Minimalistic LSTM Models for Next Day Hourly Residential HVAC Energy Usage Forecasting

Rahman Heidarykiany
Electrical and Computer Engineering Dept.
Marquette University
Milwaukee, WI, USA
rahman.heidarykiany@marquette.edu

Cristinel Ababei
Electrical and Computer Engineering Dept.
Marquette Universty
Milwaukee, WI, USA
cristinel.ababei@marquette.edu

***Abstract*—Accurate electrical energy demand forecasting is essential to optimization and operation approaches aimed at reducing the cost of electricity at the consumer and utility level. Machine learning models, such as long short-term memory (LSTM) models, have been increasingly employed in energy usage prediction for different time-horizons into the future. The prediction accuracy of such models depends a multitude of model architecture and model training parameters, that are often left at their default values or the strategy of selecting them is not even reported. In this paper, we present a thorough investigation of the impact of fifteen different such parameters on the performance of LSTM models used to forecast HVAC energy usage in typical residential homes for 24 hours. The objective is to arrive at a select number of practical LSTM models, which are trained and tested on data generated from the equivalent of a 21 year long simulation of a testbed based on the IEEE 13 node test feeder. Our investigation reveals several remarkable characteristics that the highest ranked in terms of prediction accuracy LSTM models have in common: models can use as few as two layers, training should use more equivalent years of data available, batch size should include 24 days of data, and the best optimizer used during training is RMSprop.**

***Index Terms*—Minimalistic LSTM model, prediction, residential HVAC energy usage, hyperparameter sensitivity analysis, model size on disk**

I. Introduction

Energy usage forecasting has attracted a lot of interest not only for planning purposes, but, also for developing predictive optimization and operation techniques, both at grid and house levels. Among these forecasting methods, particularly those that employ machine learning (ML) models and techniques have become very popular, as discussed in several recent survey papers [1]–[6]. Specifically, recurrent neural networks (RNNs) models got significant attention in building energy forecasting because they were found to provide better performance [7]. Among the previous RNN models studied for sequence-to-sequence (seq2seq) learning, two particular models have been particularly popular: long short term memory (LSTM) and gated recurrent unit (GRU) models.

Previous studies show that despite the fact that they require longer training times, LSTM models can provide better prediction than GRU models in the context of energy forecasting for the smart grid [7]. As such, one can find significantly more studies investigating LSTMs in previous literature [8]. However, only a relatively small number of previous studies focused on the HVAC energy usage, despite the fact that HVAC represents the major component in the energy usage of a residential house, accounting for more than a third of the total energy usage [9]. Therefore, in this paper, we focus on the LSTM models used to forecast HVAC energy usage for 24 hour future horizons. More specifically, we are interested in: 1) investigating how the multitude of parameters (learning rate, momentum, activation functions, number of layers, epoch, learning algorithms, etc.) that one can configure during LSTM model definition and training affect the prediction accuracy; which ones are the most important and affect model performance the most, and 2) identifying the smallest LSTM models that would require shorter training times for very large datasets while still offering good prediction accuracy.

II. Literature Review

In this section, we present a discussion of related recent studies that focused on the use of different types of RNN models for prediction of power or energy usage in residential homes. A seq2seq RNN model was presented in [10] for one-hour horizon forecasting of residential HVAC loads. The study in [11] presented a gated RNN-Conv model and reported an accuracy of 8.99% mean absolute percentage error (MAPE) for day-ahead forecasting. The study in [12] used an LSTM model to forecast four-day residential total loads. The studies in [13]–[15] used LSTM models for 2 day (48 points) prediction of residential aggregated load prediction for 150 households. Day ahead (i.e., 24 h horizon) prediction was investigated in the studies from [16]–[21] for daily university air-conditioning energy usage, residential building load, multi-building energy usage, energy usage of office buildings, and building cooling load prediction. The study in [22] investigated prediction for hourly (24 h, 6 h, 1 h) and daily (7 days, 1 day) prediction horizons of single house power load. The study in [23] focused on prediction of one hour load in smart buildings.

An interesting line of research is the investigation of more sophisticated RNN/LSTM or LSTM based hybrid models. For example, the study in [24] used an LSTM encoder-decoder (LSTM Enc-Dec) model to forecast seven-day heating and cooling energy use of office buildings. Other recent studies that reported LSTM Enc-Dec models include [25]–[28]. Seven-day load forecasting using LSTM deep neural network (LSTM-

DNN) models for the heating, ventilation, and air-conditioning (HVAC) systems was presented in [29]. Other studies [30]–[38] investigated LSTM-DNN, LSTM domain adversarial neural network (LSTM-DANN), RNN-LSTM, convolutional neural network and LSTM (CNN-LSTM), k means CNN and LSTM (KCNN-LSTM), GRU+RNN, Deep RNN+GRU (DRNN-GRU), and CNN-GRU models.

Most of previous studies reported MAPE and/or root mean square error (RMSE) values to quantify the performance of their prediction models. However, it is difficult to directly use these metrics to compare the performance of previously reported LSTM models because their values significantly depend on the particular dataset that was used for training as well as on whether data normalization was used or not among other factors [22]. Another challenge related to LSTM model development is that it is generally unclear how model architecture *parameters* and model training *hyperparameters* affect model performance. Most previous works usually leave this challenge unaddressed without providing explanations of how model architecture was decided or fine-tuned, and at best, some discuss standard hyperparameter selection and optimization but for a rather small number of hyperparameters.

