# Deep Neural Network Model for Improving Price Prediction of Natural Gas

Aliyuda Ali
*Department of Computing and Data Science*
*Birmingham City University, UK*
Aliyuda.Ali@bcu.ac.uk

M. K. Ahmed
*Department of Computer Science*
*Gombe State Univesity*
Gombe, Nigeria
mkahmed@gsu.edu.ng

Kachalla Aliyuda
*Department of Geology*
*Gombe State University*
Gombe, Nigeria
aliyudakachalla@gmail.com

Abdulwahab Muhammed Bello
*Department of Earth Sciences*
*Durham University*
Durham, UK
abdulwahab.m.bello@durham.ac.uk

***Abstract*— Natural gas accounts for one of the most industriously marketed energy commodities with a meaningful impact on various financial activities around the world. As direction of price for natural gas changes over time, accurate price prediction of natural gas is essential since this prediction is useful in decision making, commodity marketing, and sustainability planning. In this paper, a deep neural network (DNN) model for monthly price prediction of natural gas is proposed. Deep neural networks are becoming the standard tools that offer a lot of values to researchers for solving different problems in the machine learning and data science community due to their ability for increasing model accuracy. The proposed DNN model presented in this paper utilizes the capability of fully connected layers for learning the dynamics in natural gas price data and the efficiency of Rectified Linear Unit (ReLU) function for performing threshold operations on each input element. A wide range of monthly data covering 281 months were used to develop and test the predictive capability of the proposed DNN model. In comparison to five recently reported mainstream machine learning models, overall results disclose that the proposed DNN model demonstrates superior performance over the mainstream machine learning models with mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination ($R^2$) of 0.0595, 0.2440 and 0.9937, respectively.**

***Keywords—Data-driven modelling, deep neural network, machine learning, natural gas industry, natural gas spot price.***

## I. Introduction

Natural gas plays a crucial role in socio-economic development across the world and its demand globally is continuously increasing in both the developed and developing countries [1]. Natural gas industries play strategic roles to ensure that the produced gas is accessible to meet the raised demand in all seasons. Industrial and residential operations are responsible for a large proportion of natural gas consumption in developed nations, and thus are targets of energy efficiency plans[2], [3]. Price prediction for natural gas is important for a number of reasons including energy sustainability planning, energy investment, hedging capabilities and policy decision making. Natural gas consumption by industrial activities is projected to grow yearly across the world [4]. As mobilizations about environment and climate change keep growing across the globe, renewable energy sources are growing but slowly and low-carbon energy sources are hard to investigate in some areas. Given that natural gas accounts for one of the major marketed energy commodities with a meaningful impact on various financial activities across the globe, knowledge of natural gas price direction is essential to a variety of stakeholders in the natural gas industry. In particular, fast and accurate models for price prediction of natural gas can be of great support in the industry and contribute to business planning with respect to effective energy policy implementation [5]. Nevertheless, fast and accurate forecasting of natural gas spot price is conventionally a complex and challenging task considering the nonlinearities of some variables upon which the natural gas price depends [4]. Benchmarks for natural gas spot price across the world include the Russian border price available in Germany, the Henry hub price available in United States of America, the Indonesian Liquified Natural Gas price available in Japan, and the National Balancing Point price available in the United Kingdom [4].

Recently, application of AI and ML approaches to uncover solutions to intricate problems is becoming attractive in the energy and other engineering related disciplines [6] [7], [8], [9]. AI and ML approaches have been attractive in the energy sector due to the fact that they have been successfully applied in well production forecasting [10], estimation of geological parameters impact on oil reservoirs recovery [11], compressors failure modes prediction in Liquid Natural Gas (LNG) compressors [12], carbon emission centric forecasting [13], impact modeling of electricity tariffs [14], forecast electricity price [15], energy ratings classification [16], to mention but a few.

A variety of machine learning algorithms have also been applied by various researchers to forecast natural gas spot price. The authors in [17] reported the application support vector regression (SVR) algorithm to predict daily and weekly spot prices for natural gas. The authors in [18] reported the application of four different machine learning algorithms including Gaussian process regression (GPR), gradient boosting machines (GBM), SVR and ANN for monthly price forecast of natural gas. The authors in [19] reported a study that applied deep neural network for loads forecasting of natural gas. In their work, 62 historical data that constitutes a variety of climates and covers a range of geographical regions in the United States were used. In our previous work [4], we reported an ensemble learning model for price prediction of natural gas based on least squares boosting.

