Dynamic Statistical Learning with Engineered Features Outperforms Deep Neural Networks for Smart Building Cooling Load Predictions

Yiren Liu
School of Data Science
City University of Hong Kong
yirenliu2-c@my.cityu.edu.hk

S. Joe Qin *
School of Data Science
City University of Hong Kong
joe.qin@cityu.edu.hk

Xiangyu Zhao *
School of Data Science
City University of Hong Kong
xianzhao@cityu.edu.hk

Yixiao Huang
City University of Hong Kong
yixihuang9-c@my.cityu.edu.hk

Shenglong Yao
City University of Hong Kong
shenglyao2-c@my.cityu.edu.hk

Guo Han
City University of Hong Kong
guohan2-c@my.cityu.edu.hk

Abstract

Cooling load predictions for smart building operations play an important role in optimizing the operation of heating, ventilation, and air-conditioning (HVAC) systems. In this paper, we report a cooling load prediction solution for real municipal buildings in Hong Kong set up in a recent global AI competition. We show that dynamic statistical learning models with engineered features from domain knowledge outperform deep learning alternatives with optimal efforts. The proposed solution for the global AI competition was conferred a Grand Prize and a Gold Award by the panel of internationally renowned experts. We report the results of data preprocessing based on cooling operation knowledge, feature engineering from HVAC system knowledge, and dynamic statistical learning algorithms to build the models. To search for the best model to predict the cooling load, deep learning models with LSTM and gated recurrent units are extensively studied and compared with our proposed solution.

1 Introduction

Deep neural networks for dynamic modeling have gained much attention and wide applications in recent years. For example, the long short-term memory (LSTM) network by Hochreiter and Schmidhuber [1997] is one of the most cited approaches since its inception. The successes in speech recognition, machine translation, and robotic control have inspired industrial and engineering domains to explore its benefit for dynamic predictions. Given a prediction problem, deep learning with a recurrent structure is often the first preferred method. Published works are often satisfied if the prediction accuracy is good, with little effort trying to interpret the models with engineering knowledge. This approach has resulted in a few setbacks. First, the complex deep neural networks and their underlying mechanisms are difficult to interpret (Zhang et al. [2021]). Second, these deep

*Corresponding authors: S. Joe Qin and Xiangyu Zhao ({joe.qin; xianzhao}@cityu.edu.hk)

I Can't Believe It's Not Better Workshop at NeurIPS 2022.

neural models are a black-box function approximator (Chakraborty et al. [2017]), while engineers prefer interpretable solutions.

In addition, for engineering and industrial data where domain knowledge is abundant, it is not clear whether the black-box deep learning models can always outperform statistical learning models based on engineering-informed features from domain knowledge. The objective of this paper is to report a cooling load prediction solution for two real municipal buildings in Hong Kong in response to a recent global AI competition. We show that dynamic statistical learning models with engineered features outperform deep learning alternatives in both interpretability and prediction accuracy. The proposed solution for the global competition was finally conferred a Gold Award and a Grand Prize as the Microsoft Outstanding AI Influencer Award by the panel of internationally renowned experts. The knowledge-informed learning solution includes data preprocessing based on cooling operation knowledge, feature engineering from control system knowledge, and interpretable statistical learning algorithms to build the models. To search for the best model to predict the cooling load, we explored LSTM, gated recurrent units (GRU), and several other AutoML models as benchmarks.

The Global AI Challenge for Building E&M Facilities was launched in October 2021 by the Electrical and Mechanical Service Department (EMSD) of the Hong Kong SAR and Guangdong Provincial Association for Science and Technology (URL: https://www.globalaichallenge.com/en/home). It is a challenge to predict the cooling loads of two high-rise office buildings located in Kowloon, Hong Kong. We are given a large dataset collected over 1.5 years after the COVID pandemic with multiple external weather variables as input and the building cooling load consumption as output. After the pre-processing of the original data, we started the model training work. In the beginning, we highly expected to obtain superior results from deep neural networks to build the model. Initially deep neural networks did outperform other routine machine learning methods. However, with the progress of research when domain knowledge was accumulated and engineered features were incorporated, we find that the statistical learning model is the best after extensive trials of deep neural networks. Our proposed model was tested by the competition organizer on real operation data that had yet to be collected for the months of October to December 2021. Our model with engineered features stood out in terms of prediction accuracy as well as interpretability.

