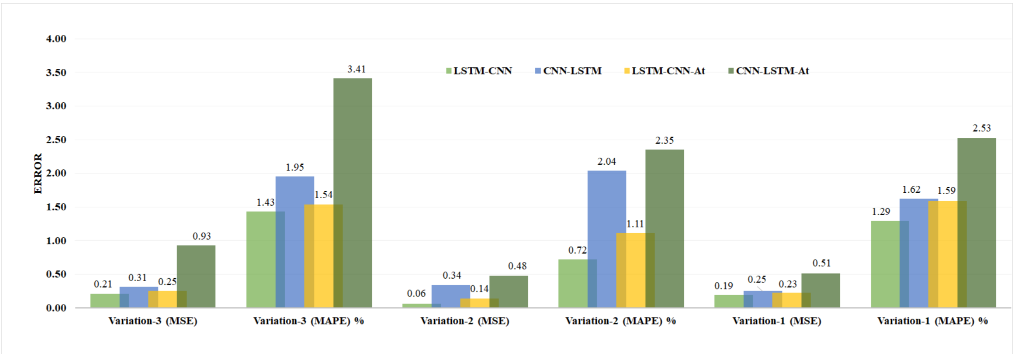


Figure 6. Performance evaluation (MAPE, MSE) of hybrid models under different input parameters

In variation-3, the number of input features considered for model training increased to six by including wind speed along with T_{out}, T_{in}, T_{amb}, m^0 and R_{sol}. The forecasting results presented in Fig. 5 demonstrate that the accuracy of CNN-LSTM increases, while the accuracy of LSTM-CNN, LSTM-CNN-At, and CNN-LSTM-At exhibit a decrease, with MAPE values of 1.95%, 1.43%, 1.54%, and 3.41% respectively. By conducting a comprehensive analysis of the forecasting results and the characteristics of the different hybrid models, valuable insights are gained regarding their suitability for predicting the outlet temperature of the borehole heat exchanger. Factors such as the number of input parameters, layer structure, and model complexities play a significant role in determining the performance of these models.

3.2 Feature importance analysis using global and local Ex-AI methods
In this section the critical input parameters for the data-driven AI models are identified through an interpretability analysis using Local and Global Explainable-AI techniques.

3.2.1 Local Ex-AI analysis. To evaluate the significance of input parameters of hybrid learning algorithms for borehole coupled heat pump systems, local-level explanations using the SHAP method was employed. This methodology leverages the Shapley value to quantify the contribution of each input measured sensor data (T_{in}, T_{amb}, m^0, R_{sol},V_{wind}) by assessing their impact on the output variable, T_{out}. Figure 7 presents the average impact of input parameters for Variation-1as measured by their corresponding SHAP values when employing the developed hybrid algorithms. The SHAP values provide insights into the range of average impact that each input parameter has on the output forecasting. Here only average impact results are discussed and the results indicate a consistent overall trend of relative average impact on the output across the considered hybrid algorithms. However, slight variations in the magnitude of their effects is observed. For the best-performing LSTM-CNN model, the order of importance of input parameters in terms of average impact is $T_{in} > T_{out} > m^0 > T_{amb}$, while for the least-performing CNN-LSTM-At model, the order is $m^0 > T_{amb} > T_{in} > T_{out}$.

In Variation-2 as depicted in Fig.8, the results demonstrate that the inclusion of solar radiation as an input parameter significantly impacts the forecasting accuracy and cannnot be disregarded. Depending on the architecture of the hybrid model, this inclusion can either lead to a decrease in accuracy (CNN-LSTM) or an improvement (LSTM-CNN, LSTM-CNN-At, CNN-LSTM-At). The order of average impact of input parameters for LSTM-CNN-At ($m^0 > T_{amb} > T_{in} > T_{out} > R_{sol}$), CNN-LSTM-At ($m^0 > T_{in} > R_{sol} > T_{amb} > T_{out}$), LSTM-CNN ($m^0 > R_{sol} > T_{in} > T_{out} > T_{amb}$), and CNN-LSTM ($T_{in} > T_{amb} > T_{out} > m^0 > R_{sol}$) highlights that m^0 and T_{in} have the most significant impact, while the impact of other parameters with reduced magnitude, such as R_{sol}, cannot be ignored.

