

# SHORT-TERM LOAD FORECASTING USING A PARTICLE SWARM OPTIMISED MULTI-HEAD ATTENTION-AUGMENTED CNN-LSTM MODEL

Paapa Kwesi Quansah<sup>1</sup> and Edwin Kwesi Ansah Tenkorang<sup>2</sup>

Department of Electrical/Electronic Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

## **ABSTRACT**

*Short-term load forecasting is of utmost importance in the efficient operation and planning of power systems, given their inherent non-linear and dynamic nature. Recent strides in deep learning have shown promise in addressing this challenge. However, these methods often grapple with hyperparameter sensitivity, opaqueness in interpretability, and high computational overhead for real-time deployment. This paper proposes an innovative approach that effectively overcomes the aforementioned problems. The approach utilizes the Particle Swarm Optimization algorithm to autonomously tune hyperparameters, a Multi-Head Attention mechanism to discern the salient features crucial for accurate forecasting, and a streamlined framework for computational efficiency. The method was subjected to rigorous evaluation using a genuine electricity demand dataset. The results underscore its superiority in terms of accuracy, robustness, and computational efficiency. Notably, its Mean Absolute Percentage Error of 1.9376 marks a significant improvement over existing state-of-the-art approaches, heralding a new era in short-term load forecasting.*

## **KEYWORDS**

*Short-Term Load Forecasting, Deep Learning, Particle-Swarm Optimization, Multi-Head Attention, CNN-LSTM Network, Electricity Demand*

## **1. INTRODUCTION**

In contemporary society, electrical energy has emerged as a pivotal resource propelling the economic and societal progress of nations worldwide. It is extensively utilised in industries, including manufacturing, mining, construction, and healthcare, among others. The provision of consistent and high-quality electrical power is not merely a convenience; rather, it is imperative to sustain investor confidence in economies and foster further development [1]. With the advent of new technological advancements, electricity demand has surged, creating an urgent need for more cost-effective and reliable power supply solutions [2].

The current energy infrastructure lacks substantial energy storage capabilities in the generation, transmission, and distribution systems [3]. This deficiency necessitates a precise balance between electricity generation and consumption. The maintenance of balance is contingent upon the utilisation of an accurate load forecasting approach. Adapting electricity generation to dynamically meet shifting demand patterns is paramount; since failure to do so puts the stability of the entire power system at risk [4]. Moreover, as the world pivots towards the increased adoption of renewable energy sources [5], power grids have witnessed a substantial transformation in their composition and structure. This integration of renewable energy sources, such as wind and solar power, introduces a degree of unpredictability into energy generation due to the stochastic nature of these sources [6]. Consequently, ensuring a stable and secure power system operation becomes an even more complex endeavour, demanding meticulous power planning and precise load forecasting.

Electric load forecasting is the practice of predicting electricity demand within a specific region. This process can be categorised into three distinct groups: short-term, medium-term, and long-term forecasting, depending on the forecasting horizon. Short-term load forecasting (STLF), which focuses on predicting electricity demand for upcoming hours, a day, or a few days, serves as the foundation for effective power system operation and analysis. It facilitates the optimization of the operating schedules of generating units, including their start and stop times, and their expected output. The accuracy of STLF is of critical importance, as it directly influences the efficient utilisation of generating units [7]. The absence of accurate short-term load forecasting can lead to many operational challenges, including load shedding, partial or complete outages, and voltage instability. These issues can have detrimental effects on equipment functionality and pose potential risks to human safety.

Short-term load forecasting methods are pivotal in achieving this precision. These methods can be broadly classified into two main categories: statistical methods and machine learning methods [8-9]. Machine learning-based load forecasting methods, such as the autoregressive integrated moving average model (ARIMA) [10], long short-term memory (LSTM) [11], generative adversarial network (GAN) [12], and convolutional neural network (CNN) [13], have gained prominence. These machine learning methods excel at capturing complex nonlinear data features within load patterns [14]. They leverage the ability to discern similarities in electricity consumption across diverse power supply areas and customer types, allowing for more accurate and feasible load forecasting through the consideration of spatial-temporal coupling correlations.

