Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 3, December 2024, pp. 1694~1700
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i3.pp1694-1700

Leveraging transformer models for enhanced temperature forecasting: a comparative analysis in the Beni Mellal region

Hamza Jdi, Noureddine Falih
LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Jul 12, 2024
Revised Aug 8, 2024
Accepted Aug 26, 2024

Keywords:

Machine learning
Recurrent neural networks
Temperature forecast
Temporal fusion transformers
Time series

ABSTRACT

The remarkable impact of transformers in artificial intelligence, exemplified by applications like GPT-3 in language processing, has sparked interest in their potential for time series analysis. This study aims to explore whether transformers, specifically temporal fusion transformers (TFT), can outperform conventional methods in this domain. The research question is whether TFT exhibits superior performance compared to conventional recurrent neural network (RNN) methods, specifically gated recurrent unit (GRU), and traditional machine learning approaches, notably autoregressive integrated moving average (ARIMA), in the context of time series analysis and temperature prediction. A comparative analysis is conducted among three models: ARIMA, GRU, and TFT. The study utilizes time series data spanning from 1984 to the end of 2022. The models' performances are evaluated using multiple metrics: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R2). The TFT model achieves the lowest MAE, indicating high accuracy in its predictions. It outperforms both the RNN and traditional machine learning in temperature prediction tasks. Integrating the TFT model with the FAO penman-monteith method could improve irrigation scheduling due to more accurate temperature predictions, potentially enhancing water efficiency and crop yields.

Corresponding Author:

Hamza Jdi
LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University
Mghila BP: 592, Beni Mellal, Morocco
Email: hamzajdi@gmail.com

1. INTRODUCTION

The arrival of transformers has brought about a paradigm shift in artificial intelligence, exemplified by the remarkable success of the generative pre-trained transformer 3 (GPT-3) in natural language processing [1], [2]. GPT-3 has demonstrated unparalleled proficiency in generating coherent and contextually relevant text, finding applications in chatbots, content creation, language translation, and text summarization.

Beyond GPT-3, another influential transformer model in the realm of natural language processing is bidirectional encoder representations from transformers (BERT) [3]. BERT's bidirectional approach to understanding word context, considering both left and right contexts, has significantly improved its performance in tasks such as question answering [4] and sentiment analysis [5].

While transformers have established themselves as powerful tools in NLP and language-related tasks, the exploration of their potential in time series analysis is ongoing. Traditional methods like autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM) have been predominant in this domain, but transformers, with their self-attention mechanism [6], show promise in capturing long-range dependencies in sequences. This suggests the potential for transformers to enhance time series forecasting and anomaly detection.

In the realm of computer vision, transformers are not confined to language-related tasks. Vision transformers (ViT) have emerged as successful models for image classification [7], [8]. ViT's approach of dividing images into patches, linearly embedding them, and processing them through transformer layers highlights the adaptability of transformers across diverse data types.

While their efficacy is well-established in diverse domains, their potential in time series analysis remains a subject of exploration. This study aims to address this by examining the performance of transformers, specifically focusing on temporal fusion transformers (TFT) models, in comparison to conventional methods such as recurrent neural network (RNN), with a focus on gated recurrent unit (GRU), and traditional machine learning approaches, particularly ARIMA. For this purpose, we shall use a time series spanning from 1984 to the end of 2022. Our primary objective is to conduct a comprehensive comparative analysis among the three models ARIMA, GRU, and TFT using established evaluation metrics, including mean absolute error (MAE) [9], root mean square error (RMSE) [10], mean absolute percentage error (MAPE) [11], and the coefficient of determination (R2) [12]. Through this investigation, we aim to shed light on the potential superiority of transformers in handling time series data for temperature prediction.

We present a summary of research carried out using TFT. Known for their transformative capabilities, they have become instrumental in reshaping the processing and modeling of diverse data types. This includes time-series data [13] such as financial market fluctuations, weather pattern changes, and biological signals, as well as spatial data like geographical information, textual data from documents and social media, and multimedia data such as images and videos, across various fields.