Therefore, in this paper our contribution is a thorough investigation of how fine-tuning of model architecture parameters and model training hyperparameters - collectively referred to as *parameters* in the remainder of this paper - affects prediction performance of LSTM models used in HVAC energy use forecasting. This sensitivity analysis provides insight into what are the most important such parameters, which can be fine-tuned to develop minimalistic LSTM models that can still provide satisfactory prediction accuracies. Using extensive simulations with a modified GridLab-D simulator, we generate training data spanning the equivalent of 21 years and use that to train LSTM models during the parameter fine-tuning process. This process leads us to the best minimalistic LSTM models for all 15 houses from the testbed constructed with the IEEE 13 bus system. Simulation results demonstrate that the identified models can provide better predictions as verified via MAPE, RMSE, and CV-RMSE (coefficient of variation of root mean square error) metrics when compared to a recently reported LSTM model.

III. Dataset Generation

We use a custom simulation tool based on GridLab-D simulation framework [39]. To generate training datasets, we have modified the IEEE 13 node testcase and attached 15 houses to the 7 buses as illustrated in Fig. 1. For diversity, all 15 homes are defined inside GridLab-D to have different total areas while their models include HVAC, water heater, lighting, and three other appliances. Particularly, the HVAC load is simulated as a multi-state load dependent on the air temperature and triggered at different temperature bands [40]. For simplicity, all setpoints for all houses's HVAC are set to the same value. Simulation of the entire testcase is done for a total period of 21 years to generate datasets for each of the 15 houses. The weather data for the simulation is from Yakima, Washington available in GridLab-D. Dataset sizes that are investigated include 2, 5, 10, and 21 years out of the total 21 year of simulation data.

Table I
Specific values for each of the 15 variable parameters.

Variable Parameter	Specific Values Explored
Learning rate	0.00025, 0.0005, 0.001, 0.002, 0.004
Number epochs	800, 1000
Batch size	12, 24, 48, 72, 96, 120, 144
Optimzer	RMSprop, Adam, Adamax, Nadam
Activation	relu, sigmoid, softmax, tanh, selu, elu, hardSigmoid
Recurrent activation	relu, sigmoid, softmax, tanh, elu, hardSigmoid
Number LSTM layers	1, 2, 3
Number units of LSTM layers	48, 72, 96
Loss function	mse, mae
Number dropout layers	0,1, 2
Value recurrent dropout	0.2, 0.4, 0.5, 0.6
Value dropout	0.2, 0.4, 0.5, 0.6
Number dense layers	1, 2, 3
Number output units of dense layers	24, 48, 72, 96
Length of dataset in equivalent years	2, 5, 10, 21

Each hourly entry in these datasets includes: timestamp, HVAC energy load, entire house energy load, heating setpoint, mass temperature, inside air temperature, outdoor temperature, system mode, and thermostat deadband. Therefore, the dataset is multivariate; however, for simplicity, in this paper we focus only on multi-step univariate input and multi-step univariate output, with the variable of interest being the HVAC energy usage. The HVAC load profiles inside GridLab-D are based on equivalent thermal parameter (ETP) models that can generate multi-state time-varying load models [40]. This model uses more than one state to describe the energy usage of the HVAC. Each state is governed by a physical model, with transitions between states determined by either internal state transition rules or external signals. For example, the HVAC State 1 represents no power draw, HVAC is off; in States 2 and 3, an electric fan motor and a compressor motor are included. Like State 3, State 4 provides heating with an associated electric fan plus heating provided by resistive heating elements. In this model, HVAC energy use depends on outdoor temperature, heating setpoint, cooling setpoint, mass temperature, inside air temperature, thermostat deadband, aux heat deadband, heating coefficient of performance (COP), cooling COP, and aux heat temperature. As examples, Fig. 2 shows the HVAC load profiles for House #1 on the 16th of February and August.

IV. Proposed Approach for Parameters Fine-tuning

We are investigating fifteen (15) variable parameters, including model architecture parameters, training hyperparameters, and training dataset size as listed in Table I. Fig. 3 illustrates two of these variables occurring inside a typical LSTM cell. These 15 parameters define a vast search space, which we propose to explore using a simple heuristic algorithm. To reduce complexity, the idea is to split the parameters into two subgroups for simplicity, then, conduct searches in the spaces defined by those subgroup variable parameters. This is a divide-and-conquer approach in order to keep the execution time of the proposed search to reasonable practical values. A