## 2.1. Motivation

Based on the existing research, the following three shortcomings need to be addressed to improve the forecasting effect of the spatial-temporal distribution of the system load: (i) the lack of flexibility and scalability of traditional statistical methods, (ii) the high computational complexity of deep learning methods, and (iii) the inability of existing methods to capture the spatial-temporal correlations in load patterns.

Considering these challenges, this paper proposes a novel short-term load forecasting model that uses a particle swarm-optimised multi-head attention-augmented CNN-LSTM network. The proposed model employs a particle swarm optimization (PSO) algorithm to identify the optimal hyperparameters of the CNN-LSTM network. This enhances the model's resilience to overfitting and its accuracy. Additionally, the multi-head attention mechanism is used to learn the importance of different features for the forecasting task. Finally, a hybrid CNN-LSTM Model is used to help the system capture the spatial-temporal correlations in load patterns, hence enhancing its forecasting accuracy.

## 2.2. Contributions

The following are the main contributions of the paper:

1. **Feature Extraction:** To improve efficiency during feature extraction for STLF, PSO is employed to optimise model hyperparameters, leading to enhanced efficiency in extracting significant features with lower computational resources.
2. **Attention-Augmented Hybrid Model:** Given that power demand is impacted by short-term fluctuations and long-term trends in data, a hybrid model is used to detect both temporal and extended dependencies, improving accuracy.
3. **Performance evaluation:** The effectiveness of PSO-A2C-LNet has been validated using three real-world electricity demand data sets (from Panama, France, and the US). Testing results demonstrate that the PSO-A2C-LNet outperforms benchmarks in terms of forecasting performance.

### 2.3. Structure of this paper

The rest of the paper is structured as follows. Section II provides comprehensive explanations and definitions of key terminology. Section III provides an in-depth exposition of the proposed framework and a comprehensive explanation of its operation. The findings of our tests on the framework are presented in Section IV. Section V concludes the paper.

## 2. FRAMEWORK COMPONENTS

### 2.1. Definitions of Key Terms

#### 2.1.1 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a deep learning model designed primarily for image-related tasks, but it can also be applied to other grid-like data, such as audio or time series data. CNNs are especially effective at capturing spatial dependencies within inputs. [14].

The CNN achieves the localization of spatial dependencies by using the following layers:

1. **Convolutional Layer:** The core operation in a CNN is the convolution operation. Convolutional layers use learnable filters or kernels to scan the input data in a localised and overlapping manner. Each filter detects specific features, like edges, textures, or more complex patterns.

Mathematically, the 2D convolution operation is defined as follows:

$$(Y * X)(i, j) = \sum_{m=1}^M \sum_{n=1}^N X(i + m - 1, j + n - 1) \cdot Y(m, n) \quad (1)$$

Here,

- $Y$  is the filter (kernel) of size  $M \times N$ .
- $X$  is the input data of size  $(W, H)$ .
- $(i, j)$  represents the coordinates of the output feature map.
- $(m, n)$  iterates over the filter dimensions.

By sliding the filter across the input, the convolution operation computes feature maps that highlight different aspects of the input. This process effectively captures spatial dependencies.

2. **Pooling Layer:** Pooling layers are predominantly used to downsample the feature maps, reducing their spatial dimensions. Common pooling operations include max-pooling and average-pooling. Pooling aids in the invariance of network translation and minimises the computational overhead.

For max-pooling, the operation is defined as:

$$\text{Max-Pooling}(x, p, q) = \max_{i,j} x(p + i, q + j) \quad (2)$$

where  $x$  is the input feature map, and  $(p, q)$  represents the pooling window position. Max-pooling retains the most significant information within the window.

**3. Fully Connected Layer:** After multiple convolutional and pooling layers, the spatial dimensions are reduced, and the network connects to one or more fully connected layers, also known as dense layers. These layers perform classification or regression tasks by learning high-level representations.