Zhu and Tang [14] try exploring the use of TFT, LSTM, and linear regression time series models for the purpose of strengthening the prediction strategy of meteorological factors weather monitoring by comparing the advantages and disadvantages.

Heerden *et al.* [15] applies the TFT directly to raw wind power output data, revealing its performance comparable to widely used RNN-based approaches to predict wind power. Notably, the model's effectiveness diminishes ten to twelve hours after the last input.

As for this study [16] the authors introduce an innovative forecasting approach integrating prominent methods of breaking down, interpretable multifactor predicting models, and optimization algorithms. The suggested model utilizes variational mode decomposition to dissect the raw wind speed sequence into intrinsic mode functions. Adaptive differential evolution is employed to optimize various parameters of TFT model, resulting in a forecasting model with satisfactory performance.

A novel technique utilizing the TFT for concurrent prediction of departure and arrival delays at multiple airports is introduced by Wang *et al.* [17]. This method effectively captures intricate time-based patterns of available input variables at the moment of prediction, enabling forecasts of specific delay metrics up to four hours ahead. To enhance weather data processing, the team developed a self-supervised learning model that condenses complex weather information into a more compact format suitable for TFT training. Preliminary results suggest that the TFT-based delay prediction model performs well, demonstrating reduced forecasting errors when tested on the validation dataset.

Santos *et al.* [18] seeks to forecast hourly day-ahead PV power generation by employing the TFT, an attention-based architecture offering interpretability in temporal dynamics and high-performance forecasting across various horizons. The proposed model underwent training and testing with data from six facilities in Germany and Australia. Comparative analysis against algorithms such as ARIMA, LSTM, multi-layer perceptron (MLP), and extreme gradient boosting (XGBoost) was conducted using statistical error measures. The results demonstrate that TFT overtakes the other algorithms in accurately forecasting PV generation across the specified facilities.

In the context of traffic flow forecasting, the transformer model presented in article [19] utilizes multi-head attention and undergoes a comparative assessment against a GRU and a LSTM model using the performance measurement system dataset. With five heads and five identical layers for both the encoder and decoder, the model incorporates square subsequent masking techniques. The findings underscore the transformer-based model's proficiency in accurately predicting extended-term traffic flow patterns when supplied with a significant volume of data.

Lime *et al.* [20] introduces TFT, an attention-based architecture with interpretable insights into temporal dynamics. TFT employs recurrent layers for local processing to capture temporal relationships at various scales and interpretable self-attention layers for long-term dependencies. Dedicated components are utilized for feature selection, and a series of gating layers effectively suppress needless components, ensuring superior performance across a diverse range of situations.

Wu and Wang [21] propose a wind speed prediction model with goals including enhanced decomposition techniques, intelligent optimization using the JADE algorithm, and interpretability. The model, incorporating the TFT framework, is empirically validated for real-world effectiveness.

Jdi and Falih [22] compared four models, namely simple RNN, LSTM, GRU, and ARIMA. The study was conducted on the region of Beni Mellal, Morocco using a 22-year daily collected record. The results showed that the deep learning models, namely GRU and LSTM, outperformed the machine-learning model, ARIMA, with the GRU model achieving the lowest MAE

As evident from the cited literature, TFT have demonstrated success and outperformed both RNN and traditional machine learning models. Its attention-based approach and interpretable insights into temporal dynamics prove valuable in predicting various fields such as traffic flow, air quality, PV power forecasting, wind speed, wind power forecasting, and multi-airport delays. However, the literature highlights a gap in the implementation of TFT in weather forecasting, particularly in predicting temperature. This work builds upon our previous work [22]-[24], where we focus on predicting temperature by comparing RNN and ARIMA models from various perspectives, we concluded that RNN had the upper hand based on MAE. The question arises: can transformers, through TFT, offer greater accuracy when it comes to temperature forecasting?

