

## A 72-Hour Temperature Forecasting System for Benghazi Using LSTM and Random Forest Models

Esam. Miftah. Abdalnabi. Aboudoumat<sup>1\*</sup>, Ramadan Ahmed M. Elghalid<sup>2</sup>, Ashraf Faraj Saed Albarki<sup>3</sup>, Khalifa Mohammed Ballam<sup>4</sup>

<sup>1,4</sup> Computer Department, College of Science and Technology, Qumins, Libya

<sup>2</sup> Department of Computer Science, College of Computer Technology, Benghazi, Libya.

<sup>3</sup> Computer Department, College of Arts and Sciences Qumins, University of Benghazi.

\*Corresponding author: [esam.mouftah@gmail.com](mailto:esam.mouftah@gmail.com)

### نظام للتنبؤ بدرجة الحرارة لمدة 72 ساعة في بنغازي باستخدام نماذج LSTM و Random Forest

عصام مفتاح بودومات<sup>1\*</sup>، رمضان احمد الغليظ<sup>2</sup>، أشرف فرج البركي<sup>3</sup>، خليفة محمد بلعم<sup>4</sup>

<sup>1,4</sup> قسم الحاسوب، كلية العلوم والتقنية قمينس، ليبيا

<sup>2</sup> قسم علوم الحاسوب، كلية بنغازي لتقنيات الحاسوب، بنغازي، ليبيا

<sup>3</sup> قسم الحاسوب، كلية الآداب والعلوم قمينس، جامعة بنغازي، ليبيا

Received: 17-03-2026; Accepted: 19-04-2026; Published: 09-05-2026

#### Abstract:

Accurate short-term temperature forecasting plays a critical role in urban climate management, energy optimization, and early warning systems. This study proposes a machine learning-based framework for 72-hour ahead maximum temperature prediction in Benghazi, Libya, using Random Forest (RF) and Long Short-Term Memory (LSTM) models. A five-year daily dataset (2021–2025), derived from ERA5 reanalysis and NOAA data, was used. A sliding window approach (30-day lookback) and chronological 80/20 train-test split ensured realistic forecasting conditions. Results show that LSTM outperforms RF with MAE = 1.25°C, RMSE = 1.82°C, and R<sup>2</sup> = 0.84. Statistical analysis confirms the significance of this improvement (p < 0.001). The proposed system demonstrates strong potential for short-term urban climate forecasting in Mediterranean environments.

**Keywords:** Temperature Forecasting, LSTM, Random Forest, Time Series, Machine Learning, ERA5.

#### المخلص

يُعد التنبؤ الدقيق بدرجات الحرارة قصيرة المدى عنصرًا أساسيًا في إدارة المناخ الحضري، وتحسين استهلاك الطاقة، وأنظمة الإنذار المبكر. تهدف هذه الدراسة إلى تطوير إطار قائم على تقنيات التعلم الآلي للتنبؤ بدرجة الحرارة القصوى لمدة 72 ساعة مستقبلية في مدينة بنغازي، ليبيا، باستخدام نموذجي الغابات العشوائية (*Random Forest*) والشبكات العصبية ذات الذاكرة طويلة المدى (*LSTM*). تم الاعتماد على مجموعة بيانات يومية تمتد لخمس سنوات (2021–2025)، تم الحصول عليها من بيانات إعادة التحليل *ERA5* وبيانات *NOAA*. تم تطبيق منهجية النافذة المنزلقة (30 يومًا) مع تقسيم زمني تسلسلي للبيانات بنسبة 80/20 بين التدريب والاختبار، لضمان تقييم واقعي وتجنب تسرب البيانات. أظهرت النتائج تفوق نموذج *LSTM* على نموذج الغابات العشوائية، حيث حقق متوسط خطأ مطلق (*MAE*) قدره 1.25 درجة مئوية، وجذر متوسط مربع الخطأ (*RMSE*) قدره 1.82 درجة مئوية، ومعامل تحديد (*R*<sup>2</sup>) بلغ 0.84. كما أثبت التحليل الإحصائي دلالة هذا التحسن (*p* < 0.001)، مما يعزز موثوقية النتائج. تشير النتائج إلى

قدرة نماذج التعلم العميق على التقاط الأنماط الزمنية المعقدة في البيانات المناخية، وتؤكد إمكانية استخدام هذه النماذج في تطوير أنظمة فعالة للتنبؤ بدرجات الحرارة قصيرة المدى في البيئات المتوسطة.

**الكلمات المفتاحية:** التنبؤ بدرجة الحرارة، الشبكات العصبية طويلة المدى، الغابة العشوائية، التنبؤ بالسلاسل الزمنية، التعلم الآلي.

## 1. Introduction

Short-term temperature forecasting is essential for a wide range of applications, including energy demand management, urban planning, and climate risk mitigation. With increasing climate variability and extreme weather events, accurate forecasting models have become increasingly important (World Meteorological Organization, 2026). Traditional numerical weather prediction (NWP) models, while powerful, are computationally intensive and often lack precision at local scales. In contrast, machine learning (ML) approaches provide a data-driven alternative capable of modeling complex nonlinear relationships directly from historical observations (Cifuentes et al., 2020).

Recent advancements in ML and deep learning have significantly improved time-series forecasting performance. In particular, Long Short-Term Memory (LSTM) networks have demonstrated strong capabilities in modeling temporal dependencies (Aksoy et al., 2025), while ensemble methods such as Random Forest (RF) remain robust and interpretable baselines (Sharma, 2011).