Specifically, we propose a dynamically engineered features with modes of operation learning (DEFMOL) model for cooling load predictions. Dynamic and interaction terms are included in the model. To the best of our knowledge, this method is the first effort made for building cooling load prediction. We give detailed results in this paper to show how our model outperforms deep learning models, which are considered to be superior in many previous studies.

2 Related Work

Several effective algorithms were applied for cooling load prediction with high accuracy in recent years. For instance, Roy et al. [2020] applied a deep neural network (DNN) to predict cooling load. Al-Rakhami et al. [2019] proposed an ensemble learning applying XGBoost to build an efficient prediction model. Sajjad et al. [2020] proposed a multi-output (MO) sequential learning model followed by utility preprocessing with a unified framework. Fan et al. [2019] proposed an efficient regression model based on sensitivity analysis and the traditional autoregressive with exogenous (ARX) model. Wang et al. [2020] proposed a twofold algorithm, which first used LSTM for short-term load prediction, then used XGBoost for long-term load prediction. In addition, Kwok et al. [2011] demonstrated that building occupancy area and rate play a critical role in cooling load prediction. The fresh air supply rate measured via Primary Air-handling Units (PAU) was used to indicate the CO_2 concentration which reflects the change in occupants' load.

As surveyed by Lu et al. [2021], LSTM has attracted the most attention among all RNNs with 42 journal articles (accounting for 48.93%). LSTM is a dominant DNN structure that outperforms other models in these 42 articles. Although previous studies provide various methods of cooling load predictions, the cooling load trend is different for each building. In this study, we devote sufficient effort to try deep learning and also search for a new effective method to tackle the task, namely, the dynamically engineered features with modes of operation learning method. Finally, the performance of the proposed DEFMOL learning model exceeds the performance of others.

3 Methodology

3.1 The DEFMOL Solution for Cooling Load Prediction

Figure 1 depicts the daily cooling load consumption traces of the workdays from Jul 1, 2021 to Sep 30, 2021. The daily traces can be divided into three modes, which are Off-mode (yellow), On-mode (red), and Shutting-off mode (green). The trend and consumption at different modes are very different. For the statistical learning method, we adopt a divide and conquer approach, in other words, we build a model for each mode.

Large prediction errors of the cooling load often occur at the start-up phase of the On-mode (red) and the Shutting-off mode (green). When the system turns on, the cooling load surges to a high point, then drops a little bit to a relatively stable working state. This surge is caused by turning on the electric motor of the compressor to respond to an increased cooling load at startup by feedback control of the room temperatures (Aswani et al. [2011]). This transition phase is basically the step response of the HVAC control system. When the system shuts off, there appear to be gradual stages in turning off the cooling. Pumps do not instantly switch to Off-mode due to the dynamics of heat pumps (Aswani et al. [2011]). Both the start-up and the shut-off processes are very similar to dynamic feedback control responses. Therefore, we propose to create engineered features to mimic the dynamic responses to step-like changes.