Even though the previously reported works that applied machine learning algorithms to forecast the spot price of natural gas have performed well to some extent, the need to improve price prediction of natural gas, as up-to-date data keeps on emerging is paramount. This need is crucial considering that the previously reported models are developed using vanilla machine learning algorithms which resulted in yielding models with high variance and in turn, unable to generalize well. To fill in this gap, this study proposes a deep neural network model that utilizes the capability of fully connected layers for learning the dynamics in natural gas price data and the efficiency of Rectified Linear Unit (ReLU) function for performing threshold operations on each input element. Thus, the objective of this paper is to develop a model that improves the price prediction accuracy of natural gas and generalizes well on the data.

The remaining parts of this paper are organized as follows. In section two, materials and methods are presented. These include the development process of the proposed DNN model and source, description and pre-processing of data used in this work. In section three, computational results and their discussions are presented. Section four presents conclusion and future work.

## II. Materials And Methods

### A. Deep Neural Network (DNN)

To understand how DNN works, a brief description of the working principle of ANN, from which DNN evolves, is paramount. Similar to the synapses in a human brain, information in ANN is processed by connecting different uncomplicated nodes to establish composite networks [2]. In ANN, signals are received as input by each node. These nodes then use activation functions to process their inputs and pass their resulting outputs to other nodes via a weighted connection. As such, the determinants of an ANN output include its architecture, activation functions, and the weight value. A diagrammatic illustration of an artificial neuron is presented in Fig.1. Given a unit of a neuron $i$, as depicted in Fig. 1, assume there is a connection from the neuron's input signals to other units say $x_i$ (where $i = 1,2,\ldots,n$) with respective weights $w_i$, summing and activating the neuron's input signals are performed as a way of processing its unit.

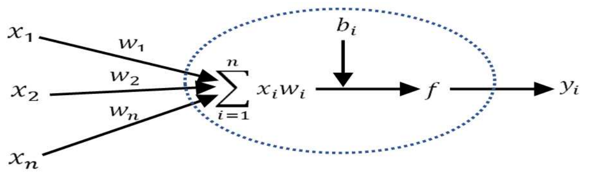


Fig. 1. A diagrammatic illustration of an artificial neuron process.

The output unit $y_i$ is defined as:

$$y_i = f\left(\sum_{i=1}^{n} x_i w_i + b_i\right) \tag{1}$$

where $b_i$ denotes a bias for the input $i$, and $f$ denotes the activation function. In this work, the ReLU function is used and can be defined as

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

Given the above description of ANN, a DNN can simply be described as a neural network that possesses some level of complexity, normally at least two hidden layers as shown in Fig. 2.

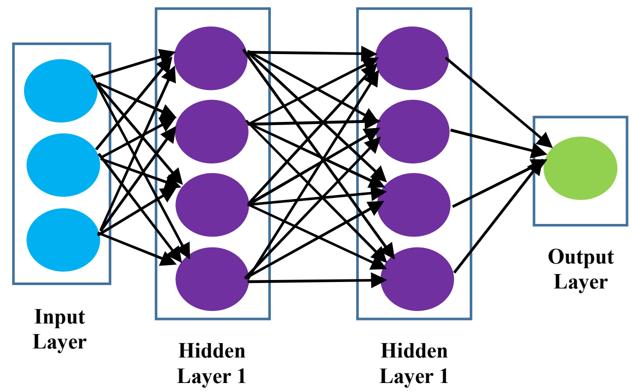


Fig. 2. A deep neural network architecture with two hidden layers.

### B. The Proposed DNN Model Development Process

The proposed DNN model development process for price prediction of natural gas involves a series of steps as shown in Fig. 3. As described in Fig. 3, the process starts with collecting and pre-processing the input/output data. The pre-processing step involves labelling each output variable against its corresponding input variables and removing records that contain missing values. The process continued by partitioning the pre-processed data into training and testing proportions. In this work, cross validation technique is used to partition the data where 80% and 20% of the entire dataset were allocated to model training and testing, respectively. Next, the DNN framework is trained using the training dataset and the performance of the trained model is analyzed. Having improved the price prediction accuracy, the trained model is then tested using a dataset that the model has not encountered during the training process.