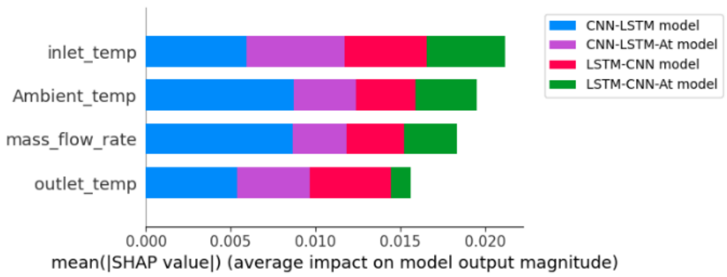


Figure 7. Quantifying input parameter impact on model outputs with SHAP values for Variation-1

Figure 9 illustrates the mean SHAP values for Variation-3 with the error analysis revealing that the inclusion of the sixth input parameter(V_{wind}), leads to a decrease in accuracy for all algorithms except CNN-LSTM, which shows an improvement. The order of parameter importance, based on the average impact value, for LSTM-CNN and LSTM-CNN-At is is $T_{in} > T_{out} > m^0 > R_{sol} > T_{amb} > V_{wind}$ and $T_{in} > T_{out} > T_{amb} > m^0 > R_{sol} > V_{wind}$, respectively. Overall, the results indicate that the addition of V_{wind} as an input parameter has the least significant impact on the forecasting accuracy of all hybrid algorithms, but its contribution cannot be overlooked. These results highlight the models' sensitivity to variations in the input parameters and emphasize the significance of obtaining precise measurements to ensure accurate forecasting of the outlet temperature.

3.2.2 Global analysis. Global interpretability of the hybrid deep learning models is investigated using the permutation feature importance method and Friedman statistical test. Fig. 10 shows global feature analysis using permutation feature importance method for three different variations. By randomly permuting the values of a feature in each variation and evaluating the model's performance on the permuted data, it was investigated how much the model's accuracy decreases. The greater the drop in accuracy, the more important that feature is for the model's multistep predictions. Overall results depict that the order of feature importance for variation-1($T_{in} > T_{out} > T_{amb} > m^0$), variation-2 ($T_{in} > T_{out} > T_{amb} > m^0 > R_{sol}$) and variation-3 ($T_{in} > T_{out} > T_{amb} > R_{sol} > V_{wind} > m^0$) very much depend on the model architecture. A higher feature importance score of inlet temperature for each variation suggests that this feature has a stronger influence on the model's predictions. The impact magnitude of other features varies depending on the architecture of hybrid model.

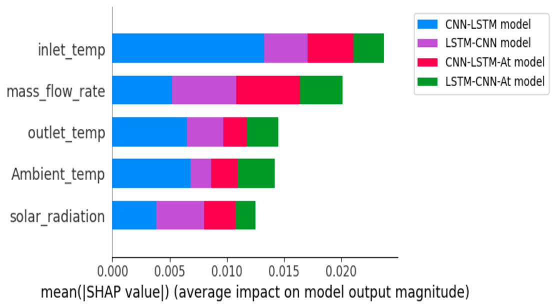


Figure 8. Quantifying input parameter impact on model outputs with local explainable SHAP values for Variation-2