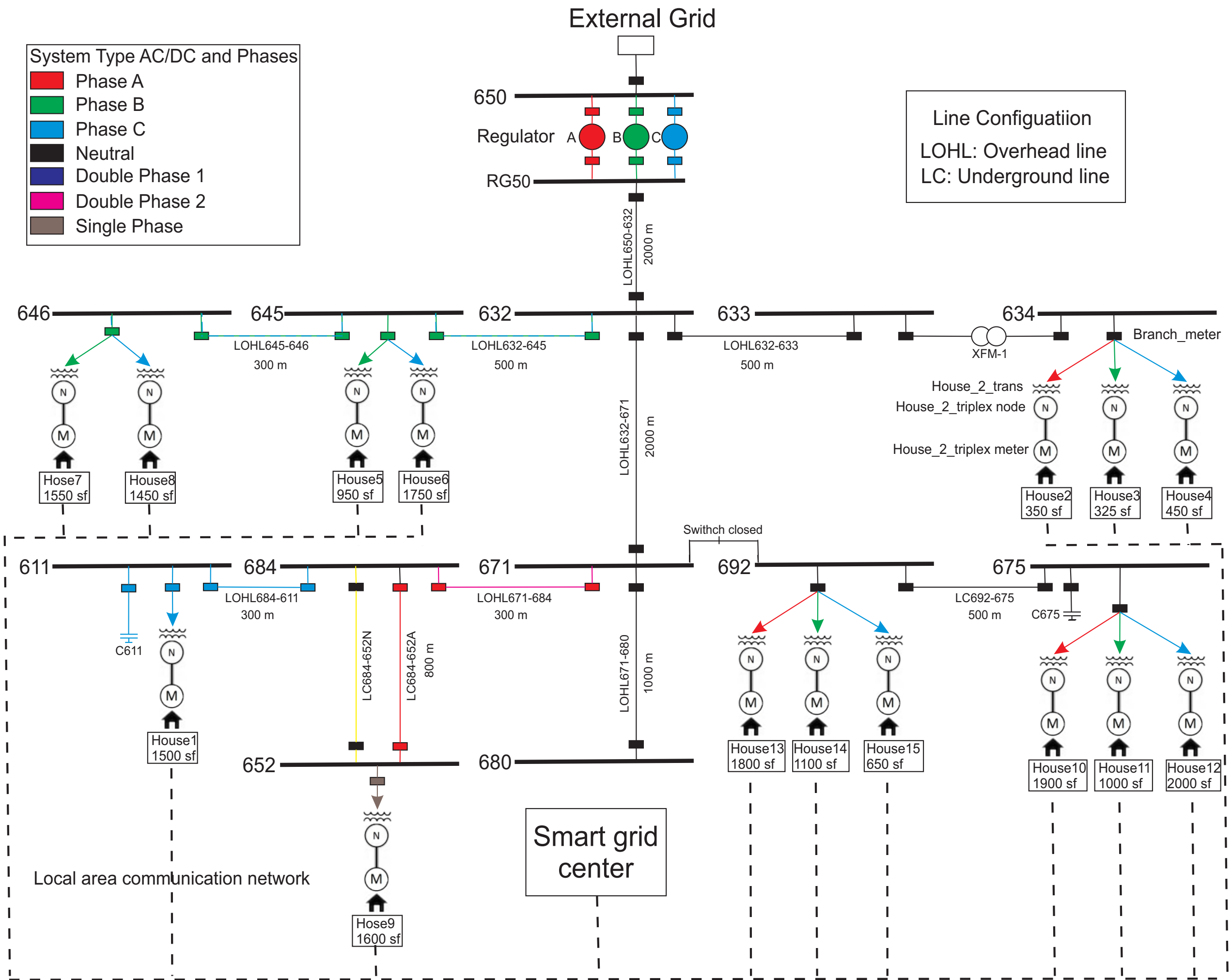
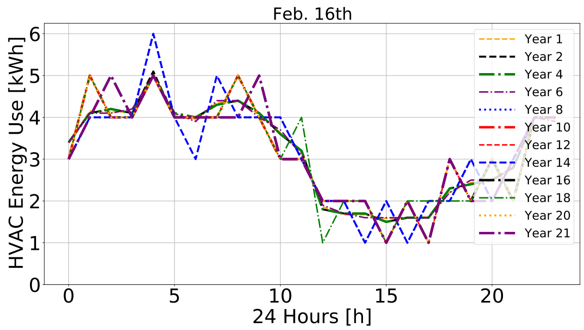


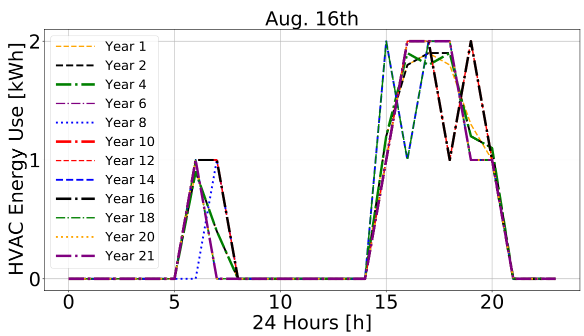
Figure 1. Diagram of the modified IEEE 13 node testcase used in simulations with modified GridLab-D to generate up to the equivalent of 21 years long datasest for each of the 15 houses.

simplified pseudocode description of the search is presented in Fig. 4. The parameters in the first subgroup include: learning rate, number of epochs, batch size, optimizer, activation function, and recurrent activation function. The remaining parameters in the second subgroup include: number LSTM layers, number units of LSTM layers, loss function, number of dropout layers, value recurrent dropout, value dropout, number of dense layers, number of output units of dense layers, and number of years in dataset. Specific values for all parameters that are explored in this search are listed in Table I.

As illustrated in Fig. 4, in the first step of the fine-tuning, an exhaustive search explores the space defined by the parameters in the first subgroup while the parameters in the second subgroup are kept constant at typical values. The top best models (e.g., 30 models - 2 for each house - but this number could be modified by the user) from this first step are recorded and stored. Sorting of best models is done based on the value of CV-RMSE. The best models are those whose CV-RMSE values are the smallest. In the second step, the parameters from the first subgroup are kept constant while the variables in the second subgroup are explored. Again, a constant number of the best models found during this step are recorded and stored. In the third step, a combination of all the previously recorded models are evaluated and ranked using MAPE, RMSE, and CV-RMSE. The best LSTM model is then selected. Best means that we look first at CV-RMSE and select the model with the lowest CV-RMSE value. If two models have the same CV-RMSE value, then, we look at MAPE for deciding about the best model. If MAPE is not available (i.e., values are nan), we finally use RMSE and select the model with the lowest RMSE value. This search process is repeated for each of the houses in the testcase used for dataset generation, which will be described in the next section.