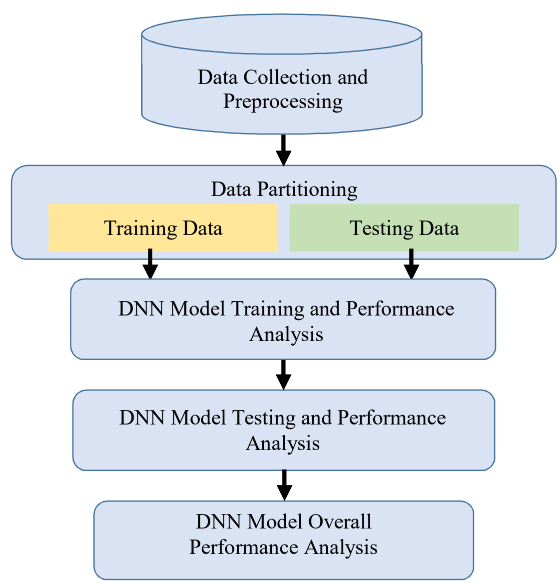


Fig. 3. DNN model development process for price prediction of natural gas

A flowchart showing the workflow of the proposed DNN model is presented in Fig. 4.

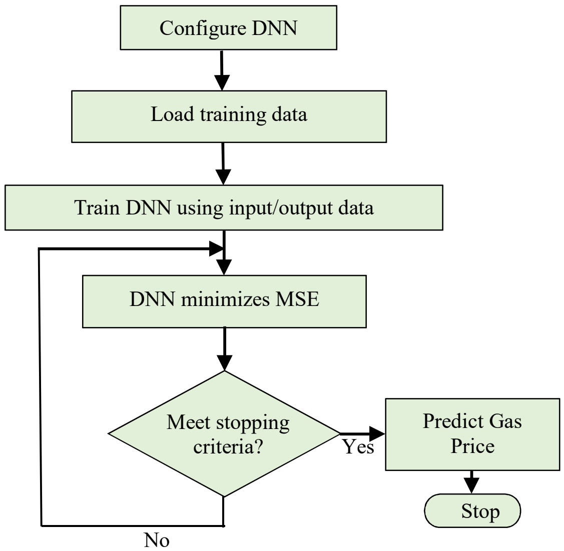


Fig. 4. A flowchart of the proposed DNN models.

The performance of the DNN model to predict price using the unseen dataset is then analyzed. Finally, the DNN overall performance on both the training and testing datasets is analyzed.

To evaluate the performance of the DNN model, two performance metrics are used namely, $R^2$ and RMSE. The $R^2$ is selected to evaluate the prediction accuracy and is mathematically given as follows [2]:

$$R^2 = \left[1 - \frac{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}{var(y)}\right] \quad (3)$$

where $N$ denotes total number of data samples, $\hat{y}_t \; and \; y_t$ denote the actual/real value and predicted value at time t, respectively. The RMSE is selected to evaluate the prediction error and is expressed as follows [2]:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \quad (4)$$

This work employed grid search technique to get the optimum values of parameters for developing the DNN model. Table I gives a description of the parameter values used to develop the DNN model.

TABLE I. PARAMETERS USED FOR DNN MODEL DEVELOPMENT.

| Parameter | Description/Value |
|---|---|
| Input variables | 8 |
| Output variables | 1 |
| Hidden layers | 2 |
| Neurons in hidden layer 1 | 30 |
| Neurons in hidden layer 2 | 20 |
| Activation function on hidden layers | Rectified Linear Unit |
| Gradient tolerance | 1e-06 |
| Loss tolerance | 1e-06 |

## C. Data Collection and Description

The observed data utilized in this work relate to the U.S. Henry hub monthly price data covering the time from January 1997 to May 2020. The data comprises of 281 data records retrieved from the official site of the U.S. Energy Information Administration [20]. Eight (8) variables were used as inputs including marketed production of natural gas, total storage capacity, natural gas production activity, imports activity of natural gas, weather change (cooling), weather change (heating), heating oil price, and crude oil price. Whereas spot price of natural gas was used as the output variable. Statistical summary is shown in Table II for all the variables used in this work.