Recognizing and exploiting spatial dependencies in CNNs is facilitated through several key mechanisms [15]. CNNs utilise local receptive fields, whereby each neuron is connected to a small region of the input data. This enables neurons to specialise in detecting specific features within their receptive fields, hence facilitating the network's ability to record spatial relationships across multiple scales. Additionally, weight sharing is a fundamental aspect of CNNs, where the same set of filters is applied consistently across the entire input. This weight sharing allows the network to learn translation invariant patterns, boosting its capacity to grasp spatial dependencies. Moreover, CNNs employ a hierarchical representation approach, where deeper layers in the network combine higher-level features derived from lower-level features. This hierarchical representation aids the network in comprehending complex spatial dependencies by gradually constructing abstractions. These mechanisms collectively empower CNNs to effectively model and exploit spatial dependencies in input data.

### 2.1.2 Long Short-Term Memory Network

The LSTM network is a type of recurrent neural network (RNN) architecture that is designed to capture and model sequential data while addressing the vanishing gradient problem that plagues traditional RNNs. LSTMs are particularly effective at locating and modelling long-term dependencies in sequential data.

LSTMs consist of multiple interconnected cells, each with its own set of gates and memory cells [16]. The primary components of an LSTM cell are:

**Forget Gate ( $f_t$ ):** Controls what information from the previous cell state ( $C_{t-1}$ ) should be discarded or kept. It takes the previous cell state and the current input ( $x_t$ ) as input and produces a forget gate output.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

**Input Gate ( $i_t$ ):** Determines what new information should be added to the cell state. It takes the previous cell state and the current input and produces an input gate output.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

**Candidate Cell State ( $\tilde{C}_t$ ):** This is a candidate new cell state, computed using the current input and a tanh activation function.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

**Cell State Update ( $C_t$ ):** The cell state is updated by combining the information retained from the previous cell state ( $(f_t \cdot C_{t-1})$ ) and the new candidate cell state ( $(i_t \cdot \tilde{C}_t)$ ).

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (6)$$

**Output Gate ( $o_t$ ):** Determines what part of the cell state should be output as the final prediction. It takes the current input and the updated cell state and produces an output gate output.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

**Hidden State ( $h_t$ ):** The hidden state is the output of the LSTM cell, which is used as the prediction and is also passed to the next time step. It is calculated by applying the output gate to the cell state.

$$h_t = o_t \cdot \tanh(C_t) \quad (8)$$

LSTMs address the vanishing gradient issue of traditional RNNs by introducing key components: the cell state ( $C_t$ ) and the forget gate ( $f_t$ ) [17]. The forget gate dynamically adjusts ( $f_t$ ) to enable LSTMs to remember or discard information from distant time steps, facilitating the capture of long-term dependencies. Meanwhile, the cell state ( $C_t$ ) acts as a memory buffer, accumulating and passing relevant information across time steps, thus enabling the model to recognize and exploit long-term patterns within input sequences.

### 2.1.2 Multi-Head Attentional Mechanism

The Multi-Head Attention mechanism [18] is a key component of Transformer-based models, such as BERT and GPT, used for various natural language processing tasks. It excels at capturing extremely long-term dependencies in sequences of data.

Multi-Head Attention extends the idea of the self-attention mechanism [19] by employing multiple attention heads in parallel. Each attention head focuses on different parts of the input sequence, enabling the model to capture various types of information and dependencies simultaneously.

The primary components of Multi-Head Attention are as follows:

**Query ( $Q$ ), Key ( $K$ ), and Value ( $V$ ) Projections:** For each attention head, we project the input sequence into three different spaces: query, key, and value. These projections are learned parameters.