(a)



(b)

Figure 2. HVAC load profiles of House #1 during the 16th of February and August for all years of the generated dataset, using an instrumented GridLab-D simulator.

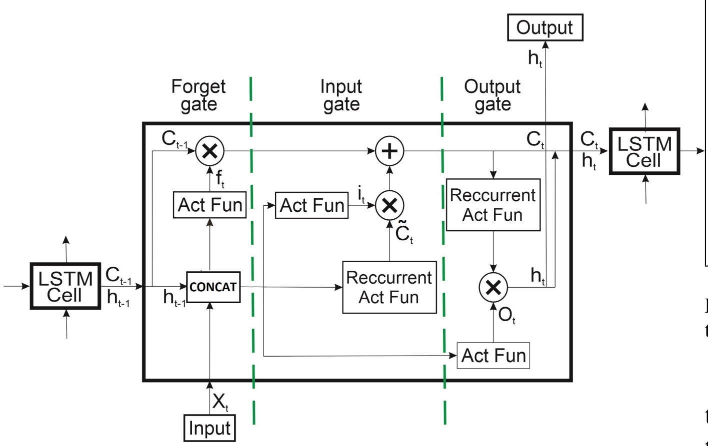


Figure 3. Simplified diagram of the LSTM cell illustrating the activation and recurrent activation function parameters.

V. Sensitivity Analysis

In this section, we conduct a parameter sensitivity analysis to identify which are the most important parameters. During the simulations in this experiment, we use early stopping to avoid overfitting during training and to save simulation time; early stopping was done based on validation loss values and on *patience* (i.e., number of epochs with no improvement after which training is stopped) value of 15% of the total number of epochs. The following discussion is based on looking at

Algorithm: Heuristic for parameter fine-tuning

```
Step 1: Record top best LSTM models from exhaustive
search in space defined by 6 variables in first subgroup
for lr ← 0.00025 to 0.004 do
  for epoch ← 800 to 1000 do
    for batch size ← 12 to 144 do
      for optimizer ← RMSprop to Nadam do
        for activation ← relu to hardSigmoid do
          for rec activation ← relu to hardSigmoid do
            Remaining parameters kept constant
            Train, test model, calculate error metrics
            Record and store top LSTM models
          end for
        end for
      end for
    end for
  end for
end for
Step 2: Record top best LSTM models from exhaustive
search in space defined by 9 variables in second subgroup
for num. LSTM layers ← 1 to 3 do
  for output units LSTM layers ← 48 to 96 do
    for loss function ← mse to mae do
      for num. dropout layers ← 1 to 2 do
        for rec. dropout ← 0.2 to 0.6 do
          for dropout ← 0.2 to 0.6 do
            for num. dense layers ← 1 to 3 do
              for output units dense layers ← 24 to 96 do
                for length dataset ← 2 to 21 do
                  Remaining parameters kept constant
                  Train, test model, calculate error metrics
                  Record and store top LSTM models
                end for
              end for
            end for
          end for
        end for
      end for
    end for
  end for
end for
Step 3: Re-evaluate all recorded models
Retain best LSTM model for each given house
```

Figure 4. Pseudocode description of the three-step heuristic search to identify the best LSTM model.

the top best 16 LSTM models from among the collection of top 3 models for all 15 houses. The number 16 was chosen arbitrarily. We report in Table II the main findings of this investigation and make the following observations:

- A learning rate of 0.001 was the most likely to help in identifying the best LSTM models.
- Despite using early stopping to stop the training once models' loss starts to increase, a number of 1000 epochs lead to the best models (including the top 2). Only 12.5% of models were obtained when training used 800 epochs.
- The batch size of 24 was used during training of 15 out of 16 best models.
- Contrary to the popular previous approach of using the Adam optimizer, all 16 best models were obtained using RMSprop optimizer.

Table II
RESULTS BASED ON THE TOP 16 BEST LSTM MODELS.

Variable Parameter	Sensitivity Analysis Result
Learning rate (lr)	50% of models obtained with $lr = 0.001$; 25% obtained with $lr = 0.002$
Number epochs	87.5% of models obtained with 1000 epochs
Batch size	batch size of 24 led to 15 out of 16 best models
Optimzer	RMSprop was the optimizer that led to all best 16 models
Activation	Top two models used relu and tanh
Recurrent activation	tanh used by 75% of best 16 models
Number LSTM layers	Two layer LSTM networks are found to be best
Number units of LSTM layers	72 units used by 75% of best 16 models
Loss function	mse was used for identifying all best models
Number dropout layers	1 dropout layer used by all best models
Value recurrent dropout	0.4 and 0 impact equally the fine-tuning
Value dropout	0.2 used by two thirds of best models
Number dense layers	2 dense layers used by all
Number output units of dense layers	Top 2 models used 24 and 96 dense output units
Length of dataset in equivalent years	21 equivalent years lead to best models

- No activation function seems to be used by the majority of the best 16 models. relu and tanh are used by the top 2 models; however, selu, elu, and sigmoid functions helped achieve good models too.
- 12 out of 16 best models use tanh as recurrent activation functions. The remaining models used elu, relu.
- All best models had two layers.
- 75% of the 16 best models were fine-tuned to use 72 units for LSTM layers, while 25% of the models used 96 units.
- The mse loss function was used in the training of all best 16 models.
- One dropout layer was used by all 16 best models.
- 50% models were found with a value of 0.4 for recurrent dropout, while the remaining 50% of models were found with a value of 0.0.
- 62.5% of models used a value of 0.2 for dropout, while 25% of models used a value of 0.4 and only 12.5% of models used a value of 0.5. The top 2 models used 0.2 and 0.5 dropout values.
- Two dense layers were used by all best models.
- 37.5% of models use 96 output units, 25% of models use 24 output units, 18.75% of models use 72 output units, and 12.5% of models used 48 output units. The top 2 models were identified for 24 and 96 dense output units.
- As expected, the longest datasets (i.e., the equivalent of 21 years) resulted in 15% better prediction in contrast with 10 year long datasets.

