

Evaluating hybrid and standard deep learning models for maximum temperature forecasting in a semi-arid region

Oussama Zemnazi¹, Sanaa El Filali¹, Sara Ouahabi¹, Abderrahim Mouhtadi²

¹Laboratory of Artificial Intelligence and Systems, Faculty of Sciences Ben M'Sick, University Hassan II – Casablanca, Casablanca, Morocco

²General Directorate of Meteorology, Casablanca, Morocco

Article Info

Article history:

Received Jul 28, 2025

Revised Feb 27, 2026

Accepted Mar 4, 2026

Keywords:

Deep learning

Hybrid models

Semi-arid climates

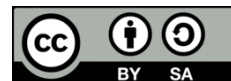
Temperature forecasting

Time series

ABSTRACT

Temperature forecasting is important for industries affected by climate, especially in semi-arid regions where the weather can change quickly and is hard to predict over time. Many studies have examined various deep learning models, including long short-term memory (LSTM), gated recurrent unit (GRU), convolutional neural networks (CNNs), and transformer-based hybrids. However, their performance in data-limited semi-arid environments is often unclear and inconsistent. This study compares six deep learning methods for predicting daily maximum temperatures in Settat, Morocco. It uses 11 years of ground-observed meteorological data. The models examined include a baseline artificial neural network (ANN) and five hybrid structures: ANN-LSTM, ANN-GRU, ANN-CNN, ANN-random forest (RF), and ANN-transformer. The results indicate that the ANN performs the best overall, with MAE = 0.0432, root mean square error (RMSE) = 0.0543, and $R^2 = 0.8820$. It surpasses all hybrid models. When using a relative improvement metric, the ANN shows accuracy gains of 32% to 42% compared to the recurrent, convolutional, and attention-based hybrids. These results suggest that in semi-arid climates, where maximum temperature mainly depends on the same-day atmospheric conditions, simpler feedforward models work better than more complex temporal models. The study underscores the need to match model complexity with climatic factors and dataset size, offering a useful benchmark for temperature forecasting in regions with limited data.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Oussama Zemnazi

Laboratory of Artificial Intelligence and Systems, Faculty of Sciences Ben M'Sick

University Hassan II – Casablanca

Casablanca, Morocco

Email: oussama.zemnazi-etu@etu.univh2c.ma

1. INTRODUCTION

Temperature forecasting plays a critical role in agriculture, water resource management, energy planning, and climate-related decision-making. Its importance is particularly pronounced in semi-arid regions, where meteorological conditions are characterized by strong day-to-day variability, large thermal amplitudes, and irregular atmospheric behavior. These characteristics reduce long-term predictability and make temperature evolution largely dependent on same-day atmospheric conditions, as documented in studies on daily temperature variability and climate extremes [1].

Traditional numerical weather prediction (NWP) models rely on complex physical equations and large-scale simulations; however, their performance is often limited in data-sparse regions and remains

highly sensitive to initial conditions [2]. As a result, data-driven approaches based on machine learning and deep learning have emerged as promising alternatives capable of capturing non-linear relationships directly from historical meteorological observations [3]–[6].

In recent years, several deep learning applications to temperature and weather prediction have been proposed. Recurrent neural networks (RNNs), in particular long short-term memory (LSTM) and gated recurrent unit (GRU) networks, are commonly used for their capacity to model time dependency among sequential data [7], [8]. Their predictability studies and limitations in prediction applications have been well examined in previous works [9]. Convolutional neural networks (CNNs) have also been considered in meteorological applications, especially in weather radar and nowcasting tasks, where localized spatial or feature-level patterns play an important role [10], [11]. More recently, attention-based architectures such as transformers have demonstrated strong performance in long-sequence forecasting problems and have attracted growing interest in weather and climate prediction research [12]–[14]. Nevertheless, the effectiveness of these models remains strongly dependent on dataset size, climatic stability, and the temporal structure of the target variable.

Despite their growing popularity, deep learning-based approaches exhibit several limitations reported in the literature. The majority of previous work has focused on temperate, urban, or data-rich regions, where meteorological variables show relatively stable periodic behavior. Such conditions differ substantially from semi-arid climates, where temperature fluctuations are more erratic and less governed by long-term temporal correlations. Furthermore, many studies rely on reanalysis products or gridded datasets that do not adequately represent station-based observations in data-sparse regions. Consequently, the effectiveness of complex hybrid architectures such as ANN–LSTM, ANN–GRU, ANN–CNN, or transformer-based models remains unclear in semi-arid environments. In addition, heterogeneous preprocessing strategies, feature sets, and evaluation protocols are commonly employed, making fair and consistent comparisons between models difficult and often leading to inconclusive findings regarding the superiority of complex models over simpler ones [3], [15], [16]. This lack of unified benchmarking represents a clear research gap.