TABLE II. STATISTICAL SUMMARY OF ALL VARIABLES USED IN THIS STUDY.

| Variable | Unit | Minimum | Maximum | Average |
|---|---|---|---|---|
| Marketed production of natural gas | Million cubic feet | 1400941 | 3194898 | 1979874.38 |
| Natural gas total storage capacity | Million cubic feet | 7952224 | 9264128 | 8625721.92 |
| Natural gas production activity | Count | 79 | 1585 | 690.76 |
| Imports of natural gas | Million cubic feet | 189403 | 426534 | 291789.34 |
| Cooling weather change | Number | 3 | 404 | 112.48 |
| Heating weather change | Number | 3 | 996 | 358.97 |
| Heating oil price | Dollars per Gallon | 0.304 | 3.801 | 1.6335 |
| Crude oil price | Dollars per Barrel | 11.35 | 133.88 | 55.71 |
| Spot price of natural gas | Dollars per Million Btu | 1.72 | 13.42 | 4.23 |

# III. Computational Results And Discussion

The proposed DNN model for price prediction of natural gas is trained and tested using 281 data samples. The results that demonstrate the model's predictive capability are presented in Table III.

TABLE III. RESULTS OF PRICE PREDICTION OF NATURAL GAS BY THE PROPOSED DNN MODEL.

| Training | | Testing | | Overall | |
|---|---|---|---|---|---|
| $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE |
| 0.9951 | 0.2175 | 0.9865 | 0.3296 | 0.9937 | 0.2440 |

The results presented in Table III demonstrate the predictive performance of the proposed DNN model developed in this work. From the results, it can be observed that the model performed well in terms of training, testing, and overall performance. In particular, the DNN model performed well by generalizing well on the data which prevents it from overfitting the data. This can be observed in the model's ability to predict the natural gas price with high accuracy using the testing dataset that the model has not encountered during the training process. To perceive the performance of the DNN model visually, graphical representations of model's predictive performance with respect to training, testing, and overall are presented in Figures 5, 6, and 7, respectively.

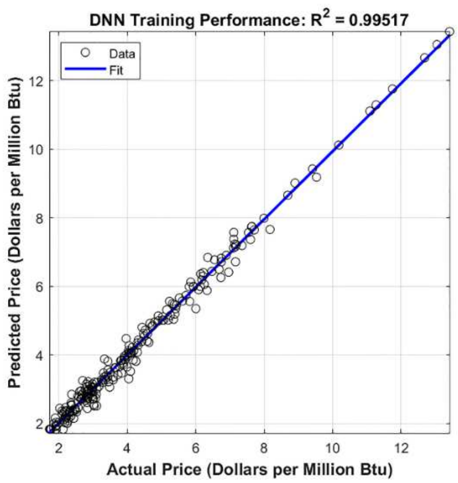


Fig. 5. Graphical representation of the proposed DNN model's training performance.

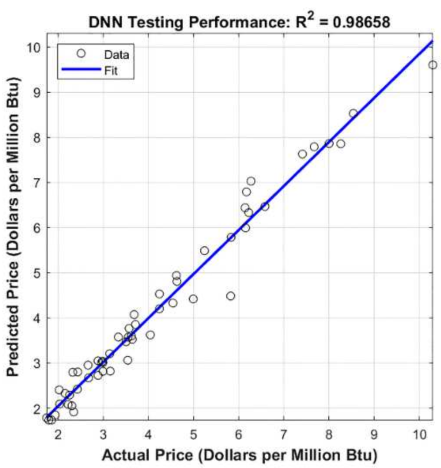


Fig. 6. Graphical representation of the proposed DNN model's testing performance.

Furthermore, Fig. 8 displays graphically the actual natural gas price values against the ones predicted by the proposed DNN model for the entire data samples used in this work which represent the period from January 1997 to May 2020. From Fig. 7, it can be observed that the price values predicted by the proposed DNN model match well with the actual price values with insignificant errors.

To evaluate the predictive performance of the proposed DNN model, comparison of the model's overall performance is made to recently reported studies on price prediction of natural gas using machine learning algorithms. Three statistical tools including $R^2$, MSE, and RMSE are used as metrics to evaluate the proposed DNN model's predictive performance against five mainstream machine learning algorithms reported recently and used the same dataset for price prediction of natural gas. Table IV presents the comparative results.

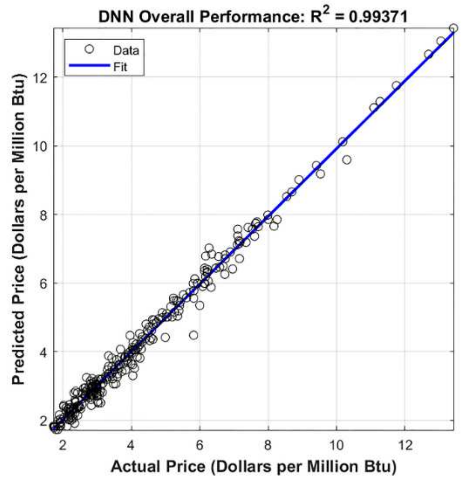


Fig. 7. Graphical representation of the proposed DNN model's overall performance.

