# Predicting the performance of hybrid ventilation in buildings using a multivariate attention-based biLSTM Encoder – Decoder neural network

**Gaurav Chaudhary[1*], Hicham Johra[2], Laurent Georges[1], Bjørn Austbø[1]**

[1]Department of Energy and Process Engineering & ENER-SENSE, NTNU, Trondheim, Norway
[2]Department of the Built Environment, Aalborg University, Aalborg Øst, Denmark
[*]Corresponding author: gaurav.chaudhary@ntnu.no

**Abstract**. Hybrid ventilation (coupling natural and mechanical ventilation) is an energy-efficient solution to provide fresh air for most climates, given that it has a reliable control system. To operate such systems optimally, a high-fidelity control-oriented model is required. It should enable near-real time forecast of the indoor air temperature and humidity based on operational conditions such as window opening and HVAC schedules. However, widely used physics-based simulation models (i.e., white-box models) are labour-intensive and computationally expensive. Alternatively, black-box models based on artificial neural networks can be trained to be good estimators for building dynamics. This paper investigates the capabilities of a multivariate multi-head attention-based long short-term memory (LSTM) encoder-decoder neural network to predict indoor air conditions of a building equipped with hybrid ventilation. The deep neural network used for this study aims to predict indoor air temperature dynamics when a window is opened and closed, respectively. Training and test data were generated from detailed multi-zone office building model (EnergyPlus). The deep neural network is able to accurately predict indoor air temperature of five zones whenever a window was opened and closed.

## 1. Introduction

Buildings are responsible for over 40% of global energy use and 36% of greenhouse gas emissions, with Heating, ventilation, and air conditioning (HVAC) operation accounting for almost half of this [1,2].[3]. Reducing the overall building energy demand and footprint has thus become an urgent task to meet the current sustainability goals and tackle current energy crises. To that end, natural ventilation is seen as one of the most effective passive energy-saving measures for buildings. HVAC systems in buildings coupled with natural ventilation (hybrid ventilation) can theoretically provide the most efficient energy system for any climate, given they have a fast and reliable control system. For such systems, a high-fidelity control-oriented prediction model is required. It should be able to forecast building dynamics in near-real time for given operational conditions such as window opening and HVAC operational schedules. This is, however, challenging due to internal time-varying building dynamics, disturbances from occupants, lighting and plug-in loads, and external factors like outdoor weather and HVAC operations. Developing efficient prediction models accounting for these buildings dynamics has been a harsh bottleneck to implementing predictive-based control strategies [4].

The case of buildings with hybrid ventilation is particularly challenging for building dynamic modelling. In a scenario where natural ventilation is allowed, e.g., by opening a window, the indoor temperature variation depends on many parameters, such as the indoor-outdoor temperature gradient, the effective window opening area, the HVAC mode (heating , cooling , fan or dry mode), and the internal loads. Such building dynamics can be modelled using well-established laws of physics (white box approach) [4], or by using simplified first-principle conservation equations (first law of thermodynamics) and empirical data for model calibration (grey box approach) [5]. A typical white box modelling tool like EnergyPlus or IDA-ICE requires time and effort to define and collect the required parameters to model and then calibrate the building model. A grey box model, such as a resistor-capacitor (RC) network, requires the robust estimation of its parameters. Reinforced by the massive deployment of metering and sensing technologies in buildings, the black box approach employing data-driven models for building dynamics prediction has increased in popularity in recent years [6]. Black box models also have the advantage of low development costs. However, such models require a large amount of training data. This, however, can be solved by using transfer learning methods that couple data generated from white box modelling tools like EnergyPlus [7] and system identification techniques. A black box model trained on various operating conditions and scenarios simulated with the white box model of a generic building could, in theory, be operational for real building applications after only tuning the former with a very small dataset [7–9].

Following that principle, the current study aims to investigate the use of deep neural networks (DNNs) as a black box model for the prediction of the indoor environment. DNN models are suitable for complex building dynamics as they can handle non-linear multivariable modelling problems. It has been shown that a neural network with enough hidden units can approximate arbitrary continuous functions defined on a closed and bounded set [10,11]. This paper investigates if and how a DNN can be trained to predict the effect of window opening on the indoor air temperature dynamics in a conditioned building. The training data for this DNN was generated with the building energy simulation tool EnergyPlus.

## 2. Model architecture

Deep learning models based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) like long short-term memory (LSTM) [12] and gated recurrent units (GRUs) [13] have been widely used in applications like speech recognition [14], natural language processing [15] and, computer vision [16]. Similarly to building energy models, these applications have data in the form of time-series. In RNNs such as LSTMs or GRUs, each input corresponds to an output for the same time step. However, in many real cases, there is a need to predict an output sequence given an input sequence of different length without correspondence between each input and output. This situation is called Sequence-to-Sequence Mapping model and lies behind commonly used applications like Machine Translation [17], Question Answering, Chat-bots, and Text Summarization. Sequence-to-Sequence models are also known as the Encoder-Decoder models, where, as the name suggests, the encoder takes the input sequence and the decoder decodes that information to give out the output sequence.