2. METHOD

In this study, our objective is to predict temperatures using the TFT model and subsequently compare its performance against the results obtained from RNN and ARIMA models from our previeus studies [22], [23]. The data utilized in this analysis is sourced from the modern-era retrospective analysis for research and applications, version 2 (MERRA-2), a state-of-the-art NASA atmospheric reanalysis product. MERRA-2 integrates various satellite observations and numerical models to provide a comprehensive, global dataset of atmospheric conditions. This reanalysis product is particularly valuable for climate studies due to its consistent methodology and extensive temporal coverage.

For our study, we extract daily temperature data from MERRA-2 for a specific geographical location (latitude: 32.33, longitude: -6.37). This location corresponds to the Beni Mellal region in Morocco [25], as shown in Figure 1.

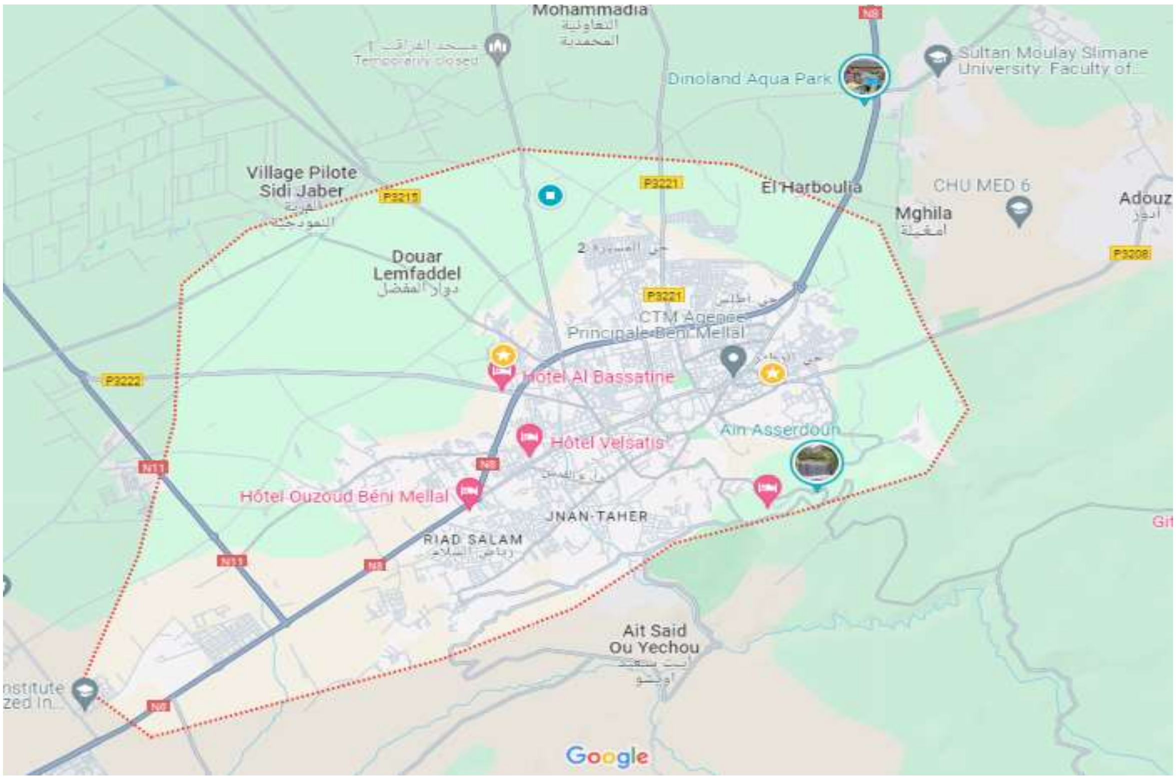

Figure 1. Google Maps screenshot of the Beni Mellal region, Morocco

The TFT model is fed a CSV file containing two primary columns: DATE and temperature. The DATE column follows the ISO 8601 format (YYYY-MM-DD) to ensure precise temporal referencing.