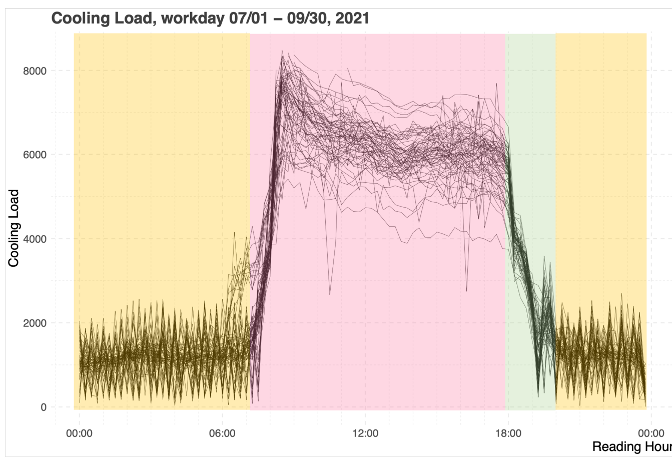

Figure 1: Normal workday cooling load trend

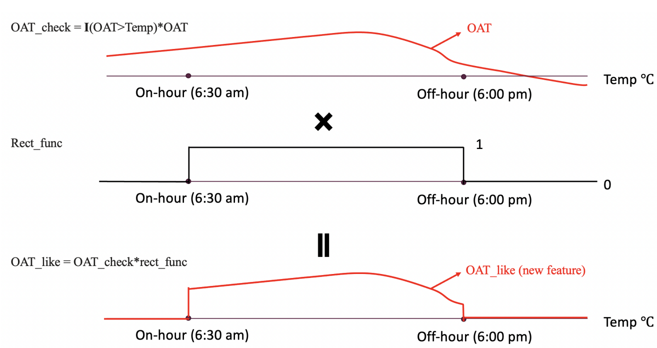

Figure 2: A schematic of the OAT_like feature

Two key innovative ideas of the DEFMOL model are the engineered dynamic features to capture changes in the outdoor ambient temperature (OAT) and their interaction with changes in the operation modes. We propose Algorithm 1 to engineer these features that augment the input dataset. The schematic of OAT_like feature is shown in Figure 2, which captures OAT and its interaction with the modes of on and off hours.

Let $\boldsymbol{x}_k = [x_1,\ x_2, \cdots, x_p]^\top \in \mathbb{R}^p$ be the augmented features including time-lagged variables and y_k be the response variable to be predicted, where p is the augmented input dimension. Statistical learning methods can be applied to estimate this model, such as the well-known ridge regression by Hoerl and Kennard [1970],

$$\hat{\boldsymbol{\beta}}_\lambda = \arg\min_{\boldsymbol{\beta}} \frac{1}{2N} \sum_{k=1}^{N} (y_k - \beta_0 - \boldsymbol{x}_k^\top \boldsymbol{\beta})^2 + \frac{1}{2}\lambda \|\boldsymbol{\beta}\|_2^2 \tag{1}$$

where λ is the hyperparameter to be optimized via cross validation, N is the number of observations, β_0 and β represent intercept and coefficients from ridge regression respectively. The model can be nonlinear but is linear-in-parameters. Other methods such as the Lasso can be adopted to select relevant variables.

Modes of operations: We divide all days into workdays and non-workdays (including weekends & holidays) and a workday into three modes. For On-mode and Shutting-off-mode, we need to distinguish between workdays and non-workdays and build separate models since they behave differently due to the people load. However, there is no need to distinguish between workdays and non-workdays for the Off mode, as the chiller is turned off for both modes. Therefore, we need to build five separate models for the all-time forecast. Further, the South Tower and North Tower each tower has its own equipment, hence, it is necessary to build a model for each tower separately. To sum up, we need to build ten small models in total. After we get the prediction of South Tower cooling load and North Tower cooling load, sum them up to get the whole cooling load prediction and calculate the root mean squared error (RMSE) to benchmark the performance, which is the official criterion announced by the competition.

Feature engineering: Besides the dynamically engineered features introduced in Algorithm 1, we did more actions on the original variables. Because the weather factors (e.g. outside temperature, and rainfall) have a lagging influence on the environment. We add 1-4 hours lag variables for each weather factor. Plus, because the working intensity is different on each weekday, weekend, and holiday, we add one-hot variables for each day-of-the-week and holiday_check (to check if it is a holiday).

Algorithm 1 Dynamically Engineered Features

Make the OAT dynamic signalized, and augment them into input
$On\_hour \leftarrow 6:30, Off\_hour \leftarrow 18:00$
$lags \leftarrow 40, temp\_bias \leftarrow 14°C, i \leftarrow 1$
Create $Rect\_func$
if time in $range(On\_hour, Off\_hour)$ **then**
    $Rect\_func = 1$
else
    $Rect\_func = 0$
end if
Create OAT_check
if $OAT \leq temp\_bias$ **then**
    $OAT\_check = OAT$
else
    $OAT\_check = 0$
end if
Build OAT_like
$OAT\_like = rect\_func * OAT\_check$
Build interaction terms between OAT_like and day-of-the-week
for Day in $range(Mon, Sun)$ **do**
    if Current day-of-the-week is Day **then**
        $OAT\_like\_Day = OAT\_like$
    else
        $OAT\_like\_Day = 0$
    end if
    Add OAT_like_Day as feature
end for
while $i \leq lags$ **do**
    Add $OAT\_like\_Day\_\{lag_i\}$ for Day in $range(Mon, Sun)$ as features
end while

3.2 Deep Neural Networks

We choose GRU by Chung et al. [2014] and LSTM to build dynamic recurrent models. Figure 3 depicts the architecture of these units. We use the same features used in DEFMOL as input for deep neural networks. As deep neural networks are highly non-linear, we do not divide the whole day into multiple modes. As in DEFMOL, we build separate models for the South Tower and North Tower, then sum the cooling load predictions to obtain the final output.