The most commonly used neural network unit in Encoder-Decoder models is the LSTM unit. However, they seem to suffer from "short-term memory" over longer time series sequences. Advancements like the "attention mechanism" [18,19] and the "transformer" [20,21] used in conjunction with RNNs have improved their prediction performance [22]. The attention mechanism performs better by giving higher weights to relevant parts of the sequence while minimizing irrelevant parts by allocating them lower weights, thus improving the model's accuracy [23]. The multi-head attention (MHA) module introduced in the transformer model [20] runs through the attention mechanism several times in parallel, attending to different parts of the sequence differently. Compared to LSTMs, MHA retains direct connections to all previous timestamps in the sequence, allowing information to propagate over much longer sequences. To summarize, the recurrent networks are excellent at capturing the local temporal characteristics of a sequence, while the transformer model can learn long-term

dynamics. The encoder-decoder model used in this study utilizes both RNNs, specifically bi-directional LSTMs (biLSTM) [24], and several components of the transformer model like the Self – MHA, Cross – MHA and Gated Residual Networks (GRNs).

All neural networks can, in theory, produce multiple outputs for the same set of hyperparameters due to both randomness of the optimization used during training, and the overall distribution of input data. In line with other time series forecasting models [17], this prediction model also generates prediction intervals on top of point forecasts. This is done by simultaneous prediction of various percentiles ($50^{th}$, $90^{th}$, $95^{th}$ and $99^{th}$) at each time step using quantile regression [25].

This model takes the past seven days of data as input and predicts 24 hours into the future. The data has a temporal granularity of 15 minutes, so the model takes 672 data points from the past and predicts 96 data points. The model takes two sets of inputs: "*Known past inputs*" for the input encoder and *"Known future inputs"* for the input decoder. "*Known past inputs*" are the past seven days of weather data, time information, and zone-specific information like occupancy, external loads, temperature setpoints in zones, and actual indoor air temperatures (IATs) of zones. Specifically for this study, the data includes the opening signals for all windows. *"Known future inputs"* are the future one day of weather forecast, targeted heating setpoints of HVAC, and a targeted window opening signal. These inputs are similar to what a zone level controller can access to condition a zone. The forecasted weather is created from the real weather data of the future, scrambled with Gaussian noise of zero mean and standard deviation of 0.01, as the exact future weather data would not be known in real scenarios. The model's output is the future one day of IATs of all zones. A schematic the model is shown in Figure 1.

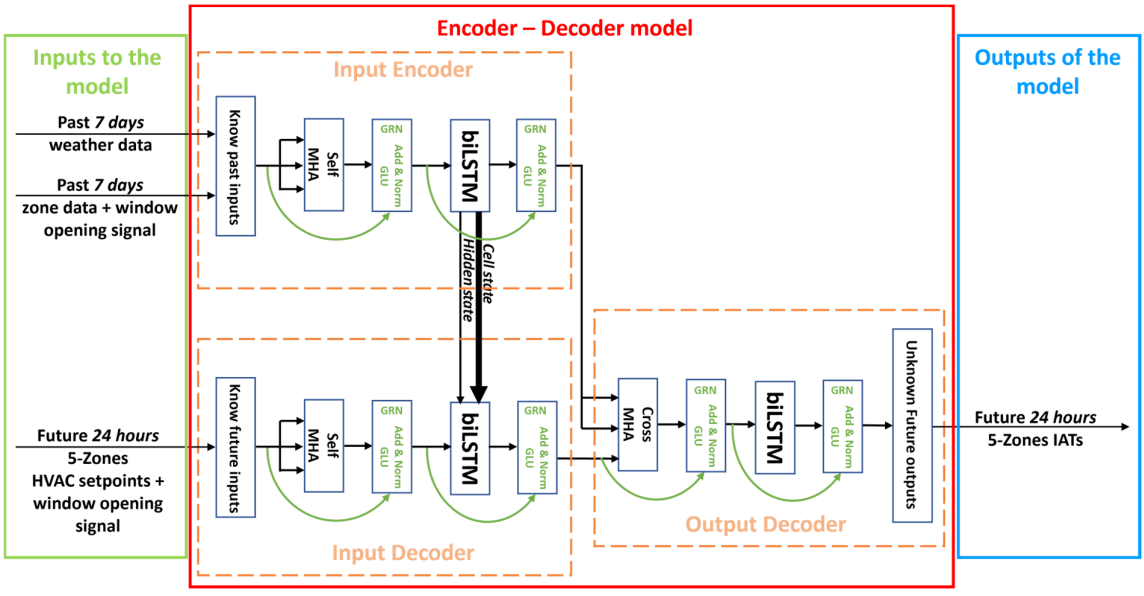


Figure 1: Final encoder-decoder model for predicting building dynamics using both recurrent networks and components of the transformers model.

## 3. Training

### *3.1. Study case description and training data generation*

Training data was produced using the building energy simulation tool EnergyPlus v22.1.0. For this study, a small-size office building with 12-hour working schedule was modelled. The building model is a generic 5-zone EnergyPlus example file geometry. As shown in Figure 2, the building is a single-floor rectangle of dimensions 30 m x 15 m. It has four exterior and one interior zones with a ceiling height of 2.4 m. There are windows on all four facades, with the south and north facades also having glass doors.