To provide background for the present work, existing deep learning-based temperature forecasting studies are summarized and compared. Table 1 presents a non-exhaustive selection of representative studies, chosen to illustrate major methodological trends and commonly reported limitations in deep learning-based temperature and weather forecasting, rather than to provide a comprehensive survey. This comparison highlights the absence of systematic assessments conducted under identical experimental conditions, particularly for station-based data in semi-arid regions.

Table 1. Representative deep learning-based temperature and weather forecasting studies

Year	Method	Data type	Strengths	Limitations
2019	LSTM, GRU [15]	Multivariate time series	Effective temporal modeling	Requires large datasets; sensitive to climatic irregularity
2020	CNN-based nowcasting [10], [11]	Radar meteorological data	Captures localized spatial patterns	Limited suitability for station-level tabular data
2021	Transformer/Informer [13], [14]	Long-sequence time series	Strong long-range dependency modeling	High data requirements; sensitivity to noise
2020	WeatherBench [17]	Global gridded reanalysis	Unified benchmarking framework	Not designed for station-based semi-arid regions
2018	Machine learning for climate and weather [3]	Global climate datasets	Highlights machine learning potential	Limited evaluation in data-scarce environments

The objective of this study is to conduct a unified and comprehensive comparison of six deep learning models for daily maximum temperature forecasting in a semi-arid climate. Using 11 years of ground-observed meteorological data from Settat, Morocco, the performance of a baseline artificial neural network (ANN) is evaluated against five hybrid architectures: ANN–LSTM, ANN–GRU, ANN–CNN, ANN–random forest (RF), and ANN–transformer. All models are assessed under identical preprocessing procedures, feature sets, and evaluation metrics to determine whether increased model complexity leads to improved forecasting accuracy in a data-limited semi-arid context.

The main contributions of this study are summarized as follows:

- A unified and systematic benchmarking framework for comparing multiple deep learning architectures under identical preprocessing procedures, feature sets, and evaluation metrics.

- An empirical evaluation demonstrating that a simple feedforward ANN can outperform more complex hybrid architectures in daily maximum temperature forecasting for a semi-arid climate.
- A quantitative analysis of relative performance differences, highlighting the limited benefits of temporal and attention-based models in environments characterized by weak multi-day dependencies.
- A climate-aware interpretation of the results, explaining why same-day atmospheric conditions dominate temperature variability in semi-arid regions.
- A practical reference benchmark for future temperature forecasting studies conducted in data-scarce, station-based semi-arid environments.

The remainder of this paper is structured as follows. In section 2 reviews related work on temperature prediction using machine learning and deep learning approaches. In section 3 describes the study area, dataset, preprocessing steps, and model architectures. In section 4 presents the experimental results, while section 5 discusses the findings in relation to semi-arid climate characteristics. Finally, section 6 concludes the paper and outlines directions for future research.

2. RELATED WORKS

2.1. Traditional and machine learning approaches for temperature forecasting

Most of the early works on temperature forecasts were still dominated by classical statistical methods and the NWP model. NWP methods are those performing explicit numerical integration of equations that represent atmospheric thermodynamic and dynamic physics and demand high computational resources along with precise initial data. These models have generally performed well at global and regional scales; however, their performance is still limited when applied to site-level stations, particularly in semi-arid climates characterized by strong meteorological variability [2].

To overcome these limitations, data-driven methods such as machine learning approaches have become more extensively studied. machine learning models can imitate non-linear relationships between input and output from the historical meteorological data without physics-based modeling. However, classical machine learning approaches are still in need of heavy feature engineering and have low generalization performance under non-stationary or highly varying climatic conditions [3].

2.2. Deep learning models for temperature and weather forecasting

Owing to the recent developments of deep learning, models using neural networks have been widely used in temperature and weather prediction problems [18]. RNNs [9], and in particular LSTM networks [7] and GRU [8], are widely used since they can capture the temporal dependencies present in time-series data. These models have achieved promising performance on the tasks where meteorological features can be observed with relatively stable temporal patterns and long-term dependence.

CNNs have also been applied in meteorological applications, especially in weather radar and nowcasting tasks, where localized spatial or feature-level patterns play a crucial role [10], [11]. More recently, attention-based models such as transformers have proven successful for long-sequence prediction problems by explicitly capturing long-range dependencies in self-attention mechanisms [12], [14]. More efficient variants like Informer have also made transformer-based models applicable to long time-series forecasting problems [13]. Despite this progress, however, such architectures are often data greedy and noise sensitive, and hence their use is not widespread in station-based and data-sparse scenarios.

2.3. Hybrid architectures and comparative studies

To enhance the accuracy of forecasting, various approaches such as combining multiple deep learning components (ANN-LSTM, CNN-LSTM) and attention-based RNN models have been proposed. These methods attempt to exploit the complementary nature of strengths, such as the fusion of feature extraction and temporal modeling. However, there have been conflicting results in the literature on whether hybrid models are indeed superior to simpler models. Thorough reviews reveal that model performance is largely influenced by data size, preprocessing methods, feature selection, and evaluation protocols [15], [19].