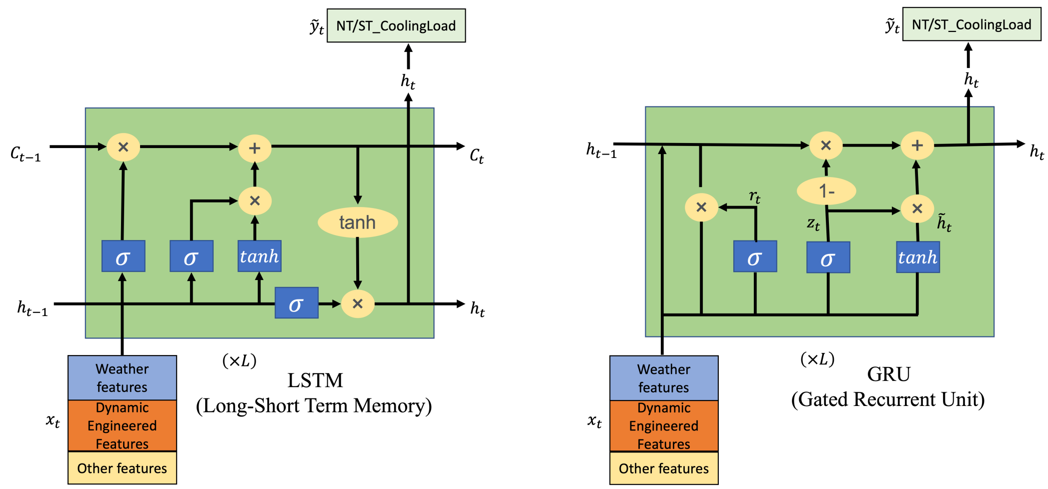

Figure 3: Architecture of LSTM and GRU for cooling load predictions

Table 1: Dataset overview

Variable Names	# Missing	# Abnormal
Average_OAT	4,395	0
Humidity	8,474	0
UV_Index	0	0
Average_Rainfall	0	0
NT_CoolingLoad	5,101	73
ST_CoolingLoad	5,337	643
CoolingLoad	6,222	5

4 Experiments and Evaluation

In this section, we applied methods presented in Section 3 on the dataset. We present results on data pre-processing, feature engineering, model hyper-parameter tuning, and prediction performance evaluations.

4.1 Dataset

We first describe the dataset from the /em Global AI Challenge For Building E&M Facilities released by Hong Kong EMSD. Table 1 describes the overview of the dataset. There are four initial variables and three outputs: NT-CoolingLoad, ST-CoolingLoad, and CoolingLoad (equal to NT-CoolingLoad + ST-CoolingLoad). The data were recorded every 15 minutes from Apr 1, 2020 to Sep 30, 2021, which has 52,608 samples overall. After pre-processing, we reserve Sep 2021 as the testing set and the rest of the data as the training set. We use the performance of the testing set to compare the trained models.

Data Pre-processing: The are some issues with the original data that need to be tackled. Due to peaks of the COVID-19, the people load in the commercial buildings was highly irregular due to working-from-home for a large part of 2020, which also caused missing values. In addition, there

Table 2: Performance of learning with feature engineering (hourly data)

Method	Training time/s	Training RMSE	Validation RMSE
DEFMOL	120	225	**264**
GRU	475	181	434
LSTM	569	**173**	401
CatBoost	200	307	426
LightGBMXT	17	306	429
WeightedEnsemble	300	294	440
RandomForest	163	338	454
NeuralNetFastAI	39	303	472
XGBoost	33	316	487

were some abnormal (some negative values appeared on Cooling Load) or NaN values appeared. As long as abnormal/NaN values appeared in Cooling Load (output), we choose to delete these samples of data. For NaN values that appeared in input but not in Cooling Load, we use interpolation to fill in the missing records to make full use of existing data.