Overhangs shade the south-facing window and door. There are no internal openings like doors inside the zones. The U-value of internal and external walls used was 1.6 W/m²K and 2.8 W/m²K, respectively. All fenestrations are high-performance low-E Planibel triple pane argon-filled windows with U-value of 0.7 W/m²K. To reduce overheating, automatic window shading control lowers the interior shade when the outside temperature exceeds 23 °C. The building uses a variable refrigerant flow (VRF) HVAC system for conditioning the zones, whereas the ventilated air for the building is delivered by a dedicated outdoor air system (DOAS).

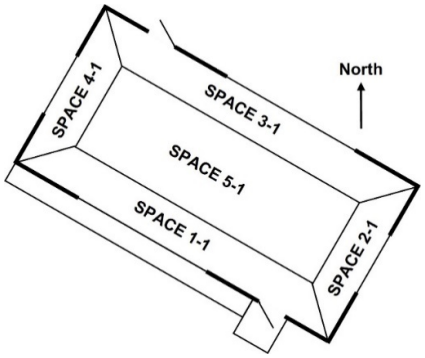


Figure 2: 5-zone office building simulated in EnergyPlus for generating training dataset.

The schedules for occupancy and miscellaneous electric loads (MELs) were produced using an agent-based stochastic occupancy simulator [26] to represent the dynamic and stochastic movement of occupants in an office building. The schedule of lighting was kept to working hours, i.e., 07:00 – 19:00. Figure 3 compares a typically used occupancy schedule for energy simulation and a stochastic occupancy schedule.

For temperature heating setpoints of five zones, a multi-pseudo random sequence (m-PRS) input signal was generated to excite the deep neural network model for all possible changes in heating setpoints. The signal excitation method is one of the system identification methods aiming to excite one input with a signal that does not correlate with other inputs. This way, the deep neural network learns the underlying dynamics of thermal setpoints with indoor temperature and all boundary conditions. In several studies [27–30], m-PRS input signals modelled as random signals were used to excite indoor air temperature for a large office building.

Five different short-period (sixth-order) signals were used for this training. In the m-PRS signals, the value will change randomly from any value in the range of 18 to 22 °C (with 0.5 °C increment) and stay at that value for any random amount of time, which is determined by the "order" of the signal. For this study, the cooling setpoint was set 5 °C above the heating setpoint (i.e., 23 to 27 °C). During occupied hours, the mPRS signal was used for heating and cooling, whereas for non-occupied hours, the setpoints fell to a setback of 15 and 30 °C, respectively.

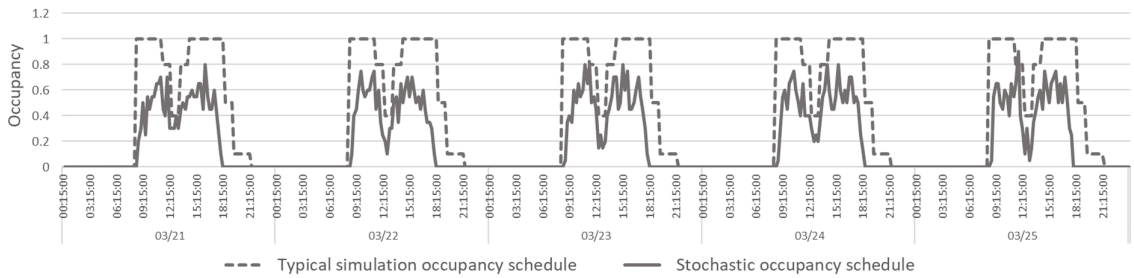


Figure 3: A typically used occupancy schedule for energy simulation and a stochastic occupancy schedule for 5 days

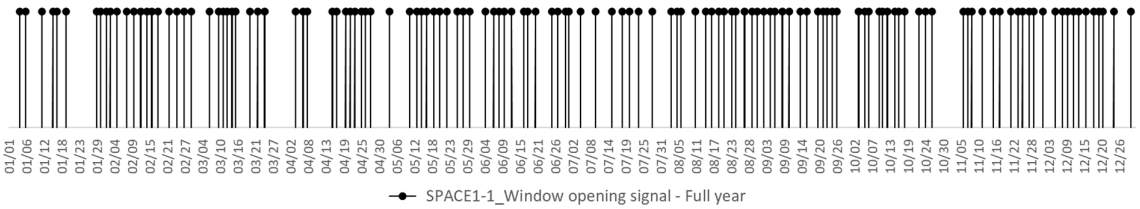


Figure 4: One of the PRBS input signals for window opening and closing for one full year.

The modelling of opening/closing of windows in EnergyPlus was done using the *ZoneVentilation:WindandStackOpenArea* object [31]. This object allows users to define the limits on the outdoor conditions (temperature, wind speed) that determine whether the window is open or closed. The equation used to calculate the ventilation rate driven by wind is based on "*Wind and Stack with Open Area*" model [32]. Similar to exciting the neural network for the temperature heating setpoints, four different PRBS input signals were used to excite the neural network for the effect of opening and closing of four windows for different boundary conditions. Each one in the signal denotes a window opening of 30 mins as shown in Figure 4. Figure 5 shows the indoor air temperature of Space 1-1 along with the mPRS input signal for heating and cooling setpoints, outdoor air temperature and opening/closing signal. The effect of opening windows can be seen with a sharp decline followed with gradual rise of indoor air temperature.