The temperature values represent daily averages in Celsius, computed from hourly data to capture diurnal variations while maintaining a manageable dataset size.

Our time series spans more than 38 years, encompassing the period from January 1, 1984, to December 31, 2022. This extensive temporal range allows for a comprehensive analysis of long-term temperature trends and patterns, including the potential effects of climate change and various climatic oscillations (e.g., El Niño Southern Oscillation, North Atlantic Oscillation).

Figure 2 illustrates the complete dataset through a time series plotted graph, showcasing the temporal variations in temperature over the study period. This visualization not only provides an intuitive understanding of the data’s structure but also reveals key features such as:

- Seasonal cycles: clearly visible annual temperature patterns.
- Long-term trends: any gradual increase or decrease in temperatures over the decades.
- Extreme events: notable temperature anomalies or heat waves.
- Inter-annual variability: year-to-year differences in temperature patterns.

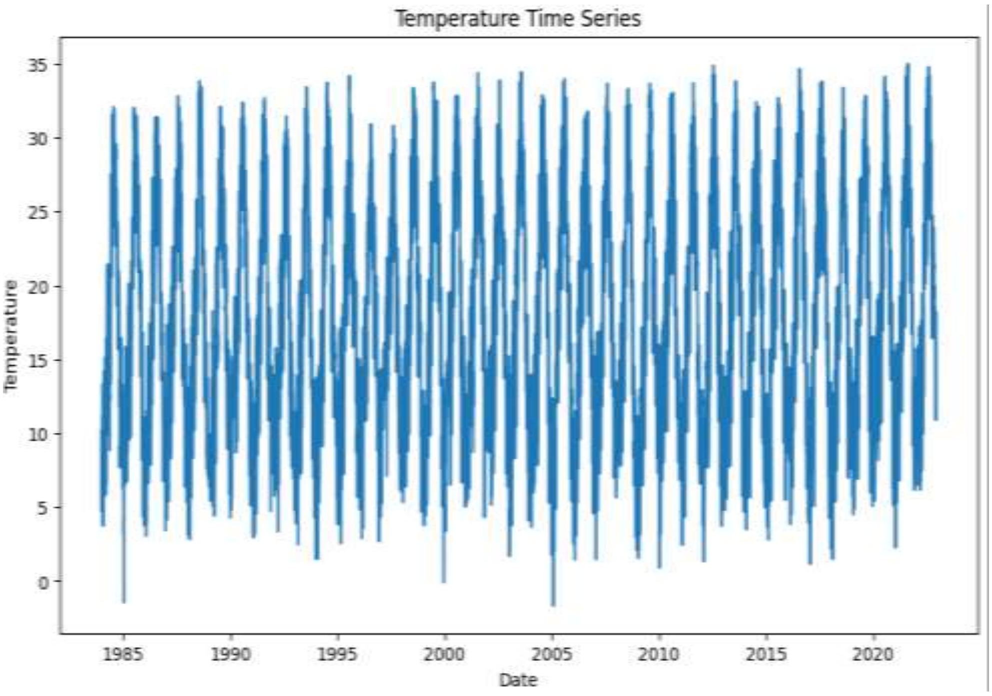

Figure 2. Evolution of daily average temperatures in Beni Mellal, Morocco (1984-2022)

The richness of this dataset, combined with its extensive temporal coverage, provides a robust foundation for training and evaluating our TFT model. It allows us to capture both short-term weather fluctuations and long-term climate trends, setting the stage for potentially accurate and insightful temperature predictions.