4.2 Overall Performance

Table 2 shows the predicted RMSEs of the testing dataset from the DEFMOL, GRU, and LSTM models. The optimal hyperparameters of LSTM and GRU are given in Appendix. Figure 4 shows the detailed prediction performance of each method on testing set. There are missing values in the morning of Sep. 24, 2021, which have been deleted from the dataset. We compare the results of three models: (1) **DEFMOL:** This model earns the best performance on the testing set. From Figure 4 it is seen that the DEFMOL model obtains much more accurate predictions in the on-hour mode than the alternatives. Moreover, this model takes less running time than LSTM and GRU, which is not only effective but also efficient. (2) **GRU:** GRU takes more time to train and has better training error than the statistical learning method. (3) **LSTM:** LSTM has better performance on training and testing sets than GRU, but the performances of LSTM and GRU on the testing dataset are far worse than the DEFMOL model. For the sake of comparison, mean absolute errors (MAE) on the testing set are also compared for these models, which are 190, 265, and 243 for DEFMOL, GRU, and LSTM, respectively. It is seen that the proposed DEFMOL model is the clear winner.

To search for better deep learning algorithms for the challenge problem, we applied CatBoost, LightGBMXT, WeightedEnsemble, RandomForest, NeuralNetFastAI, and XGBoost provided in Erickson et al. [2020] to the dataset with the same input and output settings. The training and testing results are also shown in Table 2. It is observed that all these alternatives perform far worse than the proposed DEFMOL model in terms of both training and testing RMSEs.

5 Conclusions

In this paper, we present the solution of a dynamically engineered features with modes of operation learning, or DEFMOL, for the cooling load predictions of two high-rise office buildings in Hong Kong as a response to the Global AI Challenge organized by the Hong Kong government. The DEFMOL solution includes data preprocessing based on cooling operation knowledge, feature engineering from control system knowledge, and interpretable learning algorithms to build the dynamic model. It is shown that DEFMOL models with mode-dependent features outperform deep learning alternatives including the optimized LSTM and GRU. The engineered OAT-like feature and modes of operations play a critical role in DEFMOL. In addition, the DEFMOL model is interpretable. Future work is planned to generalize the results on other similar datasets of commercial buildings.

Acknowledgements

The authors are grateful to the Electrical and Mechanical Service Department (EMSD) of the Hong Kong SAR for providing the data and the conferred Grand Prize by Microsoft.

References

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

Yu Zhang, Peter Tiňo, Aleš Leonardis, and Ke Tang. A survey on neural network interpretability. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021.

Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M Rao, et al. Interpretability of deep learning models: A survey of results. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, pages 1–6. IEEE, 2017.

Sanjiban Sekhar Roy, Pijush Samui, Ishan Nagtode, Hemant Jain, Vishal Shivaramakrishnan, and Behnam Mohammadi-Ivatloo. Forecasting heating and cooling loads of buildings: A comparative performance analysis. *Journal of Ambient Intelligence and Humanized Computing*, 11(3):1253–1264, 2020.

Mabrook Al-Rakhami, Abdu Gumaei, Ahmed Alsanad, Atif Alamri, and Mohammad Mehedi Hassan. An ensemble learning approach for accurate energy load prediction in residential buildings. *IEEE Access*, 7:48328–48338, 2019.

Muhammad Sajjad, Samee Ullah Khan, Noman Khan, Ijaz Ul Haq, Amin Ullah, Mi Young Lee, and Sung Wook Baik. Towards efficient building designing: Heating and cooling load prediction via multi-output model. *Sensors*, 20(22):6419, 2020.

Chengliang Fan, Yundan Liao, and Yunfei Ding. Development of a cooling load prediction model for air-conditioning system control of office buildings. *International Journal of Low-Carbon Technologies*, 14:70–75, 01 2019. doi: 10.1093/ijlct/cty057.

Zhe Wang, Tianzhen Hong, and Mary Ann Piette. Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, 263:114683, 2020. ISSN 0306-2619.

Simon SK Kwok, Richard KK Yuen, and Eric WM Lee. An intelligent approach to assessing the effect of building occupancy on building cooling load prediction. *Building and Environment*, 46 (8):1681–1690, 2011.