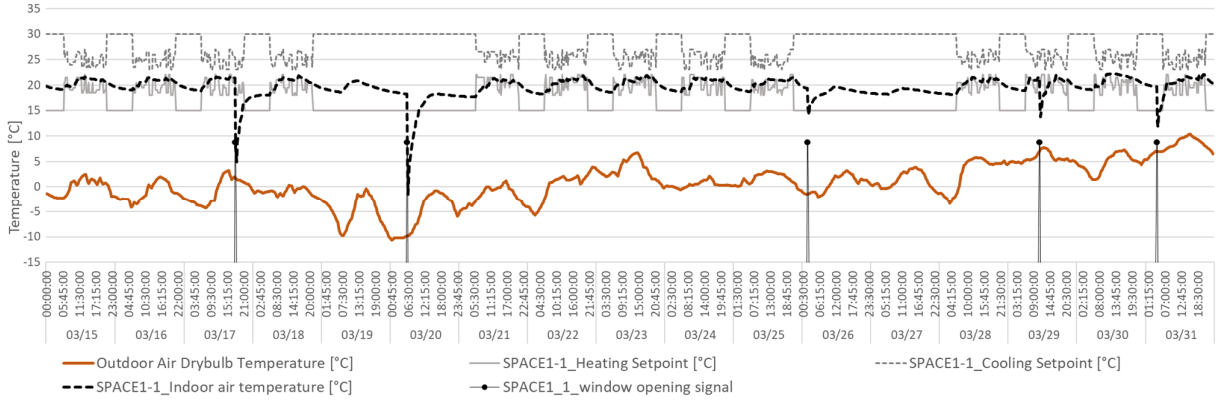


Figure 5: The indoor air temperature of Space 1-1 along with mPRS input signal for heating and colling setpoints, outdoor air temperature and opening/closing signal.

**Table 1.** Multivariate inputs-output variables used for the study

| **Input variables (*Known past inputs*) = $u_{i(t-671,\,t)}$** | | | **Interval** |
|---|---|---|---|
| $T^{out}$ | Outside Dry-bulb temperature | °C | [ -30.0, 40.0] |
| $H^{out}$ | Relative humidity of air | % | [ -30.0, 40.0] |
| $W^{out}$ | Wind speed component from East to West | m/s | [ -30.0, 40.0] |
| $I\_Norm^{out}$ | Direct normal radiation | W/m$^2$ | [ 0, 1300.0] |
| $I\_Hor^{out}$ | Diffuse radiation on horizontal surface | W/m$^2$ | [ 0, 1300.0] |
| $h$ | Sine and Cosine of the Hour of the day | - | [ -1.0, 1.0] |
| $d$ | Sine and Cosine of the Day of the week | - | [ -1.0, 1.0] |
| $m$ | Sine and Cosine of the Month of the year | - | [ -1.0, 1.0] |
| $hol$ | Holiday schedule | Boolean | [ 0, 1] |
| $E_i$ | Equipment load of the zone *i(1-5)* | W | [ 0.0, 1000.0] |
| $Occu_i$ | Occupancy in the zone *i(1-5)* | - | [ 0, 30] |
| $WS_i$ | Window opening signal of zone window *i(1-4)* | Boolean | [ 0, 1] |
| $SP_i$ | Heating Setpoint of zone *i(1-5)* | °C | [ 15.0, 30.0] |
| $T^{in}_i$ | Indoor temperature of zone *i(1-5)* | °C | [ 10, 40.0] |

| **Input variables (*Known future inputs*) = $v_{i(t+1,\,t+96)}$** | | | **Interval** |
|---|---|---|---|
| $T^{out}$ | Outside Dry-bulb temperature | °C | [ -30.0, 40.0] |
| $H^{out}$ | Relative humidity of air | % | [ -30.0, 40.0] |
| $W^{out}$ | Wind speed component from East to West | m/s | [ -30.0, 40.0] |
| $I\_Norm^{out}$ | Direct normal radiation | W/m$^2$ | [ 0, 1300.0] |
| $I\_Hor^{out}$ | Diffuse radiation on horizontal surface | W/m$^2$ | [ 0, 1300.0] |
| $h$ | Sine and Cosine of the Hour of the day | - | [ -1.0, 1.0] |
| $d$ | Sine and Cosine of the Day of the week | - | [ -1.0, 1.0] |
| $m$ | Sine and Cosine of the Month of the year | - | [ -1.0, 1.0] |
| $hol$ | Holiday schedule | Boolean | [ 0, 1] |
| $WS_i$ | Window opening signal of zone window *i(1-4)* | Boolean | [ 0, 1] |
| $SP_i$ | Heating Setpoint of zone *i(1-5)* | °C | [ 15.0, 30.0] |

| **Output variables (*Unknown future outputs*) = $y_{i(t,\,t+96)}$** | | | |
|---|---|---|---|
| $T^{in}_i$ | Indoor temperature of zone *i* | °C | [ 10, 40.0] |

To predict the indoor air temperature, essential features that could be commonly available in the building management system of typical office buildings were selected as inputs (see Table 1). In contrast, zone-specific features like equipment and occupancy can be recorded manually by check-in or $CO_2$ sensors [33–35]. The hour of the day, the day of the week, the month of the year, and the holiday schedule were also used as inputs. The hour input leads to knowing the difference between the temperature profile during the occupied and unoccupied time. Day and holiday input leads to distinguishing between business and weekend days.