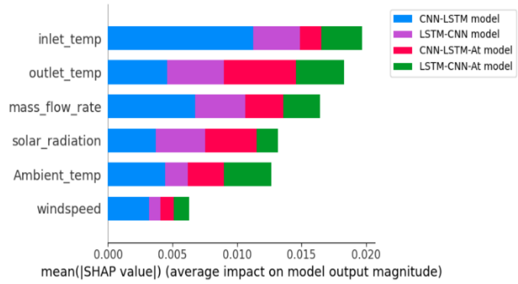
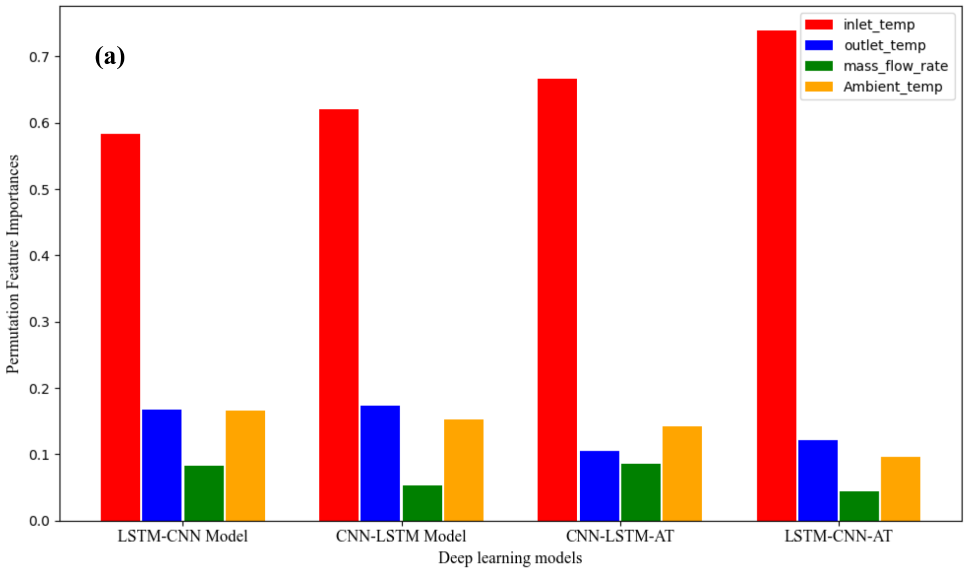
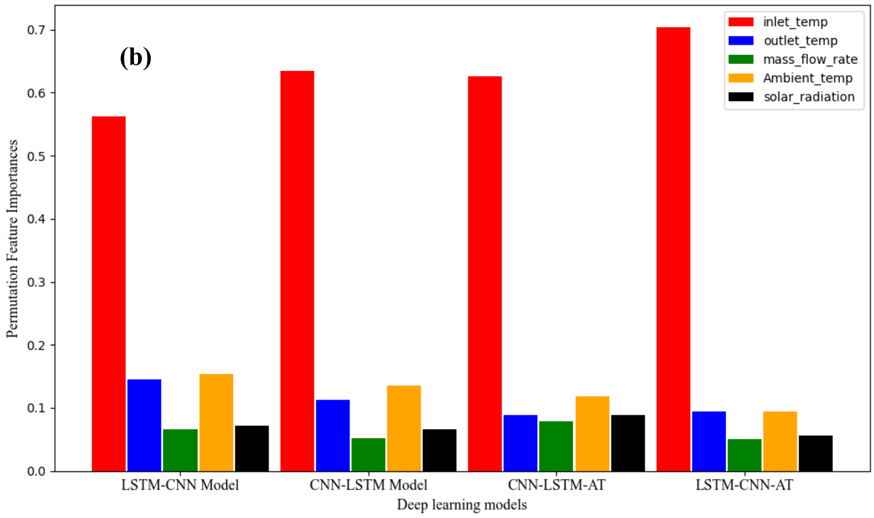


Figure 9. Quantifying input parameter impact on model outputs with local explainable SHAP values for Variation-3