To ensure robustness in our model evaluation, the dataset is partitioned into a training set (80%) and a testing set (20%). This split allows for effective model training while reserving a significant portion for performance assessment. After extensive experimentation with various hyperparameters, we found that the results are most satisfactory when employing 10 epochs for the TFT model. This optimal epoch number balances model convergence and computational efficiency. To comprehensively evaluate the TFT model’s performance and compare it with the observed data, we utilize a suite of statistical metrics. These include:

- MAE: measures the average magnitude of errors in the predictions.
- RMSE: provides a quadratic scoring rule that also measures the average magnitude of the error.
- MAPE: expresses accuracy as a percentage, allowing for relative comparisons.
- Coefficient of determination: indicates the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

In our experiments, we employ Google Collab connected to a Google Compute Engine backend with a Python 3 environment. This cloud-based setup ensures reproducibility and scalability of our analysis. The system features a GPU, which significantly accelerates the training process of the TFT model, especially beneficial for processing the extensive 38-year dataset. The available 12.67 GB of RAM facilitates efficient data handling and model operations, while the 78.19 GB of disk space allows for storage of intermediate results and model checkpoints.

3. RESULTS AND DISCUSSION

Our study reveals that the TFT model achieves the lowest MAE, indicating high accuracy in its predictions. It outperforms both the GRU and ARIMA models from our previous studies [22], [23] representing an improvement upon our earlier work, suggesting that the TFT model captures the underlying patterns in temperature fluctuations more effectively. To reach these conclusions, we used a time series of daily average temperatures in Beni Mellal, measured in Celsius. The dataset comprises 14,244 records, divided into a training set (80%, 11,395 records) and a testing set (20%, 2,848 records). To assess prediction accuracy, we primarily use the MAE, supplemented by RMSE, MAPE, and R2. We employ a TFT model for our analysis. Table 1 outlines the parameters used for the TFT model. These parameters provide a good starting point for researchers aiming to predict temperatures in the Beni Mellal region. Following the training of our TFT model with these parameters, we evaluate its performance on the testing set to assess its predictive capabilities on unseen data.

Table 1. Detailed breakdown of parameters in TFT model

Class	Parameter	Description	Value
Positional encoding	d_model	The dimension of the model	64
	dropout	The dropout value	0.1
	max_len	The maximum length	5000
Transformer model	input_dim	The dimension of the input	1
	d_model	The dimension of the model	64
	nhead	The number of heads in the multi head attention models	4
	num_layers	The number of sub-encoder-layers in the transformer encoder	2
	dropout	The dropout value	0.2

We present, in Table 2, the performance metrics of three different models: TFT, RNN, and ARIMA. These metrics include the MAE, RMSE, R2 and MAPE. The TFT model has a MAE of 1.5143, indicating that on average, the model's predictions are approximately 1.5143 Celsius away from the actual values. The RMSE for the TFT model is 1.9527, suggesting that the standard deviation of the prediction errors is around 1.9527. The MAPE is 11.6156, which means that on average, the percentage error of the model's predictions is approximately 11.6156%. Lastly, the R2 score for the TFT model is 0.9359, which is quite high and suggests that this model explains 93.59% of the variance in the target variable.

Table 2. Comparative performance metrics of TFT, simple RNN, GRU, LSTM, and ARIMA models

Model	MAE	RMSE	MAPE	R2	Predicted period
TFT	1.5143	1.9527	11.6156	0.9359	15-03-2015 To 12-31-2022
TFT	1.5743	2.0131	10.5374	0.9314	01-01-2020 To 12-31-2022
GRU [23]	3.0446	-	4.5547	-	01-01-2020 To 12-31-2022
ARIMA [23]	3.0906	-	-	-	01-01-2020 To 12-31-2022
GRU [22]	2.961	-	4.439	-	01-01-2020 To 12-31-2022
LSTM [22]	2.973	-	4.456	-	01-01-2020 To 12-31-2022
Simple RNN [22]	2.986	-	4.471	-	01-01-2020 To 12-31-2022
ARIMA [22]	3.112	-	-	-	01-01-2020 To 12-31-2022

The TFT model exhibits superior performance with an MAE of 1.5143. This suggests that the predictions made by the TFT model are, on an average, the closest to the actual values. In contrast, to the pervious prediction made by GRU, LSTM, Simple RNN and ARIMA models shared in Table 2 these models demonstrate higher MAEs, the lowest MAE is obtained by GRU [19] indicating a larger average deviation in their predictions.