Large-scale comparative studies further suggest that increased model complexity does not systematically lead to improved forecasting performance [16]. Furthermore, most of the comparative studies are not carried out under homogeneous experimental contexts; therefore, they are unable to directly compare models and generalize the results. Compounding all of this is the fact that a unified benchmarking infrastructure does not exist, so we lack solid numbers about whether these very complex architectures really are consistently better than simpler approaches.

2.4. Limitations of existing studies in semi-arid and station-based contexts

A major limitation of existing deep learning-based forecasting studies is their predominant focus on temperate, urban, or data-rich regions, often relying on large-scale gridded or reanalysis datasets. Benchmark

datasets such as WeatherBench have contributed significantly to standardized evaluation in global weather forecasting; however, they are primarily designed for large-scale gridded data and do not adequately represent station-level observations in semi-arid regions [17].

In semi-arid climates, temperature dynamics are frequently dominated by the same-day atmospheric conditions rather than strong multi-day temporal dependencies. Consequently, the suitability of complex temporal and attention-based architectures in such environments remains uncertain. Furthermore, deep learning models are particularly sensitive to data scarcity, which is a common constraint in station-based meteorological observations [20], [21]. Few studies provide climate-aware interpretations that explicitly link model performance to regional atmospheric characteristics. Motivated by these limitations, the present study conducts a unified and systematic comparison of multiple deep learning architectures for daily maximum temperature forecasting in a semi-arid context. By evaluating a baseline ANN alongside several hybrid models under identical preprocessing procedures, feature sets, and evaluation metrics, this work aims to clarify whether increased model complexity provides tangible benefits in data-limited semi-arid environments.

3. DATA AND METHODS

3.1. Study area and data description

The study area is Settat, which is a city situated in the central part of Morocco and has a semi-arid climate with marked seasons, high inter-annual variability, and an unpredictable atmosphere. These properties reduce the predictability range in the long term and induce the temperature variation to be mostly affected by the current-day atmospheric condition, which is problematic for data-driven forecasting models. The meteorological data were provided by the Moroccan General Directorate of Meteorology (DGM) through official station observations.

This work uses the 11-year daily meteorological record measured at a local weather station. The study examines daily maximum temperature, a key variable for climate-sensitive sectors such as agriculture, water resource management, and energy planning. Station-based observations are retained in order to maintain local climatic variability that is many times smoothed or even lost in larger-scale gridded data sets or reanalyses.

3.2. Data preprocessing

Meteorological data underwent a number of preprocessing steps before development of the model to achieve uniformity and reliability. Standardized quality-control procedures were used to handle missing or implausible values. For numerical stability purposes and to enable fair comparison among the various model architectures, all input variables were then normalized to a common scale.

To preserve the temporal integrity of the dataset and prevent information leakage, the data were split chronologically into training, validation, and testing subsets to ensure robust model tuning and unbiased performance evaluation [9]. All preprocessing was applied equally in all models, ensuring that any performance differences are due to the model properties rather than differences in preprocessing. The overall flowchart of the proposed workflow, which consists of data collection, preprocessing, model development, and performance assessment, is presented in Figure 1.

3.3. Model architectures

A total of six deep learning models were evaluated. A basic ANN was used to design a baseline model. This ANN denotes a forward neural network and doesn't require any direct modeling of temporal dependency. Furthermore, five other hybrid models were developed to combine multiple capabilities. These included ANN-LSTM and ANN-GRU models to use their ability to learn from temporal dependency, ANN-CNN to leverage localized extraction capabilities, ANN-RF to use its potential of an ensemble-based approach, and lastly ANN-transformer to use its potential of an attention-based learning strategy.

All models were designed to predict daily maximum temperature using an identical set of input features. This design ensures a consistent experimental framework and enables a direct assessment of whether increasing architectural complexity leads to improved predictive performance under semi-arid climatic conditions.

3.4. Model training and evaluation

The experiments were carried out using Python 3.13 within a Jupyter Notebook environment. The implementation used the TensorFlow deep learning framework, including other scientific libraries like NumPy, Pandas, and Scikit-learn for numerical computation, data preprocessing, and model evaluation. Experiments were conducted on a Lenovo Yoga laptop equipped with an Intel® Core™ i7-10510U 10th generation CPU (1.80 GHz up to 2.30 GHz) and 16 GB RAM in a 64-bit OS.

To preserve temporal dependencies intact and to avoid data leakage, the meteorological dataset (2014-2024) was chronologically split into 70% training, 15% validation, and 15% test subsets. The MinMaxScaler parameters were fitted only on the train data and used to transform validation and test sets. Gaussian noise-based data augmentation was performed exclusively on the training subset in order to increase robustness and limit overfitting risk.