### *3.2. Training procedure*

The dataset is split into training (60 % of the data), validation (20 % of the data) and testing (20 % of the data) sets from the 12 months of data. Since deep learning models can perform well with data where all the features are in the same range, all the features were scaled to the range [-1.0, 1.0] using *MinMaxScaler*. The minimum and maximum value of the scaling is mentioned in Table 1 in the interval column. The deep neural network was developed using *Tensorflow 2.11.0* and the *Keras* library on *Python 3.10.0*. The hyperparameters used for the model and training are listed in Table 2. The ADAM [36] optimizer was used in conjunction with a custom learning rate scheduler based on cosine annealing with warm restarts [20,37] to optimize the loss of the neural network.

**Table 2.** Hyperparameters for the deep neural network

| Hyperparameter name | Description | Value |
|---|---|---|
| n_past | Number of past data points for input data | 672 |
| n_future | Number of future data points predicted | 96 |
| noise_sd | Standard deviation for Gaussian noise for weather data used in *Known future input* | 0.01 |
| batch_size | Diffuse radiation on horizontal surface | 256 |
| CWR_min_lr | Minimum learning rate for Cosine Warmup Restart (CWR) learning rate scheduler | 0.00005 |
| CWR_max_lr | Maximum learning rate for CWR | 0.005 |
| CWR_cycle_length | Cosine annealing length for CWR | 10 |
| CWR_warmup_length | Restart warmup length for CWR | 3 |
| optimizer | Loss optimizer used for training | ADAM |
| loss | Loss function for the model | quantile loss |
| mha_head | Number of multi-head attention module | 4 |
| dropout | Dropout % used for training | 0.3 |
| RNN | Type of recurrent neural network units used | biLSTM |
| RNN - units | Number of recurrent neural network units whenever used | 200 |
| Activation | Activation function used in RNN units | tanh |

For evaluating actual and predicted IATs, the Coefficient of variance of root mean squared error (CVRMSE) was used as the evaluation metric. Because it is a unit-less percentage value, it is an easily comparable metric, making it the preferred choice for comparing prediction models [38,39]. It is derived by normalizing the RMSE with the mean of actual data. It is defined as follows:

$$CVRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}}$$

Here, $y_i$, $\hat{y}_i$, and $\bar{y}$ denote the actual IATs for sample $i$, the predicted IATs for sample $i$, and the mean of actual IATs of all samples, respectively.

## 4. Results

The test data for the study was the last 20% of the yearly data, which the prediction model has never seen before. There were 5 112 instances where the prediction was done for the next 96 steps using data from the previous 672 steps. Figure 5 shows five selected instances for every four zones with a window. Each subplot shows the actual IAT, heating setpoint and window opening signal (presented as a black marker) and probabilistic forecast of IAT for that instance. Each forecast has 90%, 95% and 99% confidence intervals presented in decreasing opacity of green and the 50th quantile in bold red. The

various instances of prediction were selected to have a mix of both bad and good forecasts, window opening signal being far and close to step 0 of prediction instance, window opening having and not having an effect on IAT, and degree of effect of the window opening. While Figure 6 shows the instances of prediction for the whole 96 steps, it could also be interesting how the prediction was for a certain Nth step in future for all instances combined. Figure 7 shows the 24$^{th}$ step ahead for all instances. It shows five different subplots for five different zones with the same features, as shown in Figure 6.