When we consider the MAPE for the TFT and GRU models, the GRU model surpasses the TFT model with a lower MAPE of 4.5547. This implies that the predictions of the GRU model are, on average, more accurate in terms of percentage error. The TFT model, with a higher MAPE of 11.6156, shows a larger average percentage error in its predictions.

Overall, the TFT model appears to perform well based on these metrics, with low error rates and a high R2 score. Let's present in the Figure 3 the results of the TFT model's temperature predictions for the time frame starting January 1, 2020, and ending on December 31, 2022.

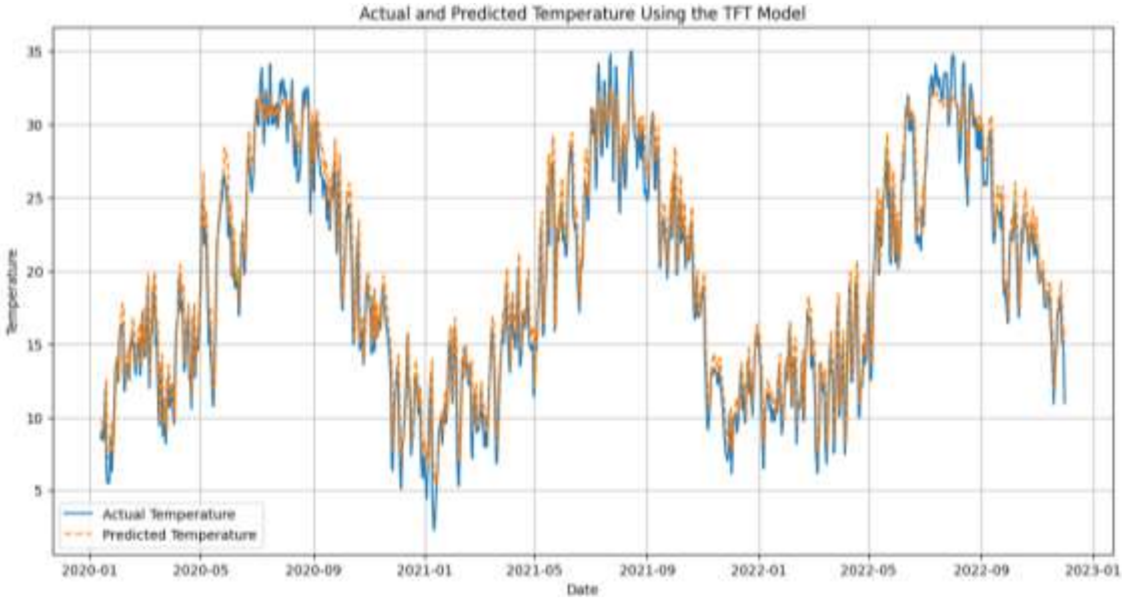

Figure 3. Beni Mellal's actual and forecasted daily temperature using TFT

4. CONCLUSION

This paper focuses on employing the TFT model to predict temperatures three years into the future. Additionally, it compares the TFT model with RNNs and ARIMA to determine whether transformers can outperform traditional RNNs and other machine learning algorithms. The findings demonstrate the superior performance of the TFT model over RNNs and other machine learning methods in forecasting long-term average temperatures in Beni Mellal, Morocco. The accurate temperature predictions achieved by the TFT model show great potential for optimizing irrigation in the region. In future research, we will utilize these enhanced forecasts to calculate reference evapotranspiration (ETo) using the FAO Penman-Monteith method, where temperature is one of the key weather parameters. This integration of TFT predictions of weather parameters with ETo calculations can potentially lead to more efficient ETo forecast, thereby improving water management in the region and contributing to sustainable agricultural practices in Beni Mellal.