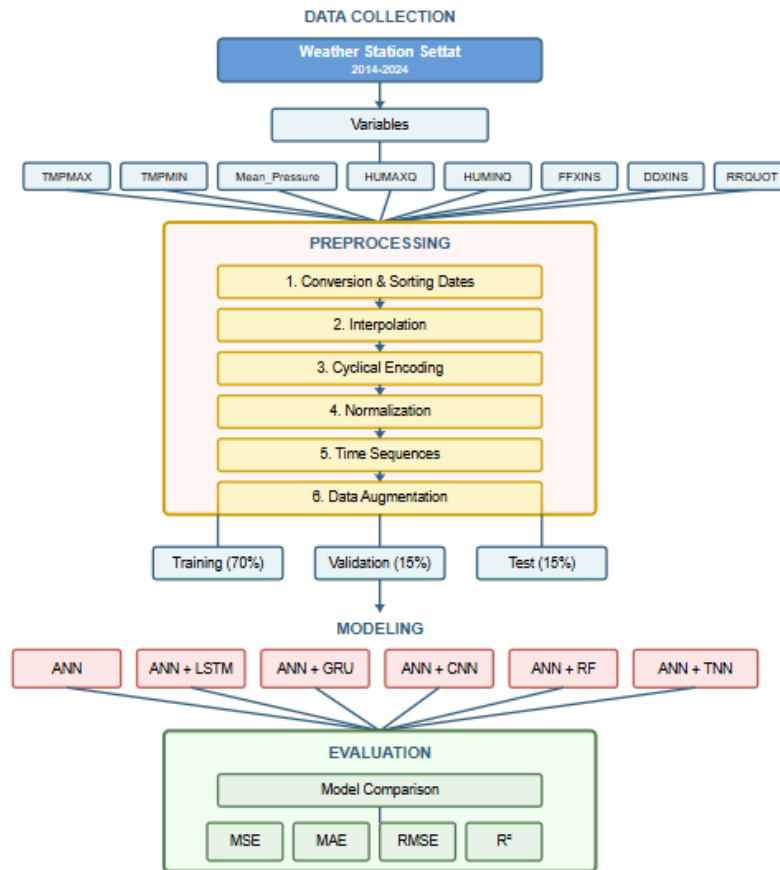


Figure 1. The proposed approach

The Adam optimizer was used for optimization of all neural network architectures. The learning rates were defined as 0.001 for the ANN, ANN-LSTM, ANN-GRU, ANN-CNN, and ANN-transformer models, while a learning rate of 0.004 was used in the case of the hybrid model (ANN-RF). Mean squared error (MSE) was applied as a training loss function, while mean absolute error (MAE) was tracked during the training. For all neural architectures, the batch size was equal to 32. The standalone ANN model contains three fully connected layers with ReLU activation and 64, 32, and 16 neurons and a linear output layer.

Batch normalization and dropout regularization were used to enhance generalization. ANN-LSTM and ANN-GRU hybrid architectures use an ANN backbone where the recurrent layers have 100 and then 50 units with tanh activation to capture temporal dependencies before feature concatenation and dense prediction layers at the end. ANN-CNN has a sequential stacking of two Conv1D layers with 64 and 32 filters of kernel size equal to 3, respectively, followed by feature flattening and dense prediction layers. The ANN-RF hybrid extracts intermediate ANN features from a dense layer of 32 neurons and uses them as input to a RF configured with 200 trees, a maximum depth of 15, and a minimum sample split equal to 5. The ANN-transformer architecture consists of a multi-head attention (2 heads, key dimension 16) followed by GlobalAveragePooling and dense prediction layers.

The training period differed for each architecture; the ANN-based hybrid was run for 150 epochs, the LSTM-, GRU-, and CNN-based hybrids were run for up to 100 epochs each, and the transformer model was limited to a maximum of 50 epochs. Early stopping and ReduceLR OnPlateau callbacks were systematically applied with patience values between 10 and 20 epochs and a minimum learning rate of 1e-6. The performance of models is quantified in terms of MAE, root mean square error (RMSE), MSE, and the

coefficient of determination (R^2), which are the most widely adopted metrics in temperature forecasting studies [22]–[24]. To further quantify the performance differences, we computed a relative improvement metric by comparing each hybrid architecture to the baseline ANN. This evaluation framework allows for a systematic consideration of the utility of model complexity in data-limited semi-arid environments.

4. RESULTS

4.1. Overall model performance

The performance of the models was evaluated by MAE, MSE, RMSE, and the R^2 on the testing dataset. The quantitative results for the baseline ANN and the five hybrid models, including ANN–LSTM, ANN–GRU, ANN–CNN, ANN–RF, and ANN–transformer, are shown in Table 2. We trained all models using the same preprocessing steps, feature set, and data splitting method to be able to compare them fairly.