**Scaled Dot-Product Attention:** Each attention head computes attention scores between the query ( $Q$ ) and the keys ( $K$ ) of the input sequence and then uses these scores to weight the values ( $V$ ). The attention scores are computed as a scaled dot product:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) \cdot V \quad (9)$$

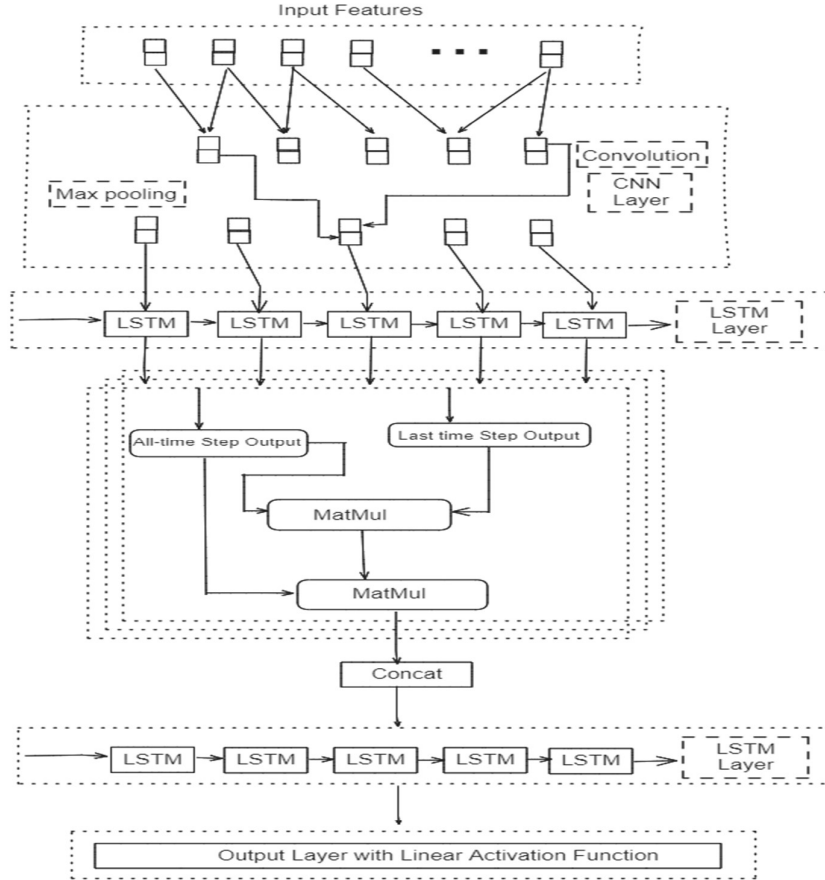
Here,  $d_k$  is the dimension of the key vectors.

**Concatenation and Linear Transformation:** After computing the attention outputs for each head, we concatenate them and apply a linear transformation to obtain the final multi-head attention output:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (10)$$

Where *Concat* concatenates the outputs from all attention heads, and  $W^0$  is a learned linear transformation.

### 3. PSO-A2C-LNET ARCHITECTURE



**Figure 1. PSO-A2C-LNet Structure Diagram**

PSO-A2C-LNet utilises the various aforementioned components to extract more relevant features of the power load data to provide better predictions. This algorithm effectively improves the prediction accuracy of STLF. The model architecture is described below.

The model starts with an input layer designed to accept sequential data. Following the input layer, a Convolutional layer is used to capture temporal spatial patterns in the data. Subsequently, a bidirectional LSTM layer is employed to model long-term dependencies both forward and backward, enabling the capture of historical data through time. The crucial Multi-Head Attention module operates on the output of the first bidirectional LSTM layer, enabling the model to focus on the most relevant features and learn their importance. To capture intricate long-term patterns, a second bidirectional LSTM layer is employed. The final LSTM layer generates a probabilistic value. The anticipated short-term demand in kilowatts per hour is predicted by the output layer, which consists of a dense layer with one neuron and a linear activation function. A few Dropout layers were interspersed among the other model's other layers to combat overfitting. Layer Normalisation is implemented subsequent to the first bidirectional layer in order to provide consistent and steady training across different inputs. The hyperbolic tangent (tanh) activation function was used for all LSTM layers.

To optimise model performance and convergence during training, PSO was employed to fine-tune critical hyperparameters. Table I shows the optimised hyperparameters and their corresponding optimization ranges.