To compare the performance of multiple hybrid models on the considered variations, a non-parametric test i.e., Friedman statistical test is used to determine whether there are significant differences among different sets of data. The Friedman test ranked the deep learning hybrid models based on their performance on multiple runs with various input data samples. Based on the average ranks obtained from the Friedman test, the performance order from best to worst as follows: LSTM-CNN model with an average rank of 1.33 (best performing), LSTM-CNN-At with a rank of 2.0, CNN-LSTM with a rank of 3.00, and CNN-LSTM-At with a rank of 3.66 (worst performing). The small differences in the average ranks indicate that the models' performance is quite comparable, and no statistically significant distinctions were found. The test results yielded an F-value of 5.80 and a p-value of 0.12. The p-value indicates that there is no significant difference in the performance of the four hybrid deep learning models at the chosen significance level of 0.05. Therefore, the null hypothesis cannot be rejected, suggesting no significant distinctions in the models' performance.





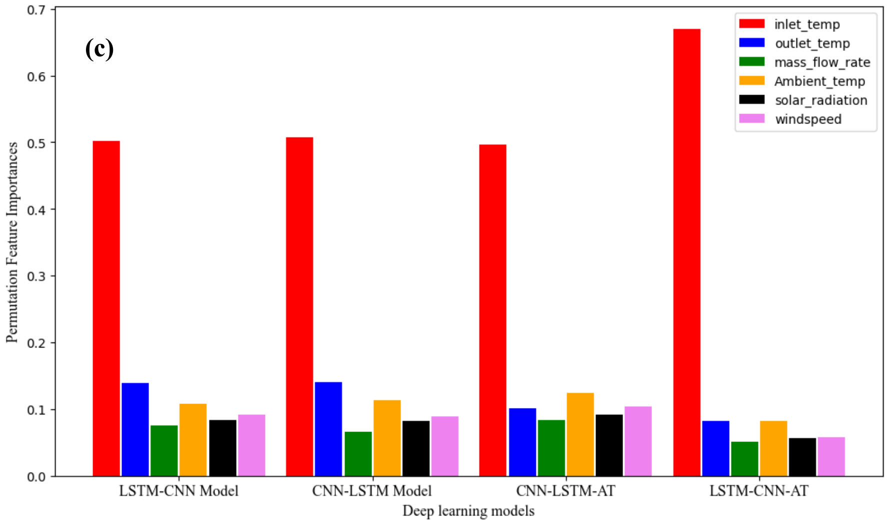


Figure 10. Global feature analysis using permutation feature importance method for (a) Variation-1, (b) Variation-2, (c) Variation-3

4. Conclusions

This study explores the impact of training algorithms, input parameters, model architecture, and exogenous variables on the multistep forecasting accuracy of data-driven models. It aims to identify key input sensors for deep learning-based load forecasting and investigates the significance of input parameters on 24-hour-ahead predictions for efficient thermal load management. The importance of input parameters is assessed using interpretability analysis with Ex-AI techniques, including SHAP analysis for local interpretability and the permutation feature importance method with the Friedman statistical test for global interpretability. The findings highlight the efficacy of attention mechanisms in capturing temporal dependencies in LSTM-CNN-At and spatial patterns in CNN-LSTM-At, may not necessarily enhance their multistep forecasting capabilities for the borehole field data in comparison to LSTM-CNN architecture. Local interpretability of the hybrid models by using SHAP method reveal that incorporating additional input features from BHE installations can improve the prediction accuracy.

However, it is important to consider the model architecture carefully, as there is a threshold beyond which adding more input parameters may not yield further improvements. Global interpretability of the hybrid deep learning models by using Permutation feature importance method show that the inlet temperature is the most influential parameter for all three variations and impact magnitude of other features varies depending on the architecture of hybrid model. Through this research, a comprehensive understanding of the interplay between model architecture, input sensors, and their identification is gained, enabling improved multistep forecasting performance in ground source heat pump applications.