4.2. Comparative analysis of model performance

The results shown by Table 2 reveal that overall, the hybrid models are inferior to the baseline ANN in terms of all evaluation measures. In particular, ANN presents the lowest MAE, MSE, and RMSE values, as well as the highest R^2 , based on other models considered. The hybrid models incorporating temporal components (ANN–LSTM and ANN–GRU) or attention mechanisms (ANN–transformer) do not yield significant improvements over the simple feedforward ANN. In some cases, these models exhibit slightly degraded performance, suggesting limited benefits from explicitly modeling long-term temporal dependencies in the studied semi-arid context. Similarly, the ANN–CNN and the ANN–RF models do not exceed the baseline ANN, indicating that increased architectural complexity is not necessarily beneficial to improve forecasting when applied to station-based temperature with weak multi-day dependencies.

Table 2. Performance comparison of the evaluated models

Metrics	Method					
	ANN	ANN-LSTM	ANN-GRU	ANN-CNN	ANN-RF	ANN-transformer
MAE	0.0432	0.0644	0.0645	0.0724	0.0742	0.0678
MSE	0.0030	0.0068	0.0069	0.0084	0.0089	0.0077
RMSE	0.0543	0.0827	0.0830	0.0918	0.0941	0.0876
R^2	0.8820	0.7270	0.7251	0.6638	0.6467	0.6933

4.3. Observed versus predicted temperature analysis

To more deeply investigate the ability of each model, Figures 2 to 7 show a comparison between observed and forecasted daily maximum temperatures for all evaluated models in the testing period. These plots provide a visual assessment of each model's ability to reproduce temporal variations and extreme temperature values. Visual observation indicates that the ANN model gives predictions closely tracking the observed temperature variations with a smoother dispersion and better alignment under normal and extreme conditions. The discrepancies with the observed data are larger for the hybrid models, especially under rapidly changing temperature conditions. Temporal or attentional-based architectures tend to smooth short-term oscillations, which can lead to delayed or attenuated responses to abrupt atmospheric variations.