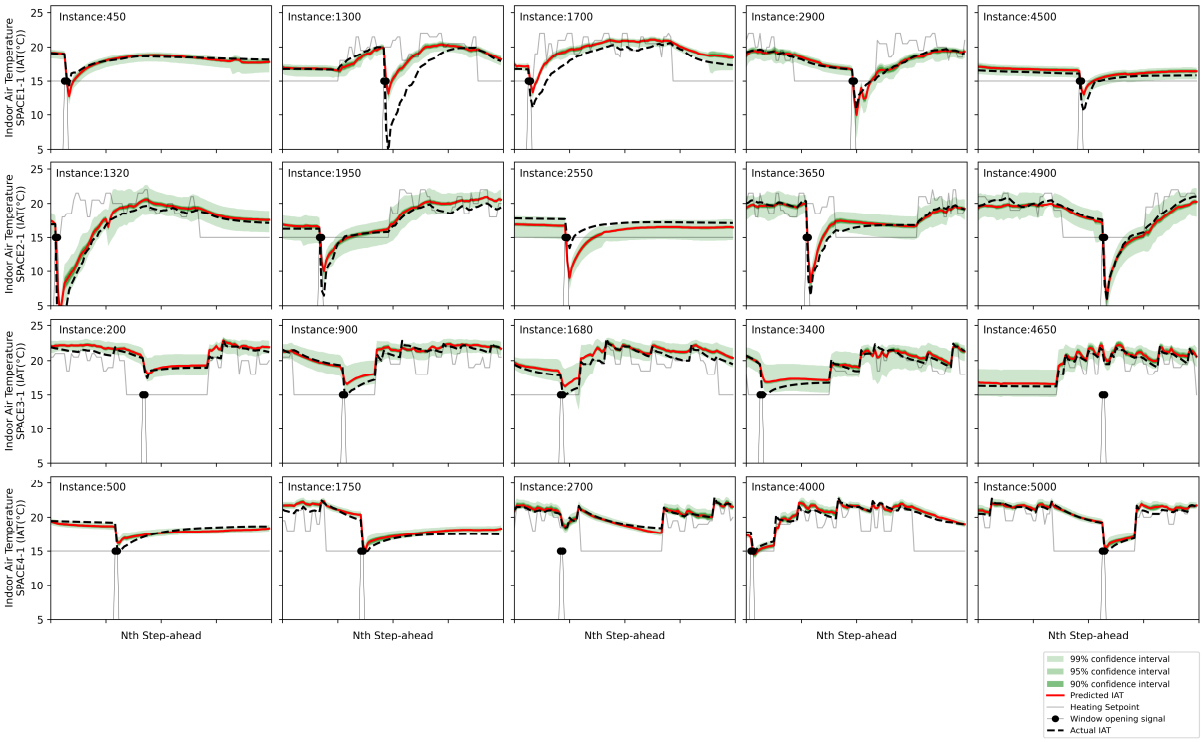


Figure 6: The real vs predicted indoor air temperature of all zones at various instances of prediction, shown by instance number. For example, Subplot Instance:900 in row 3 and column 2 shows the whole 96 future timestep prediction done at instance 900 for zone: SPACE3-1. A black marker shows the window opening signal. The high resolution version of the figure above is available in supplementary material.

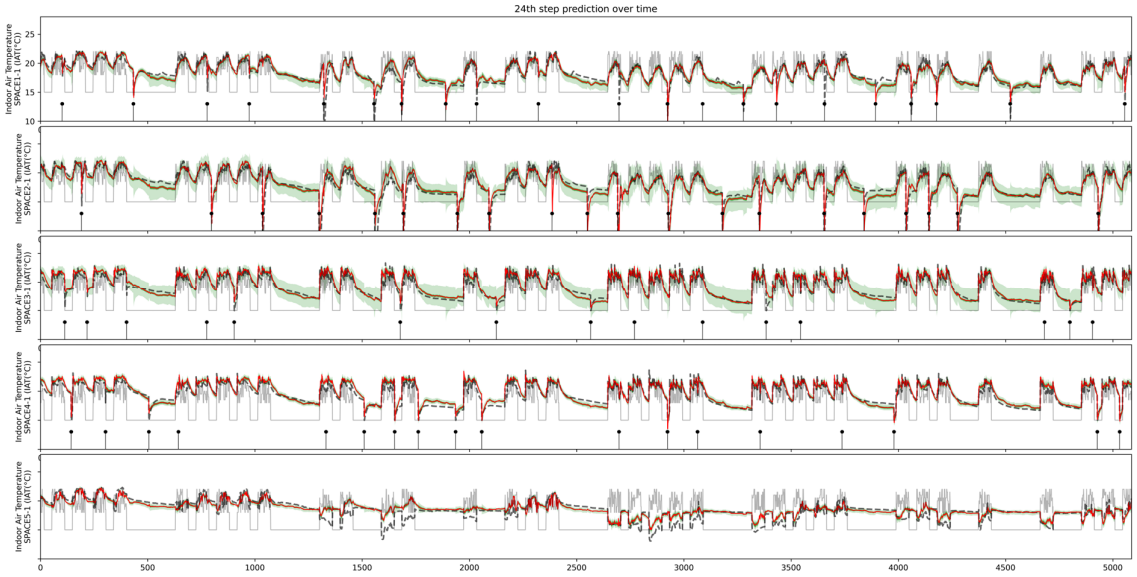


Figure 7: 24th step prediction done at all instances of prediction. The legend for the figure is same as Figure 6. The high-resolution version of the figure above along with wider figures for 1st, 24th, 48th, 72nd and 96th step prediction over time, is available in the supplementary material.

The qualitative results show that a deep neural network could predict the IAT with a random heating setpoint signal and a sharp decrease and gradual rise of temperature when a window is opened. In some instances where the error is high, like Instance:1300 of SPACE1-1 and Instance:2550 of SPACE2-1, the prediction was good when the control signal was closer to t=0 step ahead. Overall, it can be seen that error was highest for SPACE2-1. As can be observed in Figure 6, window opening severely affected IAT. This may be due to the fact that wind direction was predominantly flowing from southeast to northwest for the testing period adding more randomness to the dynamics of SPACE 2-1. This could potentially be solved by adding wind direction to input data for future studies.