**Table I**  
**Optimization Ranges**

Hyperparameter	Range
Adaptive Learning Rate	[0.001, 0.1]
Batch Size	[1, 128]
Number of Epochs	[100, 5000]
Weight Initialization Techniques	(Xavier, He, Random)
Loss Metrics	(MSE, Cross-Entropy, MAPE)

The specific implementation process of the proposed algorithm is provided in Algorithm 1.

---

**Algorithm 1:** PSO-A2C-LNet

---

**Data:** Input data  $D$   
**Result:** Output data  $O$   
 Initialize variables;  
 Extract features  $X$  and target values  $y$  from  $D$ ;  
 Do Pre-processing;  
 Define architecture of model;  
 Start PSO;  
**while** *stopping criterion not met* **do**  
     Find optimal parameters using PSO;  
     Check the fitness with defined model;  
     Update variables and data structures;  
 Update variables globally;  
 Train the model on the training set;  
 Evaluate the model with the three error metrics;  
 Post-processing steps;  
**return**  $O$ ;

---

#### 4. RESULTS AND DISCUSSION

This section comprehensively analyses the STLF results by implementing the above model and testing it extensively on three datasets; ERCOT, RTE, and the Panama Energy Dataset. Three regression evaluation metrics are introduced for quantitative analysis of the prediction results. The performance and fitting degree of the different models are measured by the following indicators:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\%$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

where  $n$ : Number of Observations,  $Y_i$ : Actual values at data point  $i$ ,  $\hat{Y}_i$ : Predicted values at data point  $i$  and  $\bar{Y}$ : Mean of the observed values.

**Table II**  
**Performance Metrics for Different Models on Various Datasets**

Dataset	Model	$R^2$	MAPE(%)	MAE
Panama Dataset	A2C-LNet	0.85	2.8	8.1
	<b>PSO-A2C-LNet</b>	<b>0.88</b>	<b>1.9</b>	<b>7.3</b>
	Hybrid CNN-LSTM	0.92	3.1	8.7
	Vanilla CNN	0.81	3.9	11.2
	Vanilla LSTM	0.86	3.4	10.0
ERCOT Dataset	A2C-LNet	0.92	3.1	8.7
	<b>PSO-A2C-LNet</b>	<b>0.87</b>	<b>2.1</b>	<b>7.5</b>
	Hybrid CNN-LSTM	0.89	2.6	9.1
	Vanilla CNN	0.95	3.2	7.8
	Vanilla LSTM	0.91	4.0	8.5
RTE Dataset	A2C-LNet	0.78	2.4	12.3
	<b>PSO-A2C-LNet</b>	<b>0.86</b>	<b>2.0</b>	<b>7.5</b>
	Hybrid CNN-LSTM	0.89	2.9	7.2
	Vanilla CNN	0.79	4.2	12.0
	Vanilla LSTM	0.81	3.1	11.2

From the results Table II, the PSO-A2C-LNet model consistently stands out. On the Panama Energy Dataset, it achieves the highest coefficient of determination ( $R^2$ ) at 0.88, indicating strong predictive accuracy, along with the lowest mean absolute percentage error (MAPE) of 1.9% and the smallest mean absolute error (MAE) of 7.3, making it the top-performing model. In the ERCOT Dataset, PSO-A2C-LNet also delivers competitive results with an ( $R^2$ ) of 0.87, a MAPE of 2.1%, and a MAE of 7.5. Similarly, on the RTE Dataset, it outperforms other models with an ( $R^2$ ) score of 0.86, a lower MAPE of 2.0%, and a MAE of 7.5. These consistent results suggest that PSO-A2C-LNet exhibits robust predictive capabilities across diverse datasets.

While PSO-A2C-LNet excels on all datasets, the other models exhibit varying levels of performance. These comparative results emphasise the importance of model selection based on the specific dataset and application, with PSO-A2C-LNet emerging as a robust choice for diverse predictive tasks.