Based on the sensitivity, we observe that the length of the datasets had the largest impact on the performance of models; 21 year long datasets helped arrive to the best models, confirming that larger datasets are desirable for model development. Next in importance are number of epochs, recurrent activation function, learning rate, and batch size.

VI. PERFORMANCE OF IDENTIFIED MINIMAL LSTM MODELS

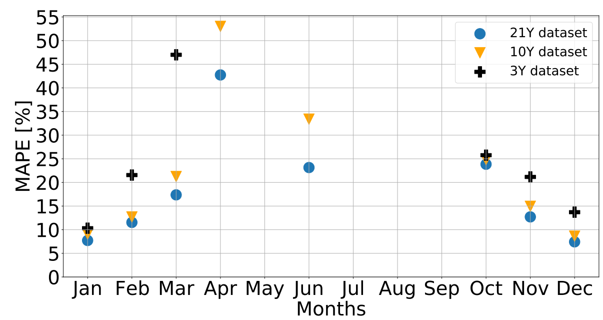
Once the minimal LSTM models for each house are identified as described in section IV, they are tested using 30% of the datasets generated using the procedure from section III (the other 70% of each dataset is used for training and validation). The performance of the best LSTM models is measured using three performance metrics (i.e., MAPE, RMSE, and CV-RMSE) and the best LSTM model is selected for each house. MAPE, RMSE, and CV-RMSE values are calculated and reported monthly, i.e., looking at each 24 h prediction in contrast with the 24 h expected value separately for each day of a given month of the year. Table III reports the average values of the four metrics over all 15 houses for each month. Important to note is the column with CV-RMSE values, which for the most part are less than 30%, which indicates an acceptable prediction accuracy [41]–[43]. Specific examples of MAPE, RMSE, and CV-RMSE plots for House #1 for three different dataset lengths are shown in Fig. 5. We note that the larger the dataset, the smaller the values of these metrics, which indicate better models. Also, specific examples of actual daily predictions using the developed LSTM models are shown in Fig. 6 for House #1 on the 16th of February and August of year 21.

VII. COMPARISON WITH STATE-OF-THE-ART MODELS

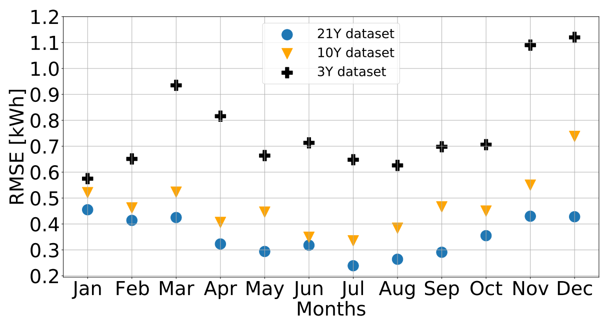
In this section, we present a comparison of the developed minimalistic LSTM models to a previous recent work that also investigated LSTM models [22], but in the context of predicting the entire house energy usage. For this comparison, we use the SHINES dataset [44], which was also used by [22]. The SHINES dataset was also split into 70% for training and validation and 30% for testing. We obtained the implementation of the LSTM models from [22], and used it in our comparison for prediction of HVAC energy usage instead, which is the focus of our paper. To make the comparison fair, both LSTM models were trained for the same umber of epochs (i.e., 150) using only three years (used by [22]) worth of data.

Table III
AVERAGE OVER ALL 15 HOUSES OF MAPE, RMSE, AND CV-RMSE RESULTS.

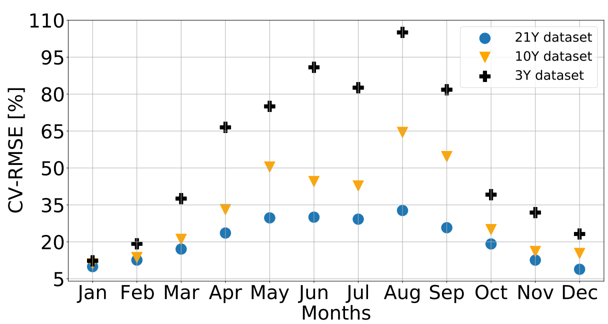
Month	MAPE (%)	RMSE (kWh)	CV-RMSE(%)
January	13.8	0.31	10.16
February	22.3	0.32	17.43
March	33.68	0.30	24.04
April	nan	0.25	30.82
May	nan	0.21	36.1
June	50.78	0.21	43
July	nan	0.21	30.1
August	nan	0.18	48.1
September	nan	0.22	35.02
October	nan	0.26	28.74
November	21.01	0.32	16.61
December	14.23	0.31	9.99



(a)