The quantitative error in the predictions is given for all zones in Figure 8. The left subplot shows the normal distribution of CVRMSE (%) of all quantiles with actual IAT for all prediction instances done for the testing period. At the same time, the right subplot shows how the error has changed for certain steps into the future for all instances. The CVRMSE (%) shown in the right plot is that between the actual IAT and the 50th quantile of the predicted IAT. SPACE2-1 had the largest error distribution, followed by SPACE5-1. For SPACE3-1, the error was high for the 95% and 99% confidence intervals, which is also evident from Figure 7.

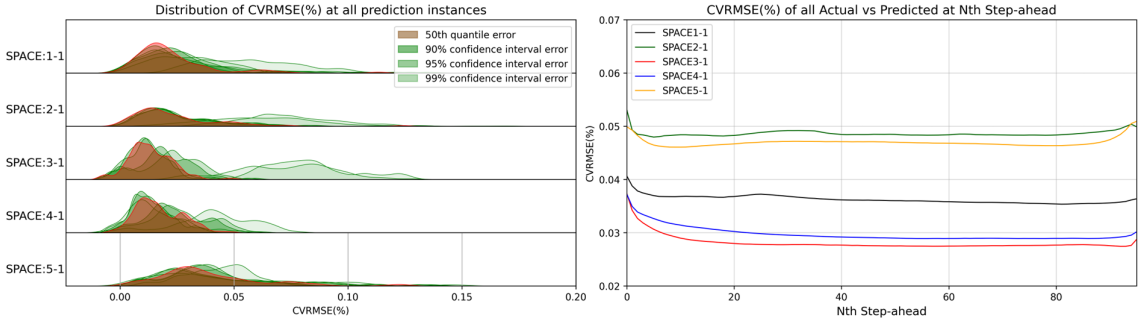


Figure 8: The left plot shows the distribution of CVRMSE for all 96 steps prediction instances done for the full testing period. The right plot of the figure shows the forecast error of actual temperature vs 50$^{th}$ quantile predicted for all certain Nth step- ahead.

## 5. Discussion and conclusion

The model used in this paper is fairly advanced in comparison to other time series prediction models and Sequence-to-Sequence neural networks applied to building energy systems in the literature. Other statistical time series models such as autoregressive exogenous models (ARX) and autoregressive moving average models with exogenous inputs (ARMAX) are linear and time-variant in nature. They perform badly when presented with nonlinearities and sudden uncertainties in the system. Sequence-to-sequence neural networks using RNNs in their basic form, although nonlinear in nature, were not able to capture varying building dynamics. Both the above-mentioned models were also not able to take targeted input such as heating setpoints and window opening signals separately. The complexity of the final model was the result of many model iterations developed after analyzing the limitations of other model and the need of control-oriented model. The use of the transformer model components enabled capturing long term building dynamics with high accuracy, where specifically components like cross-MHA enabled using targeted input data separately forming its dynamics relationships with output data and other input data.

This paper demonstrates that a deep learning-based neural network can be used as an estimator of building indoor air temperature given a heating setpoint and control signal for the window opening. It was observed that the prediction model was able to predict the incoming drop in IATs when it expected a window to be open. This can also be used inversely, where a window opening can be used as a control strategy to regulate comfort criteria and reduce energy consumption. The training of the deep learning model can be extended to predict other necessary comfort criteria-related features like indoor humidity levels and $CO_2$ levels, as well as energy use. Such trained models can be used in model predictive control [40] or a reinforcement learning architecture where all controllable parameters in a building can be regulated to get targeted comfort with minimum energy use.

This paper also showed how a deep learning model could be trained for various possibilities using an energy model of a building. The prediction model was tested on boundary conditions, occupancy schedules, heating setpoints and window opening signals, which the prediction model had not seen before. This clearly shows the signs of transfer learning. This training approach using system identification methods can be extended where merely a simple representative energy model of a building is required.

## Supplemantary material

The high resolution version of Figure 6: https://github.com/gaurav306/NSB23-Predicting-the-performance-of-hybrid-ventilation-in-buildings-using-a-multivariate-attention-/blob/main/Figure-06.png

The high resolution version of Figure 7: https://github.com/gaurav306/NSB23-Predicting-the-performance-of-hybrid-ventilation-in-buildings-using-a-multivariate-attention-/blob/main/Figure-07.pdf

**Acknowledgements**
The authors acknowledge the support from the strategic research program ENERSENSE at the Norwegian University of Science and Technology (NTNU).