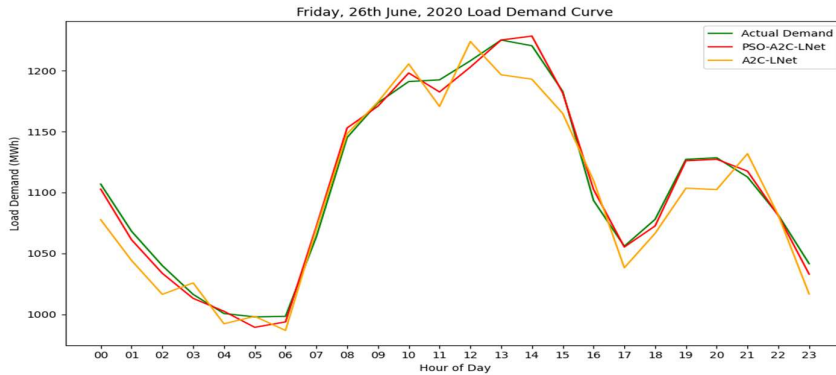
#### 4.1 Comparison of results with results in literature

Table III shows the results of the A2C-LNet and the PSO-A2C-LNet on the testing dataset compared to other models in scientific literature.



**Table III**  
**Comparison of MAPE Results with Literature**

Model	Best MAPE(%) Results
A2C-LNet	2.8097
<b>PSO-A2C-LNet (Our Proposed Model)</b>	<b>1.9376</b>
Integrated CNN and LSTM Network [20]	3.49
LSTM network considering attention mechanism [10]	2.26
ANN-IEAMCGM-R [21]	3.59
TCN-LightGBM [22]	2.64
nonAda-GWN [23]	7.42
Ada-GWN [23]	6.83
Stacked XGB-LGBM-MLP [23]	2.69
GRU-CNN Hybrid Neural Network Model [24]	2.8839



## 5. CONCLUSIONS

In conclusion, this research paper has introduced a novel neural network architecture for short-term load forecasting, amalgamating Convolutional Neural Network and Long Short-Term Memory models, reinforced by a Multi-Head Attention Mechanism. Empirical assessments confirm its superiority over traditional methods and standalone neural network models, with demonstrated applicability to real-world datasets.

Future work will focus on optimising the proposed architecture, exploring further hyperparameter tuning, and investigating additional data preprocessing techniques for enhanced forecasting. Additionally, integrating robust data privacy measures, such as federated learning or secure enclaves, into the architecture is essential to address emerging privacy concerns in load forecasting, ensuring secure and privacy-preserving predictions while advancing the scalability and adaptability of the framework to diverse forecasting challenges and datasets.

## DECLARATION OF COMPETING INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

## ACKNOWLEDGEMENTS

The authors would like to thank Professor Philip Yaw Okyere for guiding the research.