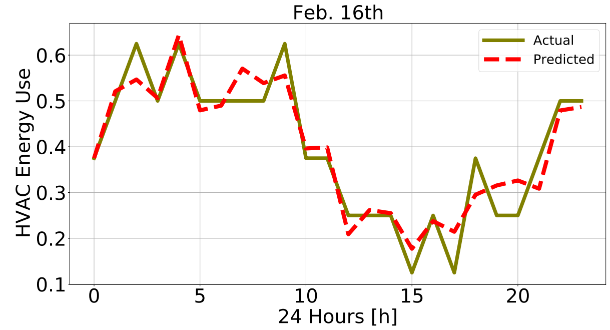


(b)

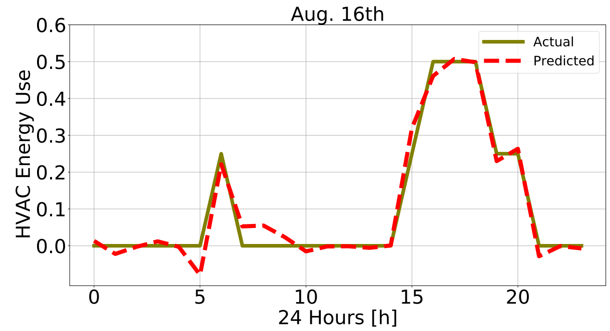


(c)

Figure 5. Examples of MAPE, RMSE, and CV-RMSE results for House #1 for three different dataset lengths.



(a)



(b)

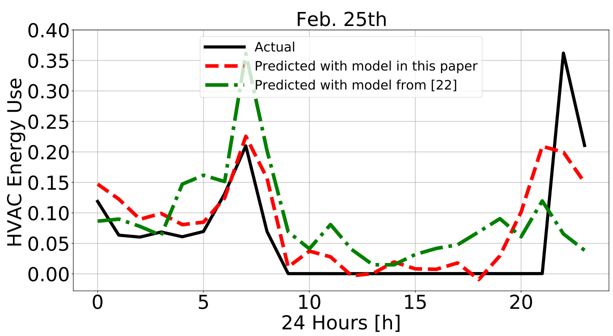
Figure 6. Examples of normalized HVAC load predictions for House #1 in year 21 of the dataset.

Our investigation focuses on 24 h future horizon prediction with 24 h input because this is the main focus of our models in this paper. The results of this testing are summarized in Table IV, where we report MAPE, RMSE, and CV-RMSE for House #1. First, we observe that looking again at the CV-RMSE column, we can see that the performance of our models is not as good as in the case of the 21 equivalent years dataset discussed before; more values in this column are higher than 30%. We attribute this to the fact that the SHINES dataset is shorter (only 3 years) and that the HVAC profiles exhibit increased variations. This is further aggravated by the fact that we use different models for different months, and this requires us to split datasets into 12 months, which in turn makes the training data for each month even smaller.

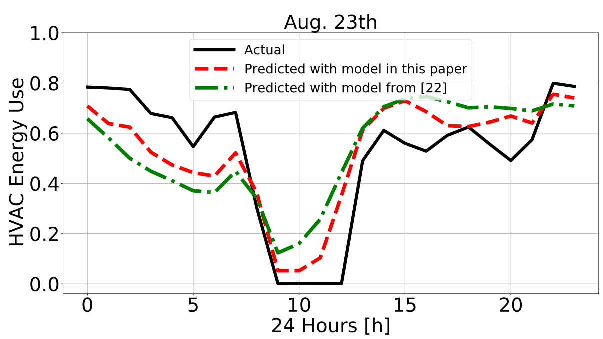
Second, we observe that the LSTM model developed in this work performs better (14% in terms of CV-RMSE) than the LSTM model from [22]. For a more detailed comparison of both models, in addition to the average values of MAPE, RMSE, and CV-RMSE, we report in Table III the model size and the memory footprint of the model (when saved on disk). We observe that our models have a very small memory footprint. One advantage of smaller trained models is that they can be more efficient and thus more suitable for realtime predictions and optimizations. Finally, examples of specific plots obtained with both models for selected days are shown in Fig. 7.

Table IV
TESTING OF DEVELOPED LSTM MODEL USING THE SHINES DATASET. COMPARISON AGAINST LSTM MODEL FROM [22].

	This work	[22]	This work	[22]	This work	[22]
Month	**MAPE (%)**	**MAPE (%)**	**RMSE (kWh)**	**RMSE (kWh)**	**CV-RMSE (%)**	**CV-RMSE (%)**
January	967	1317	1.72	1.69	88	87
February	1193	1514	1.05	1.09	83	87
March	368	534	2.77	4.15	130	218
April	903	1132	2.73	3.54	89	115
May	346	348	2.38	2.58	44	48
June	311	358	2.22	2.81	25	33
July	20	22	1.65	2.04	18	23
August	1075	1408	3.15	3.54	41	45
September	1048	1025	2.26	2.58	30	34
October	2736	2364	3.57	3.59	74	76
November	1765	1791	1.69	1.80	132	139
December	913	940	2.03	1.99	154	153



(a)



(b)

Figure 7. Examples of normalized HVAC load predictions for House #1 using SHINES dataset; predictions for Feb 2020 and Aug 2019.

Table V
QUALITATIVE COMPARISON OF DEVELOPED LSTM MODELS WITH PRIOR WORK USING THE SHINES DATASET .