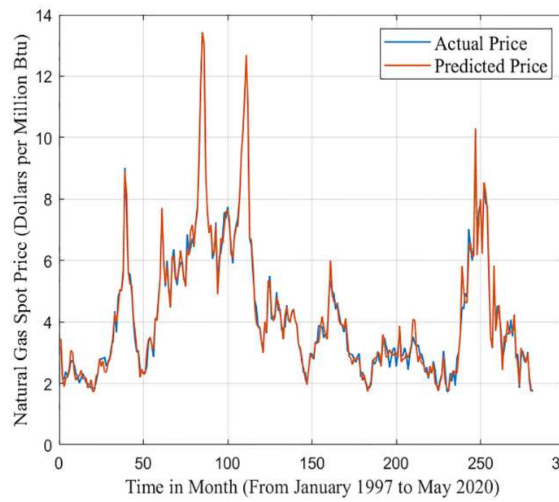


Fig. 8. Graphical representation of actual against predicted price values of natural gas for the entire period covered in this work.

TABLE IV. COMPARATIVE RESULTS OF THE PROPOSED DNN MODEL TO OTHER MAINSTREAM MACHINE LEARNING ALGORITHMS.

| Author | Algorithm | $R^2$ | MSE | RMSE |
|---|---|---|---|---|
| Su et al., 2019 [18] | ANN | 0.8904 | 0.5363 | 0.7247 |
| | SVN | 0.8437 | 0.7673 | 0.8757 |
| | GBM | 0.8006 | 0.9786 | 0.9888 |
| | GPR | 0.8374 | 0.7980 | 0.8932 |
| Ali, 2020 [4] | LSBoost | 0.9668 | 0.3248 | 0.5699 |
| **This study** | **DNN** | **0.9937** | **0.0595** | **0.2440** |

Looking at the comparative results displayed in Table IV, it can be observed that the proposed DNN model demonstrates superior performance over the five mainstream machine learning algorithms reported in previous studies. The superior performance of the proposed DNN model can be seen in its ability to record the highest $R^2$ which represent prediction accuracy and least MSE and RMSE which represent prediction errors. To this end, it can be seen that the proposed DNN model has improved the price prediction of natural gas over the previous reported studies for the dataset used in this work.

## IV. Conclusion and Future Work

In this work, a DNN model for improving price prediction of natural gas is advocated. The predictive capability of the proposed DNN model is evaluated using a collection of time series data from the U.S. Henry hub price data. A collection of 281 data samples were used to train and test the performance of the proposed model. In comparison to five mainstream machine learning algorithms recently reported for price prediction of natural gas, the proposed DNN model recorded superior performance with the highest prediction accuracy of > 99% and least MSE and RMSE of 0.0595 and 0.2440, respectively. Considering that natural gas accounts for one of the major marketed commodities in the energy domain with a meaningful impact on various financial activities across the world, the model proposed in this work can be of great support to a variety of stakeholders in the natural gas industry. In particular, the proposed DNN model can serve as an effective tool to decision makers in terms of business planning and effective policy implementation. The future scope of this work will consider evaluating the capability of the proposed DNN model on datasets of other marketed commodities in and outside the energy domain.

## Acknowledgment

The authors are grateful to the University of Bahrain for supporting the presentation of this work at the 2021 International Conference on Data Analytics for Business and Industry (ICDABI). Special thanks to the three anonymous reviewers for their constructive comments and recommendations towards improving the quality of this paper.

## References


[1] A. Ali and L. Guo, "Data-driven based investigation of pressure dynamics in underground hydrocarbon reservoirs," *Energy Reports*, vol. 7, pp. 104–110, 2021, doi: 10.1016/j.egyr.2021.02.036.

[2] A. Ali, "Data-driven based machine learning models for predicting the deliverability of underground natural gas storage in salt caverns," *Energy*, vol. 229, p. 120648, 2021, doi: 10.1016/j.energy.2021.120648.