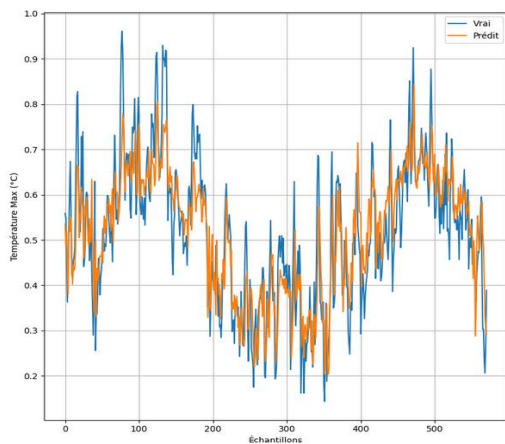


Figure 2. ANN-LSTM temperature: real vs. pred

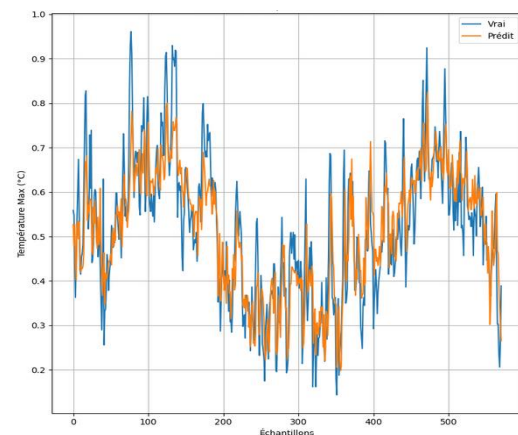


Figure 3. ANN-GRU temperature: real vs. pred

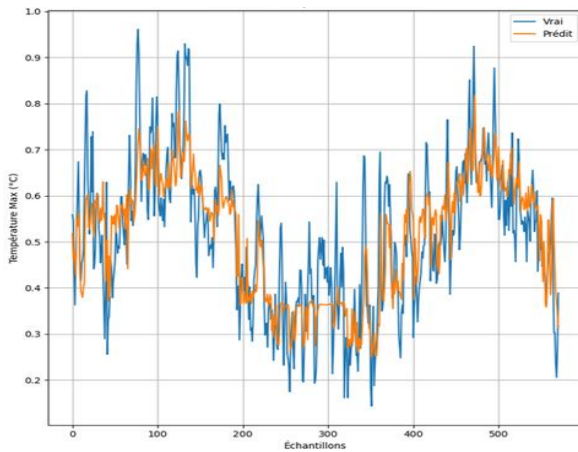


Figure 4. ANN-CNN temperature: real vs. pred

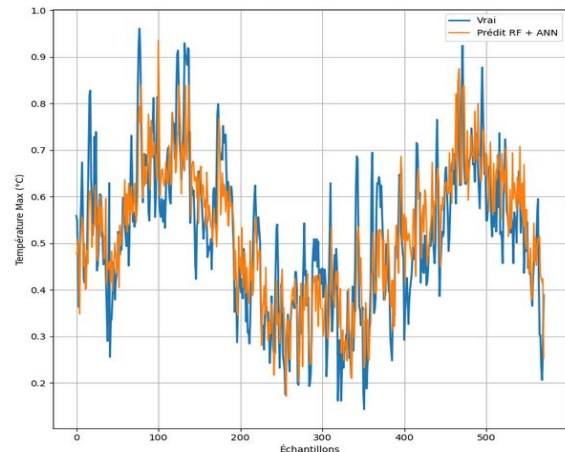


Figure 5. ANN-RF temperature: real vs. pred

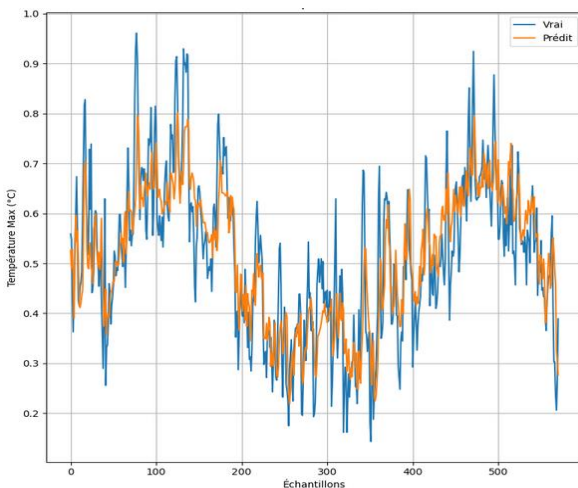


Figure 6. ANN-transformer temperature: real vs. pred

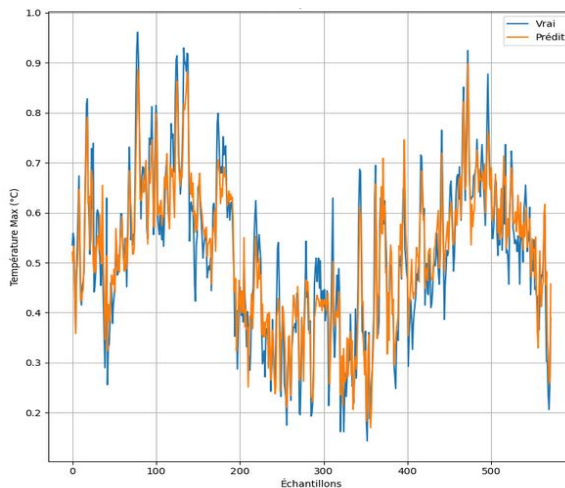


Figure 7. ANN temperature: real vs. pred

4.4. Summary of results

In general, the results indicate that the simple feedforward ANN is the most accurate and stable model to predict daily maximum temperature when compared with all models considered. Hybrid architectures that involve temporal, convolutional, or attention-based components do not provide noticeable performance improvements in the semi-arid region, despite their higher complexity. These findings suggest that, for station-based temperature forecasting in data-scarce semi-arid regions, same-day atmospheric information plays a dominant role, and increasing model complexity may not be justified.

5. DISCUSSION

5.1. Overall model performance comparison

The experimental results reveal that the baseline ANN consistently outperforms more complex hybrid architectures in daily maximum temperature forecasting. Though hybrid models, including the ANN–LSTM, ANN–GRU, ANN–CNN, and ANN–transformer, are simplified to learn temporal or hierarchical dynamics, their improvement effects are also weak for the present semi-arid, station-based context. These findings are consistent with earlier literature that suggests more complex models in general do not always lead to better forecasting accuracy if data size and variability are limited [3], [5], [15], [16].

These results are also in line with the general forecasting literature, where simpler models often compare favorably or even outperform state-of-the-art methods when fitted under limited data scenarios [16]. It appears that in semi-arid regions this relationship between atmospheric predictors and maximum

temperature can be approximated well by same-day nonlinear transformations, rather than more complicated temporal dependencies.

5.2. Why simpler models outperform complex architectures

A key explanation for the superior performance of ANN is due to the climatic conditions of semi-arid areas. Temperature variability in such environments is frequently dominated by short-term atmospheric conditions rather than strong multi-day temporal persistence. Observational studies have shown that daily temperature extremes exhibit high variability and weak autocorrelation in dry and semi-arid climates [1]. Under these conditions, recurrent and attention-based architectures, which are explicitly designed to exploit long-term temporal dependencies, may provide limited added value.

Moreover, recent climate and Earth system science studies emphasize that deep learning models should be interpreted in the context of underlying physical processes rather than being treated as universally superior black-box solutions [5]. From this perspective, the strong performance of a simple feedforward ANN demonstrates that same-day predictors prevail in controlling temperature evolution under semi-arid climates. This also agrees with findings in other comparative studies between deep learning frameworks and traditional NWP systems, which indicate that deep learning does not outperform simpler or physics-based models unequivocally across all cases [4].