Metric	Model in this work	Model from [22]	Difference (%)
MAPE (%)	970	1063	9
RMSE (kWh)	2.28	2.61	12.5
CV-RMSE (%)	75.66	88.1	14
Model Trainable Params	38788	23278	66.5
Memory Footprint (KB)	335	331	1.5

VIII. CONCLUSION

We presented an investigation of the impact of fifteen different model architecture and model training parameters on the performance and memory footprint of LSTM models used to forecast HVAC energy usage in typical residential homes for 24 hours. The objective was to identify minimal practical LSTM models, which could offer satisfactory performance with the benefits of being small and providing super-fast inference times. Testing of the best LSTM models developed during this investigation against a similar state-of-the-art LSTM model was done using the publicly available SHINES dataset demonstrated that minimalistic LSTM models that are more rigorously refined can offer better performance.

REFERENCES

[1] S. Aslam, H. Herodotou, N. Javaid, N. Ashraf, and S. Aslam, "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," *Elsevier. on Renewable and Sustainable Energy Reviews*, vol. 144, Jul. 2021.

[2] K. Alanne and S. Sierla, "An overview of machine learning applications for smart buildings," *Elsevier. on Sustainable Cities and Society*, vol. 76, Jan. 2022.

[3] D. Mariano-Hernández, F.S. Hernández-Callejo, S. Duque-Perez, and A.L. Zorita-Lamadrid, "A review of energy consumption forecasting in smart buildings: methods, input variables, forecasting horizon and metrics," *Applied Sciences. on Multidisciplinary Digital Publishing Institute*, vol. 10, no. 23, Jan. 2020.

[4] X. Yang, L. Shuli, Y. Zou, W. Ji, Q. Zhang, A. Ahmed, X. Han, Y. Shen, and S. Zhang, "Energy-saving potential prediction models for large-scale building: a state-of-the-art review," *Elsevier. on Renewable and Sustainable Energy Reviews*, vol. 156, Mar. 2022.

[5] J. Devaraj, R. Madurai Elavarasan, G. Shafiullah, T. Jamal, and I. Khan, "A holistic review on energy forecasting using big data and deep learning models," *Wiley Online Library. on International Journal of Energy Research*, vol. 45, no. 9, pp. 13489-13530, Jul. 2021.

[6] A. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: a comprehensive review," *Wiley Online Library. on International Journal of Energy Research*, vol. 44, no. 9, pp. 7136-7157, Jul. 2020.

[7] M. Massaoudi, H. Abu-Rub, S.S. Refaat, I. Chihi, and F.S. Oueslati, "Deep learning in smart grid technology: a review of recent advancements and future prospects," *IEEE Access*, vol. 9, pp. 54558-54578, Apr. 2020.

[8] J. Runge and R. Zmeureanu, "A review of deep learning techniques for forecasting energy use in buildings," *Energies. on Multidisciplinary Digital Publishing Institute*, vol. 14, no. 3, Jan. 2021.

[9] Department of Energy, "Chapter 5 — Increasing Efficiency of Buildings Systems and Technologies," [Online]. Available: *https://www.energy.gov/downloads/chapter-5-increasing-efficiency-buildings-systems-and-technologies*, 2015.

[10] A. Rahman, V. Srikumar, and A.D. Smith, “Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,” *Elsevier. on Applied energy*, vol. 212, pp. 372-385, Feb. 2018.

[11] M. Cai, M. Pipattanasomporn, and S. Rahman, “Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques,” *Elsevier. on Applied energy*, vol. 236, pp. 1078-1088, Feb. 2019.

[12] X. Lin, H. Yu, M. Wang, C. Li, Z. Wang, and Y. Tang, “Electricity consumption forecast of high-rise office buildings based on the long short-term memory method,” *Energies. on Multidisciplinary Digital Publishing Institute*, vol. 14, no. 16, Jan. 2021.

[13] L. Sehovac and K. Grolinger, “Deep learning for load forecasting: sequence to sequence recurrent neural networks with attention,” *IEEE Access*, vol. 8, pp. 36411-36426, Feb. 2020.

[14] T. Hou, R. Fang, J. Tang, G. Ge, D. Yang, J. Liu, and W. Zhang, “A novel short-Term residential electric load forecasting method based on adaptive load aggregation and deep learning algorithms,” *Energies. on Multidisciplinary Digital Publishing Institute*, vol. 14, no. 22, Jan. 2021.

[15] Z. Jiantao and J. Huang, “Application and research of short-term Power load forecasting based on LSTM,” *IEEE Int. Information Science, Parallel and Distributed Systems (ISPDS)*, 2020.

[16] C. Zhou, Z. Fang, X. Xu, X. Zhang, Y. Ding, X. Jiang, and Y. ji, “Using long short-term memory networks to predict energy consumption of air-conditioning systems,” *Elsevier. on Sustainable Cities and Society*, vol. 55, Apr. 2020.

[17] H. Zang, R. Xu, L. Cheng, T. Ding, T. Liu, T. Wei, and T. Sun, “Residential load forecasting based on LSTM fusing self-attention mechanism with pooling,” *Elsevier. on Energy*, vol. 229, no. 9, Aug. 2021.

[18] W. Wang, T. Hong, X. Xu, J. Chen, Z. Liu, and N. Xu, “Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm,” *Elsevier. on Applied Energy*, vol. 248, pp. 217-230, Aug. 2019.

[19] J. Liu, Q. Zhang, Z. Dong, G. Li, Y. Xie, and K. Li, “Quantitative evaluation of the building energy performance based on short-term energy predictions,” *Elsevier. on Energy*, vol. 223, May. 2021.

[20] Z. Zheng, H. Chen, and X. Luo, “Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network,” *Elsevier. on Energy Procedia*, vol. 158, pp. 2713-2718, Feb. 2019.