[3] R. O'Shea, R. Lin, D. M. Wall, J. D. Browne, and J. D. Murphy, "Dummy ReferenceAssessing the impact of using biogas from whiskey by-products to reduce natural gas consumption and greenhouse gas emissions at a large distillery (under review)," *Appl. Energy*, 2020, doi: 10.1016/j.apenergy.2020.115812.

[4] A. Ali, "Ensemble Learning Model for Prediction of Natural Gas Spot Price Based on Least Squares Boosting Algorithm," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325615.

[5] A. Ali, K. Aliyuda, M. K. Ahmed, and S. Saleh, "Data-driven-based pressure field decomposition and reconstruction for single-phase flow model," *Science Forum (Journal of Pure and Applied Sciences),* vol. 21, no. 2, pp. 357 – 364, 2021, doi: http://dx.doi.org/10.5455/sf.81326

[6] A. Ali and L. Guo, "Adaptive neuro-fuzzy approach for prediction of dewpoint pressure for gas condensate reservoirs," *Pet. Sci. Technol.*, vol. 38, no. 9, pp. 673–681, 2020, doi: 10.1080/10916466.2020.1769655.

[7] F. Chiarello, P. Belingheri, and G. Fantoni, "Data science for engineering design: State of the art and future directions," *Comput. Ind.*, vol. 129, p. 103447, 2021, doi: 10.1016/j.compind.2021.103447.

[8] A. Aliyuda and M. K. Ahmed, "Collaboration among Agents to Detect Fault in Power Distribution System," vol. 6, no. 8, pp. 102–107, 2016.

[9] A. Aliyuda, "Towards the Design of Cyber-Physical System via Multi-Agent System Technology," *Int. J. Sci. Eng. Res.*, vol. 7, no. 10, pp. 155–161, 2016.

[10] A. Ali and L. Guo, "Neuro-Adaptive Learning Approach for Predicting Production Performance and Pressure Dynamics of Gas Condensation Reservoir," *IFAC-PapersOnLine*, vol. 52, no. 29, pp. 122–127, 2019, doi: 10.1016/j.ifacol.2019.12.632.

[11] K. Aliyuda, J. Howell, A. Hartley, and A. Ali, "Stratigraphic controls on hydrocarbon recovery in clastic reservoirs of the Norwegian Continental Shelf," *Pet. Geosci.*, pp. petgeo2019-133, 2020, doi: 10.1144/petgeo2019-133.

[12] F. Hidalgo-Mompeán, J. F. Gómez Fernández, G. Cerruela-García, and A. Crespo Márquez, "Dimensionality analysis in machine learning failure detection models. A case study with LNG compressors," *Comput. Ind.*, vol. 128, 2021, doi: 10.1016/j.compind.2021.103434.

[13] S. Kumar, A. K. Shukla, and P. K. Muhuri, "Anomaly based novel multi-source unsupervised transfer learning approach for carbon emission centric GDP prediction," *Comput. Ind.*, vol. 126, p. 103396, 2021, doi: 10.1016/j.compind.2021.103396.

[14] M. Heleno, D. Sehloff, A. Coelho, and A. Valenzuela, "Probabilistic impact of electricity tariffs on distribution grids considering adoption of solar and storage technologies," *Appl. Energy*, vol. 279, no. June, p. 115826, 2020, doi: 10.1016/j.apenergy.2020.115826.

[15] M. Narajewski and F. Ziel, "Ensemble Forecasting for Intraday Electricity Prices: Simulating Trajectories," *Appl. Energy*, vol. 279, no. April, p. 115801, 2020, doi: 10.1016/j.apenergy.2020.115801.

[16] U. Ali *et al.*, "A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making," *Appl. Energy*, vol. 279, no. September, p. 115834, 2020, doi: 10.1016/j.apenergy.2020.115834.

[17] E. Ceperi, "Short-term forecasting of natural gas prices using machine learning and feature selection algorithms," vol. 140, 2017, doi: 10.1016/j.energy.2017.09.026.

[18] M. Su, Z. Zhang, Y. Zhu, D. Zha, and W. Wen, "Data driven natural gas spot price prediction models using machine learning methods," *Energies*, vol. 12, no. 9, 2019, doi: 10.3390/en12091680.

[19] G. D. Merkel, R. J. Povinelli, and R. H. Brown, "Short-Term Load Forecasting of Natural Gas with Deep Neural Network Regression †," *Energies*, 2018, doi: 10.3390/en11082008.

[20] U. S. Energy Information Administration (EIA), available online: https://www.eia.gov/ (accessed on 31 Augist 2020).