5.3. Impact of data scarcity and temporal dependency

Data availability is also another key factor that impacts the performance of the model. Recurrent and transformer-based architectures have been shown to require a large amount of data to learn the temporal structures effectively and avoid overfitting [9], [14]. In the current work, the application of 11 years of station-level observations may also not fully exploit the representational capacity of more complex architectures, despite being enough to train simpler ones.

This limitation is further supported by benchmarking efforts such as WeatherBench, which demonstrate strong deep learning performance primarily in data-rich, gridded reanalysis settings rather than station-based environments [9]. Consequently, the inferior performance of hybrid-based models and attention-based models observed in this work should not be considered as an indication or shortcoming of the approach taken but rather as evidence for insufficient data to match model complexity.

5.4. Practical implications and study limitations

Some practical implications could be drawn from the findings of this study. First, they imply that in data-limited semi-arid regions for which operational temperature prediction is crucial, simpler models ANN can offer reliable and robust forecasts with lower computational cost and reduced risk of overfitting. This is consistent with the general opinion in both NWP and machine learning, which advocate for context-aware model selection rather than default reliance on highly complex architectures [2], [3].

However, there are some limitations in this study. The study is limited to one meteorological station and only deals with the daily maximum temperature. The generalization of these observations across sites, longer periods, and more climate variables should be a focus of future research. In addition, integrating physical constraints or hybrid physics–data-driven approaches may help bridge the gap between model interpretability and predictive performance, as suggested by recent advances in Earth system science [5].

6. CONCLUSION

This study presented a unified and systematic comparison of six deep learning architectures for daily maximum temperature forecasting in a semi-arid, station-based context. Using 11 years of ground-observed meteorological data from Settat, Morocco, the performance of a baseline ANN was evaluated against five hybrid models, namely ANN–LSTM, ANN–GRU, ANN–CNN, ANN–RF, and ANN–transformer, under identical preprocessing procedures, feature sets, and evaluation metrics.

The experimental results demonstrate that the simple feedforward ANN consistently outperforms the more complex hybrid architectures across all evaluation metrics. These findings indicate that, in semi-arid climates, daily maximum temperature variability is largely governed by the same-day atmospheric conditions rather than strong multi-day temporal dependencies. As a consequence, models specifically designed to capture long-term temporal patterns or attention-based relationships do not necessarily provide performance gains when applied to data-limited station observations.

Beyond model accuracy, this study demonstrates a need to match model complexity with climatological attributes and data availability. The results confirm that additional architectural complexity may not offer any enhanced predictive performance and may bring in undesirable complexities and overfitting concerns when only limited historical data are available. In terms of the operational side, the

ANN, with its high-performance capability, computational speed, and efficiency, is a suitable and practical option for temperature prediction in semi-arid areas.

Despite these advances and contributions, there are several limitations that need to be addressed. The analysis is limited to a single meteorological station and just considers the daily maximum temperature. Future work should extend this framework to multiple stations, longer temporal records, and additional climatic variables. Furthermore, including physical constraints or hybrid physics-informed learning approaches could potentially enhance the interpretability and generalization of models while maintaining the efficiency in computation. Overall, this study establishes a well-defined benchmark for deep learning-based temperature prediction in data-limited semi-arid regions and contains useful insights to select an appropriate model complexity when working on climate-sensitive tasks.

ACKNOWLEDGMENTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Oussama Zemnazi	✓	✓	✓	✓	✓	✓		✓	✓		✓			✓
Sanaa El Filali	✓	✓		✓	✓					✓		✓		✓
Sara Ouahabi	✓	✓		✓	✓					✓		✓		
Abderrahim Mouhtadi						✓	✓	✓		✓				

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the General Directorate of Meteorology (Morocco). Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the corresponding author upon reasonable request and with permission of the General Directorate of Meteorology.




REFERENCES

- [1] L. V. Alexander *et al.*, "Global observed changes in daily climate extremes of temperature and precipitation," *Journal of Geophysical Research: Atmospheres*, vol. 111, no. D5, Mar. 2006, doi: 10.1029/2005JD006290.
- [2] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, Sep. 2015, doi: 10.1038/nature14956.
- [3] P. D. Dueben and P. Bauer, "Challenges and design choices for global weather and climate models based on machine learning," *Geoscientific Model Development*, vol. 11, no. 10, pp. 3999–4009, Oct. 2018, doi: 10.5194/gmd-11-3999-2018.
- [4] M. G. Schultz *et al.*, "Can deep learning beat numerical weather prediction?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200097, Apr. 2021, doi: 10.1098/rsta.2020.0097.
- [5] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019, doi: 10.1038/s41586-019-0912-1.