[21] C. Fan, J. Wang, W. Gang, and S. Li, “Assessment of deep recurrent neural network-based strategies for short-term building energy predictions,” *Elsevier. on Applied energy*, vol. 236, pp. 700-710, Sep. 2019.

[22] A. Rosemary, H. Gong, C. Ababei, and D.M. Ionel, “LSTM forecasts for smart home electricity usage,” *IEEE Int. International Conference on Renewable Energy Research and Application (ICRERA)*, 2020.

[23] S. Hadri, Y. Naitmalek, M. Najib, B. Mohamed, Y. Fakhri, and M. Elaroussi, “A comparative study of predictive approaches for load forecasting in smart buildings,” *Elsevier. Procedia Computer Science (PCS)*, 2019.

[24] C.H. Kim, M. Kim, and Y.J. Song, “Sequence-to-sequence deep learning model for building energy consumption prediction with dynamic simulation modeling,” *Elsevier. on Journal of Building Engineering*, vol. 43, Nov. 2021.

[25] Z.A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S.W. Baik, “Towards efficient electricity forecasting in residential and commercial buildings: a novel hybrid CNN with a LSTM-AE based framework,” *Sensors. on Multidisciplinary Digital Publishing Institute*, vol. 20, no. 5, Jan. 2020.

[26] S. Abhishek and S.K. Jain, “LSTM-based short-term electrical load forecasting framework with improved input feature space,” *IEEE Int. International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, 2021.

[27] R. Sendra-Arranz and A. Gutiérrez, “A long short-term memory artificial neural network to predict daily HVAC consumption in buildings,” *Elsevier. on Energy and Buildings*, vol. 216, Jun. 2020.

[28] A. Bellahsen and H. Dagdougui, “Aggregated short-term load forecasting for heterogeneous buildings using machine learning with peak estimation,” *Elsevier. on Energy and Buildings*, vol. 237, Apr. 2021.

[29] Z. Jingwen, Z. Ling, Y. Kang, L. You, Y. Zhao, Z. Xiao, and X. Chen, “HVAC load forecasting in office buildings using machine learning,” *IEEE Int. International Conference on Power and Renewable Energy (ICPRE)*, 2021.

[30] A. Shaqour, T. Ono, A. Hagishima, and H. Farzaneh, “Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building,” *Elsevier. on Energy and AI*, vol. 8, May. 2022.

[31] X. Fang, G. Gong, G. Li, L. Chun, W. Li, and P. Peng, “A hybrid deep transfer learning strategy for short term cross-building energy prediction,” *Elsevier. on Energy*, vol. 215, Sep. 2021.

[32] J. Barzola-Monteses, M. Espinoza-Andaluz, M. Mite-León, and M. Flores-Morán, “Energy consumption of a building by using long short-term memory network: a forecasting study,” *IEEE Int. International Conference of the Chilean Computer Science Society (SCCC)*, 2020.

[33] T.-Y. Kim and S.-B. Cho, “Predicting residential energy consumption using CNN-LSTM neural networks,” *Elsevier. on Energy*, vol. 182, pp. 72-81, Sep. 2019.

[34] X. Shao, C. Pu, Y. Zhang, and C.S. Kim, “Domain fusion CNN-LSTM for short-term power consumption forecasting,” *IEEE Access*, vol. 8, pp. 188352-188362, Sep. 2020.

[35] N. Somu, G.R. MR, and K. Ramamritham, “A deep learning framework for building energy consumption forecast,” *Elsevier. on Renewable and Sustainable Energy Reviews*, vol. 137, Mar. 2021.

[36] M. Palak, G. Revati, and A. Sheikh, “Smart building energy management using deep learning based predictions,” *IEEE. on North American Power Symposium (NAPS)*, 2021.

[37] L. Wen, K. Zhou, and S. Yang, “Load demand forecasting of residential buildings using a deep learning model,” *Elsevier. on Electric Power Systems Research*, vol. 179, Feb. 2020.

[38] M. Sajjad, Z.A. Khan, A. Ullah, T. Hussain, W. Ullah, M.Y. Lee, and S.W. Baik, “A novel CNN-GRU-based hybrid approach for short-term residential load forecasting,” *IEEE Access*, vol. 8, pp. 143759-143768, Jul. 2020.

[39] D.P. Chassin, K. Schneider, and C. Gerkensmeyer, “GridLAB-D: An open-source power systems modeling and simulation environment,” *IEEE/PES Transmission and Distribution Conference and Exposition*, 2008.

[40] K.P. Schneider, J.C. Fuller, and D.P. Chassin, “Multi-state load models for distribution system analysis,” *IEEE Trans. on Power Systems*, vol. 26, no. 4, pp. 2425-2433, Sep. 2011.

[41] C. Zhang, J. Li, Y. Zhao, T. Li, Q. Chen, and X. Zhang, “A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process,” *Elsevier. on Energy and Buildings*, vol. 225, Oct. 2020.

[42] Q. Meng and M. Mourshed, “Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures,” *Elsevier. on Energy and Buildings*, vol. 155, pp. 260-268. 2017.

[43] J.S. Haberl, D.E. Claridge, and C. Culp, “ASHRAE’s Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine what was really saved by the retrofit,” [Online]. Available: *http://esl.tamu.edu*, 2005.

[44] E.P.R. Institute, “DOE SHINES Residential Demonstration,” [Online]. Available: *https://dashboards.epri.com/shines-residential/dashboard*, 2020.