## REFERENCES

- [1] D. Frederick and A. E. Selase, "The effect of electric power fluctuations on the profitability and competitiveness of smes: A study of smes within the accra business district of Ghana," *Journal of Cryptology*, vol. 6, pp. 32–48, 2014.
- [2] N. D. Rao and S. Pachauri, "Energy access and living standards: Some observations on recent trends," *Environmental Research Letters*, vol. 12, no. 2, p. 025 011, 2017.
- [3] T. M. Letcher, 11-storing electrical energy, editor (s): Trevor M. Letcher, managing global warming, 2019.
- [4] P. Jiang, F. Liu, and Y. Song, "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting," *Energy*, vol. 119, pp. 694–709, 2017.
- [5] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renewable and sustainable energy reviews*, vol. 39, pp. 748–764, 2014.
- [6] D. Ahmed, M. Ebeed, A. Ali, A. S. Alghamdi, and S. Kamel, "Multi-objective energy management of a micro-grid considering stochastic nature of load and renewable energy resources," *Electronics*, vol. 10, no. 4, p. 403, 2021.
- [7] Y. Chen, P. B. Luh, C. Guan, et al., "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 322–330, 2009.
- [8] Y. Hu, B. Qu, J. Wang, et al., "Short-term load forecasting using multimodal evolutionary algorithm and random vector functional link network based ensemble learning," *Applied Energy*, vol. 285, p. 116 415, 2021.
- [9] Y. Kim, H.-g. Son, and S. Kim, "Short term electricity load forecasting for institutional buildings," *Energy Reports*, vol. 5, pp. 1270–1280, 2019.
- [10] J. Lin, J. Ma, J. Zhu, and Y. Cui, "Short-term load forecasting based on LSTM networks considering attention mechanism," *International Journal of Electrical Power And Energy Systems*, vol. 137, p. 107 818, 2022.
- [11] N. Huang, Q. He, J. Qi, et al., "Multi Nodes interval electric vehicle day-ahead charging load forecasting based on joint adversarial generation," *International Journal of Electrical Power And Energy Systems*, vol. 143, p. 108 404, 2022.
- [12] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using cnn-lstm neural networks," *Energy*, vol. 182, pp. 72–81, 2019.
- [13] Y. Liu, Q. Wang, X. Wang, et al., "Community enhanced graph convolutional networks," *Pattern Recognition Letters*, vol. 138, pp. 462–468, 2020.
- [14] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.
- [15] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.
- [16] R. C. Staudemeyer and E. R. Morris, "Understanding lstm—a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.
- [17] S. Chandar, C. Sankar, E. Vorontsov, S. E. Kahou, and Y. Bengio, "Towards non-saturating recurrent units for modelling long-term dependencies," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3280–3287.
- [18] J.-B. Cordonnier, A. Loukas, and M. Jaggi, "Multihead attention: Collaborate instead of concatenate," *arXiv preprint arXiv:2006.16362*, 2020.

- [19] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *Advances in neural information processing systems*, vol. 32, 2019.
- [20] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated cnn and LSTM network," *IEEE Access*, vol. 9, pp. 32 436–32 448, 2021. DOI: 10.1109/ACCESS.2021.3060654.
- [21] P. Singh, P. Dwivedi, and V. Kant, "A hybrid method based on neural network and improved environmental adaptation method using controlled gaussian mutation with real parameters for short-term load forecasting," *Energy*, vol. 174, pp. 460–477, 2019, ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2019.02.141>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544219303408>.
- [22] Y. Wang, J. Chen, X. Chen, et al., "Short-term load forecasting for industrial customers based on TCN LightGBM," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1984–1997, 2021. DOI: 10.1109/TPWRS.2020.3028133.
- [23] W. Lin, D. Wu, and B. Boulet, "Spatial-temporal residential short-term load forecasting via graph neural networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5373–5384, 2021. DOI: 10.1109/TSG.2021.3093515.
- [24] L. Wu, C. Kong, X. Hao, and W. Chen, "A Short-Term load forecasting method based on GRU-CNN hybrid neural network model," *Mathematical Problems in Engineering*, vol. 2020, p. 1 428 104, Mar. 2020.

#### Authors

**Paapa Kwesi Quansah** is a Machine Learning Researcher and a Teaching and Research Assistant of Electrical/Electronics Engineering at the Kwame Nkrumah University of Science and Technology. He is also the founder of Rune, Inc. a start-up focused on low resource machine learning systems capable of providing personalised recommendations to farmers in the Sub-Saharan region of Africa. His research primarily focuses on learning mechanisms that allow autonomous agents to behave intelligently and coordinate actions for the advancement and use of real-world autonomous systems.



Edwin Kwesi Ansah Tenkorang completed his Bachelor's degree in Electrical/Electronic Engineering in 2022 from the Kwame Nkrumah University of Science and Technology (K.N.U.S.T). He is an aspiring research who has worked on various research projects spanning applied machine learning, wireless network efficiency, power system anomaly detection and networking paradigms. Edwin is currently a Cybersecurity Research Assistant at the University Information Technology Services centre, K.N.U.S.T, where he explores applicable open-source security solutions.