**References**

[1] Grözinger J, Boermans T, John A, Wehringer F, Seehusen J. Overview of Member States information on NZEBs: Background paper—Final Report. Cologne, Germany: ECOFYS GmbH 2014.
[2] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy and Buildings 2008;40:394–8.
[3] Balaras CA, Dascalaki EG, Gaglia AG, Droutsa K, Kontoyiannidis S. Energy performance of European buildings. vol. 47977, 2007, p. 387–96.
[4] Cígler J, Gyalistras D, Široky J, Tiet V, Ferkl L. Beyond theory: the challenge of implementing model predictive control in buildings. vol. 250, 2013.
[5] Blum DH, Arendt K, Rivalin L, Piette MA, Wetter M, Veje CT. Practical factors of envelope model setup and their effects on the performance of model predictive control for building heating, ventilating, and air conditioning systems. Applied Energy 2019;236:410–25. https://doi.org/10.1016/j.apenergy.2018.11.093.
[6] Drgoňa J, Picard D, Kvasnica M, Helsen L. Approximate model predictive building control via machine learning. Applied Energy 2018;218:199–216. https://doi.org/10.1016/j.apenergy.2018.02.156.
[7] Pinto G, Messina R, Li H, Hong T, Piscitelli MS, Capozzoli A. Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics. Energy and Buildings 2022;276:112530. https://doi.org/10.1016/j.enbuild.2022.112530.
[8] Chen Y, Tong Z, Zheng Y, Samuelson H, Norford L. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. Journal of Cleaner Production 2020;254:119866. https://doi.org/10.1016/j.jclepro.2019.119866.
[9] Gao Y, Ruan Y, Fang C, Yin S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy and Buildings 2020;223:110156. https://doi.org/10.1016/j.enbuild.2020.110156.
[10] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Networks 1991;4:251–7.
[11] Ruano AE, Crispim EM, Conceiçao EZ, Lúcio MMJ. Prediction of building's temperature using neural networks models. Energy and Buildings 2006;38:682–94.
[12] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9:1735–80.
[13] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. ArXiv Preprint ArXiv:14123555 2014.
[14] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks, Ieee; 2013, p. 6645–9.
[15] Devlin J, Zbib R, Huang Z, Lamar T, Schwartz R, Makhoul J. Fast and robust neural network joint models for statistical machine translation, 2014, p. 1370–80.
[16] Davis A, Rubinstein M, Wadhwa N, Mysore GJ, Durand F, Freeman WT. The visual microphone: Passive recovery of sound from video 2014.
[17] Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. ArXiv Preprint ArXiv:160908144 2016.
[18] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. ArXiv Preprint ArXiv:14090473 2014.
[19] Luong M-T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. ArXiv Preprint ArXiv:150804025 2015.

[20] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in Neural Information Processing Systems 2017;30.
[21] Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting 2021;37:1748-1764 %@ 0169-2070.
[22] Shih S-Y, Sun F-K, Lee H. Temporal pattern attention for multivariate time series forecasting. Machine Learning 2019;108:1421-1441 %@ 1573-0565.
[23] Abbasimehr H, Paki R. Improving time series forecasting using LSTM and attention models. Journal of Ambient Intelligence and Humanized Computing 2022;13:673-691 %@ 1868-5145.
[24] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 2005;18:602–10. https://doi.org/10.1016/j.neunet.2005.06.042.
[25] Koenker R, Bassett Jr G. Regression quantiles. Econometrica: Journal of the Econometric Society 1978:33-50 %@ 0012-9682.
[26] Chen Y, Hong T, Luo X. An agent-based stochastic Occupancy Simulator. vol. 11, Springer; 2018, p. 37–49.
[27] Prívara S, Váňa Z, Gyalistras D, Cigler J, Sagerschnig C, Morari M, et al. Modeling and identification of a large multi-zone office building, IEEE; 2011, p. 55–60.
[28] Li X. Net-zero Building Cluster Simulations and On-line Energy Forecasting for Adaptive and Real-Time Control and Decisions. Drexel University; 2015.
[29] Royer S, Thil S, Talbert T, Polit M. A procedure for modeling buildings and their thermal zones using co-simulation and system identification. Energy and Buildings 2014;78:231–7. https://doi.org/10.1016/j.enbuild.2014.04.013.
[30] Zhang L. Data-driven whole building energy forecasting model for data predictive control. Drexel University; 2018.
[31] Winkelmann FC. Modeling windows in EnergyPlus. Proc IBPSA, Building Simulation 2001.
[32] ASHRAE I. 2009 ASHRAE handbook: Fundamentals. American Society of Heating, Refrigeration and Air-Conditioning Engineers; 2009.
[33] Han H, Jang K-J, Han C, Lee J. Occupancy estimation based on CO2 concentration using dynamic neural network model. Proc AIVC 2013;13.
[34] Zuraimi M, Pantazaras A, Chaturvedi K, Yang J, Tham K, Lee S. Predicting occupancy counts using physical and statistical Co2-based modeling methodologies. Building and Environment 2017;123:517–28.
[35] Arief-Ang IB, Salim FD, Hamilton M. DA-HOC: semi-supervised domain adaptation for room occupancy prediction using CO2 sensor data, 2017, p. 1–10.
[36] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization 2017.
[37] Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts 2016. https://doi.org/10.48550/ARXIV.1608.03983.
[38] Wu D, Hur K, Xiao Z. A GAN-Enhanced Ensemble Model for Energy Consumption Forecasting in Large Commercial Buildings. IEEE Access 2021;9:158820–30. https://doi.org/10.1109/ACCESS.2021.3131185.
[39] Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012.
[40] Drgoňa J, Arroyo J, Cupeiro Figueroa I, Blum D, Arendt K, Kim D, et al. All you need to know about model predictive control for buildings. Annual Reviews in Control 2020;50:190–232. https://doi.org/10.1016/j.arcontrol.2020.09.001.