Chujie Lu, Sihui Li, and Zhengjun Lu. Building energy prediction using artificial neural networks: A literature survey. *Energy and Buildings*, page 111718, 2021.

Anil Aswani, Neal Master, Jay Taneja, David Culler, and Claire Tomlin. Reducing transient and steady state electricity consumption in hvac using learning-based model-predictive control. *Proceedings of the IEEE*, 100(1):240–253, 2011.

Arthur Hoerl and Robert Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 04 1970.

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*, 2014.

Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate AutoML for structured data. *arXiv preprint arXiv:2003.06505*, 2020.

Appendix

Hyper-parameter tuning

All three models have hyper-parameters to be optimized. For the proposed DEFMOL, we set the range of λ as $\exp(seq(-8, 2, 50))$, where $seq(-8, 2, 50)$ means to pick 50 numbers from -8 to 2 with equal intervals. For GRU and LSTM, we pick layers from (2,3,4), pick hidden size from (128, 256, 512) and pick dropout from seq(0,1,10). Then we randomly split 10% from the training set as the validation set to optimized the hyper-parameters. Table 3 shows the selected hyper-parameters of GRU and LSTM.

Table 3: GRU and LSTM Hyperparameters

Method	GRU	LSTM
Learning Rate	0.01	0.01
Batch Size	500	500
Weight Decay	1e-6	1e-6
Epoch	50	50
Number of Layers	3	3
Hidden Size	128	256
Dropout	0.3	0.2

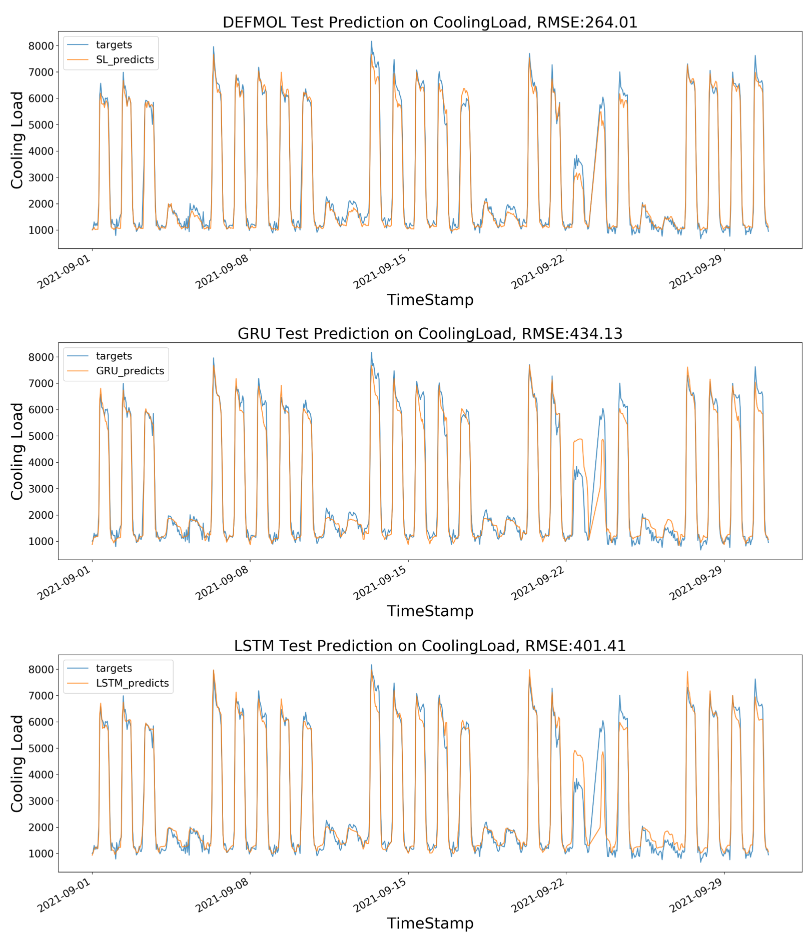

Figure 4: Prediction performance on the testing dataset using DEFMOL (Top panel), GRU (Middle panel), and LSTM (Bottom panel). DEFMOL outperform others by at least 51% in RMSE.