REFERENCES

[1] C. Grévisse, "Comparative quality analysis of GPT-based multiple choice question generation," in *Communications in Computer and Information Science*, vol. 1874 CCIS, 2024, pp. 435–447.

[2] J. Behrmann, E. M. Hong, S. Meledathu, A. Leiter, M. Povelaitis, and M. Mitre, "Chat generative pre-trained transformer's performance on dermatology-specific questions and its implications in medical education," *Journal of Medical Artificial Intelligence*, vol. 6, pp. 16–16, Sep. 2023, doi: 10.21037/jmai-23-47.

[3] J. Xiong *et al.*, "Efficient reinforcement learning-based method for plagiarism detection boosted by a population-based algorithm for pretraining weights," *Expert Systems with Applications*, vol. 238, p. 122088, Mar. 2024, doi: 10.1016/j.eswa.2023.122088.

[4] J. Yang *et al.*, "BERT and hierarchical cross attention-based question answering over bridge inspection knowledge graph," *Expert Systems with Applications*, vol. 233, p. 120896, Dec. 2023, doi: 10.1016/j.eswa.2023.120896.

[5] I. Amellal, A. Amellal, H. Seghiouer, and M. R. Ech-Charrat, "An integrated approach for modern supply chain management: Utilizing advanced machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction," *Decision Science Letters*, vol. 13, no. 1, pp. 237–248, 2024, doi: 10.5267/j.dsl.2023.9.003.

[6] N. Fishman and L. Hancox-Li, "Should attention be all we need? The epistemic and ethical implications of unification in machine learning," in *ACM International Conference Proceeding Series*, Jun. 2022, pp. 1516–1527, doi: 10.1145/3531146.3533206.

[7] M. Xu, W. Zhou, X. Shen, J. Qiu, and D. Li, "Temporal-spatial cross attention network for recognizing imagined characters," *Scientific Reports*, vol. 14, no. 1, p. 15432, Jul. 2024, doi: 10.1038/s41598-024-59263-5.

[8] H. Yao, T. Liu, G. Li, and C. Qin, "Screen-shot and Demoiréd image identification based on DenseNet and DeepViT," *Expert Systems with Applications*, vol. 240, p. 122580, Apr. 2024, doi: 10.1016/j.eswa.2023.122580.

[9] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Information Sciences*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/j.ins.2021.11.036.

[10] Q. Li *et al.*, "LandBench 1.0: A benchmark dataset and evaluation metrics for data-driven land surface variables prediction," *Expert Systems with Applications*, vol. 243, p. 122917, Jun. 2024, doi: 10.1016/j.eswa.2023.122917.