- [6] D. Rolnick *et al.*, “Tackling climate change with machine learning,” *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–96, Feb. 2023, doi: 10.1145/3485128.
- [7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [8] K. Cho *et al.*, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179.
- [9] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, Jan. 2021, doi: 10.1016/j.ijforecast.2020.06.008.
- [10] J. Cuomo and V. Chandrasekar, “Use of deep learning for weather radar nowcasting,” *Journal of Atmospheric and Oceanic Technology*, pp. 1641–1656, Jul. 2021, doi: 10.1175/JTECH-D-21-0012.1.
- [11] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, “Convolutional LSTM network: a machine learning approach for precipitation nowcasting,” *Advances in Neural Information Processing Systems*, vol. 2015-Janua, pp. 802–810, Sep. 2015.
- [12] A. Vaswani *et al.*, “Attention Is All You Need,” *arXiv preprint*, Aug. 2017.
- [13] H. Zhou *et al.*, “Informer: beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, May 2021, pp. 11106–11115. doi: 10.1609/aaai.v35i12.17325.
- [14] B. Lim, S. Ö. 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.
- [15] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, “Deep learning for time series classification: a review,” *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019, doi: 10.1007/s10618-019-00619-1.
- [16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M4 competition: results, findings, conclusion and way forward,” *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, Oct. 2018, doi: 10.1016/j.ijforecast.2018.06.001.
- [17] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Moutadid, and N. Thuerey, “WeatherBench: a benchmark data set for data-driven weather forecasting,” *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 11, Nov. 2020, doi: 10.1029/2020MS002203.
- [18] S. Haykin, *Neural networks and learning machines*, 3rd ed., vol. 3. Upper Saddle River, New Jersey, USA: Pearson Education, Inc., 2008. doi: 978-0131471399.
- [19] S. B. Taieb and R. J. Hyndman, “A gradient boosting approach to the Kaggle load forecasting competition,” *International Journal of Forecasting*, vol. 30, no. 2, pp. 382–394, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.005.
- [20] H. Hersbach *et al.*, “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.
- [21] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: the state of the art,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, Mar. 1998, doi: 10.1016/S0169-2070(97)00044-7.
- [22] C. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Climate Research*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.
- [23] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed. OTexts, 2021.
- [24] I. T. Jolliffe and D. B. Stephenson, *Forecast verification: a practitioner’s guide in atmospheric science*, 2nd ed. John Wiley & Sons, 2012. doi: 10.1002/9781119960003.

BIOGRAPHIES OF AUTHORS






Oussama Zemnazi    holds a bachelor’s degree in Mathematical and Computer Sciences and a Master’s degree in Data Science and Big Data from Hassan II University of Casablanca, Morocco. He is currently pursuing a doctoral degree in Artificial Intelligence and Data Science at the Faculty of Sciences Ben M’Sik, Hassan II University, and is affiliated with the Laboratory of Artificial Intelligence and Systems. His research interests include artificial intelligence, deep learning, time series forecasting, and decision support systems, with applications in environmental and agricultural contexts. He can be contacted at email: oussama.zemnazi-etu@etu.univh2c.ma.






Sanaa El Filali    is a professor of Computer Science at the Faculty of Sciences Ben M’Sik, Hassan II University of Casablanca, Morocco. She received a Ph.D. in Information Processing in 2006. Her current research interests include the internet of things (IoT), artificial intelligence, machine learning, deep learning, computer vision, natural language processing (NLP), and cybersecurity. She is actively involved in promoting research-industry collaboration and technology transfer. She has received several regional and international distinctions, including the Data-Driven Research Award (MENA, 2021) and the third prize at the International Conference on Microelectronics (2022). She is also the founder of multiple scientific and educational initiatives, such as the AI Club & Dev Community and ABC Math Info. She currently supervises several Ph.D. students and contributes actively to the development of research in intelligent systems and emerging technologies. She can be contacted at email: elifalilis@gmail.com.



Sara Ouahabi    is a habilitated professor (PH) in mathematics and the computer science department of Hassan 2 University and a member of the Computer Science and Information Processing laboratory at faculty of science Ben M'sik. Her research includes artificial intelligence, computer security, semantic web, E-learning, computer communications (Networks), and the IoT. She conducts cutting-edge research in these areas while actively engaging in higher education to educate the next generation of computer professionals. She can be contacted at email: s3.ouahabi@gmail.com.



Abderrahim Mouhtadi    was a senior official at Morocco's Direction Générale de la Météorologie, where he held the position of Head of the Centre National de Prévisions Météorologiques. He played a central role in the management of national weather forecasting systems and in the provision of climate and weather services at the national level. He represented Morocco at the World Meteorological Organization (WMO), actively contributing to regional and international discussions on climate resilience, early warning systems, and atmospheric science. His professional interests included weather forecasting, climate services, artificial weather modification, and the integration of meteorological data into disaster risk reduction strategies. He has participated in several international forums, including COP28, where he took part in expert panels and events dedicated to innovation in climate adaptation and environmental monitoring.