Acknowledgments: We are indebted to the staff of the Department of Energy and Petroleum Engineering (IEP) at the University of Stavanger and for allowing to present this paper. The research leading to these results has received funding from the Norway Grants 2014–2021 via the National Centre for Research and Development in Warsaw.

References

[1] International Energy Agency (IEA) 2022 World Energy Oultlook, 1–524.
[2] Ahmed A, Assadi M, Kalantar A and Sliwa T A 2022 Critical Review on the Use of Shallow Geothermal Energy Systems for Heating and Cooling Purposes, *Energies,* 1–22.
[3] Ahmed N, Assadi M, Abbas A and Banihabib R 2023 Optimal design, operational controls, and data-driven machine learning in sustainable borehole heat exchanger coupled heat pumps : Key implementation challenges and advancement opportunities, *Energy for Sustainable Development*, **74**, 231–257, doi: 10.1016/j.esd.2023.04.004.
[4] Wang Z, Hong T and Piette M A 2020 Building thermal load prediction through shallow machine learning and deep learning, *Appl Energy*, **263**, 114683, doi: 10.1016/j.apenergy.2020.114683.
[5] Vishwarupe V, Joshi P M, Mathias N, Maheshwari S, Mhaisalkar S and Pawar V 2022 Explainable AI and Interpretable Machine Learning: A Case Study in Perspective," *Procedia Comput Sci*, **204**, 869–876, doi: 10.1016/j.procs.2022.08.105.
[6] Behzadi A, Holmberg S, Duwig C, Haghighat F, Ooka R and Sadrizadeh S 2022 Smart design and control of thermal energy storage in low-temperature heating and high-temperature cooling systems: A comprehensive review, *Renewable and Sustainable Energy Reviews*, **166**, 112625, doi: 10.1016/j.rser.2022.112625.
[7] Muzaffar S and Afshari A 2019 Short-term load forecasts using LSTM networks, *Energy Procedia*, **158**, 2922–2927, doi: 10.1016/j.egypro.2019.01.952.
[8] Jang J, Han J and Leigh S B 2022 Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks, *Energy Build*, **255**, 111647, doi: 10.1016/j.enbuild.2021.111647.
[9] Cai M, Pipattanasomporn M and Rahman S 2019 Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, *Appl Energy*, **236**, 1078–1088, doi: 10.1016/J.APENERGY.2018.12.042.
[10] Zhang W, Zhou H, Bao X and Cui H 2023 Outlet water temperature prediction of energy pile based on spatial-temporal feature extraction through CNN–LSTM hybrid model, *Energy*, **264**, 126190, doi: 10.1016/j.energy.2022.126190.
[11] Yao F, Zhou W, Al Ghamdi M, Song Y and Zhao W 2022 An integrated D-CNN-LSTM approach for short-term heat demand prediction in district heating systems, *Energy Reports*, **8**, 98–107, doi: 10.1016/j.egyr.2022.08.087.
[12] Tsoka T, Ye X, Chen Y, Gong D and Xia X 2022 Explainable artificial intelligence for building energy performance certificate labeling classification, *J Clean Prod*, **355**, 131626, doi: 10.1016/j.jclepro.2022.131626.
[13] Chung W J and Liu C 2022 Analysis of input parameters for deep learning-based load prediction for office buildings in different climate zones using eXplainable Artificial Intelligence, *Energy Build*, **276**, 112521, doi: 10.1016/j.enbuild.2022.112521.
[14] Smith D C and Elmore A C 2018 The observed effects of changes in groundwater flow on a

borehole heat exchanger of a large scale ground coupled heat pump system *Geothermics*, **74**, 240–246, doi: 10.1016/j.geothermics.2018.03.008.

[15] Crossing V 2023 Global Forecast and History Data, Accessed: Jan. 05, 2023. [Online]. Available: https://www.visualcrossing.com/weather-data-editions#

[16] *Applied Statistical Modeling and Data Analytics* 2018, doi: 10.1016/c2014-0-03954-8.