[11] M. A. Rubi, S. Chowdhury, A. A. A. Rahman, A. Meero, N. M. Zayed, and K. M. Anwarul Islam, “Fitting multi-layer feed forward neural network and autoregressive integrated moving average for dhaka stock exchange price predicting,” *Emerging Science Journal*, vol. 6, no. 5, pp. 1046–1061, Aug. 2022, doi: 10.28991/ESJ-2022-06-05-09.
[12] X. Zhang and D. Hedeker, “Defining R-squared measures for mixed-effects location scale models,” *Statistics in Medicine*, vol. 41, no. 22, pp. 4467–4483, Sep. 2022, doi: 10.1002/sim.9521.
[13] H. J. Jin, Y. P. Zhao, and M. N. Pan, “A novel method for aero-engine time-series forecasting based on multi-resolution transformer,” *Expert Systems with Applications*, vol. 255, p. 124597, Dec. 2024, doi: 10.1016/j.eswa.2024.124597.
[14] C. Zhu and Y. Tang, “An empirical analysis of the long short term memory and temporal fusion transformer models on regional air quality forecast,” in *ICCSI 2023 - 2023 International Conference on Cyber-Physical Social Intelligence*, Oct. 2023, pp. 497–502, doi: 10.1109/ICCSI58851.2023.10303941.
[15] L. Van Heerden, C. Van Staden, and H. J. Vermeulen, “Temporal fusion transformer for day-ahead wind power forecasting in the South African context,” in *Proceedings - 2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe, EEEIC / I and CPS Europe 2023*, Jun. 2023, pp. 1–5, doi: 10.1109/EEEIC/ICPSEurope57605.2023.10194737.
[16] B. Wu, L. Wang, and Y. R. Zeng, “Interpretable wind speed prediction with multivariate time series and temporal fusion transformers,” *Energy*, vol. 252, p. 123990, Aug. 2022, doi: 10.1016/j.energy.2022.123990.
[17] L. Wang, A. Tien, J. Chou, and W. Hubbard, “Multi-airport delay prediction with transformers,” Jun. 2022, doi: 10.2514/6.2022-3707.
[18] M. L. Santos, X. García-Santiago, F. E. Camarero, G. B. Gil, and P. C. Ortega, “Application of temporal fusion transformer for day-ahead PV power forecasting,” *Energies*, vol. 15, no. 14, p. 5232, Jul. 2022, doi: 10.3390/en15145232.
[19] S. Reza, M. C. Ferreira, J. J. M. Machado, and J. M. R. S. Tavares, “A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks,” *Expert Systems with Applications*, vol. 202, p. 117275, Sep. 2022, doi: 10.1016/j.eswa.2022.117275.
[20] B. Lim, S. Arık, N. Loeff, and T. Pfister, “Temporal fusion transformers for interpretable multi-horizon time series forecasting,” *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.012.
[21] B. Wu and L. Wang, “Two-stage decomposition and temporal fusion transformers for interpretable wind speed forecasting,” *Energy*, vol. 288, p. 129728, Feb. 2024, doi: 10.1016/j.energy.2023.129728.
[22] H. Jdi and N. Falih, “Comparison of time series temperature prediction with auto-regressive integrated moving average and recurrent neural network,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, pp. 1770–1778, Apr. 2024, doi: 10.11591/ijece.v14i2.pp1770-1778.
[23] H. Jdi and N. Falih, “Long-term average temperature forecast using machine learning and deep learning in the region of Beni Mellal,” in *Lecture Notes in Networks and Systems*, vol. 656 LNNS, 2023, pp. 26–34.
[24] H. Jdi and N. Falih, “Weather forecast using sliding window algorithm based on hadoop and mapreduce,” in *Lecture Notes in Networks and Systems*, vol. 357 LNNS, 2022, pp. 122–132.
[25] H. Jdi and N. Falih, “Precipitation forecasting using machine learning in the region of Beni Mellal-Khenifra,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 31, no. 1, pp. 451–458, Jul. 2023, doi: 10.11591/ijeecs.v31.i1.pp451-458.

BIOGRAPHIES OF AUTHORS

Hamza Jdi has a M.Sc. in Business Intelligence, received in 2019 from the Faculty of Science and Technologies, University Sultan Moulay Slimane, Beni Mellal, Morocco. Currently, he is a Ph.D. student in the Polydisciplinary Faculty of Beni Mellal. His research interest is about big data analytics, business intelligence, computer science, digital agriculture, and deep learning. He can be contacted at email: hamzajdi@gmail.com.

Noureddine Falih has a Ph.D. on Computer Science, received from the Faculty of Sciences and Technologies of Mohammedia, Morocco in 2013. He is an associate professor since 2014 in the Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco. He has 18 years of professional experience in several renowned companies. His research topics are about information system governance, business intelligence, big data analytics, and digital agriculture. He can be contacted at email: nourfald@yahoo.fr.