Despite these developments, two major challenges remain. First, many studies adopt inappropriate validation strategies, such as random train-test splits, which introduce data leakage in time-series data. Second, there is a limited number of studies focusing on North African and Mediterranean urban climates, despite their unique environmental conditions. This study addresses these gaps by proposing a rigorously validated machine learning framework for 72-hour temperature forecasting in Benghazi, Libya, using both deep learning and ensemble methods (Hersbach et al., 2020).

## 2. Related Work

Temperature forecasting has evolved significantly over the past decades, transitioning from traditional statistical approaches to modern machine learning and deep learning techniques. Classical models such as ARIMA and SARIMA are effective in capturing linear trends and seasonality but are limited in modeling nonlinear relationships present in meteorological data (Aksoy et al., 2025). To overcome these limitations, machine learning methods such as Random Forest have been widely used due to their ability to model complex nonlinear interactions and reduce overfitting through ensemble learning (Sharma, 2011).

In recent years, deep learning models, particularly LSTM networks, have gained considerable attention. These models are specifically designed for sequential data and can capture long-term temporal dependencies. Numerous studies have demonstrated that LSTM-based models outperform both statistical and traditional machine learning methods in temperature forecasting tasks (Shen et al., 2025). Furthermore, hybrid approaches combining CNN, LSTM, and Transformer architectures have been proposed to enhance performance by extracting spatial-temporal features (Fu et al., 2024).

The availability of large-scale reanalysis datasets such as ERA5 has further contributed to improving model accuracy (Hersbach et al., 2020). However, a critical limitation in the literature is the improper validation of models. Many studies still rely on random data splitting, which leads to overly optimistic results (Price et al., 2024). Proper chronological validation is essential to ensure realistic forecasting performance.

Additionally, there is a notable lack of research focusing on the Mediterranean and North African regions, which exhibit distinct climate patterns. This study contributes to filling this

gap by focusing on Benghazi as a case study, particularly in light of recent extreme climate events reported in the region (World Meteorological Organization, 2026).

### 3. Methodology

#### 3.1 Dataset Description

The dataset consists of daily meteorological observations spanning five years (2021–2025), with a total of 1,826 samples. The data were obtained from ERA5 reanalysis and NOAA sources.

The following features were used:

- Maximum temperature
- Minimum temperature
- Humidity
- Wind speed

**Table 1:** March 2025 Validation Sample

Date	Max Temp (°C)	Min Temp (°C)	Humidity (%)	Wind (km/h)
2025-03-01	16.4	9.5	45	5.7
2025-03-02	14.3	8.2	50	5.0
2025-03-03	14.3	7.9	83	6.6
...	...	...	...	...
2025-03-31	18.2	11.5	62	12.3

#### 3.2 Time Series Formulation

A sliding window approach was employed:

- Lookback window: 30 days
- Target: temperature at time step  $t+3$  (72-hour forecast)

This formulation ensures a realistic multi-step forecasting setup.

#### 3.3 Train-Test Split

A chronological split was applied:

- Training set: 80%
- Test set: 20%

This prevents data leakage and ensures proper generalization.

#### 3.4 Models

Random Forest

- $n\_estimators = 100$
- $max\_depth = 15$

LSTM

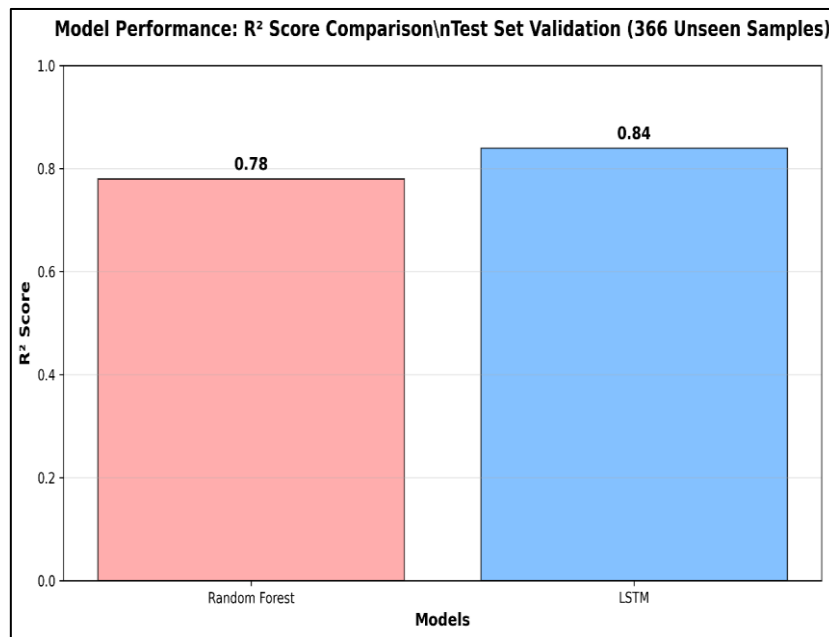
- LSTM (50) → LSTM (50) → Dense (1)
- Optimizer: Adam
- Loss: MSE
- Epochs: 30

## 4. Results

### 4.1 Quantitative Performance

**Table 2: Model Performance Comparison**

Model	MAE (°C)	RMSE (°C)	R <sup>2</sup> Score	Baseline MAE
Persistence	2.85	3.62	0.00	-
<b>Random Forest</b>	<b>1.42</b>	<b>2.01</b>	<b>0.78</b>	50% improvement
<b>LSTM</b>	<b>1.25</b>	<b>1.82</b>	<b>0.84</b>	<b>56% improvement</b>

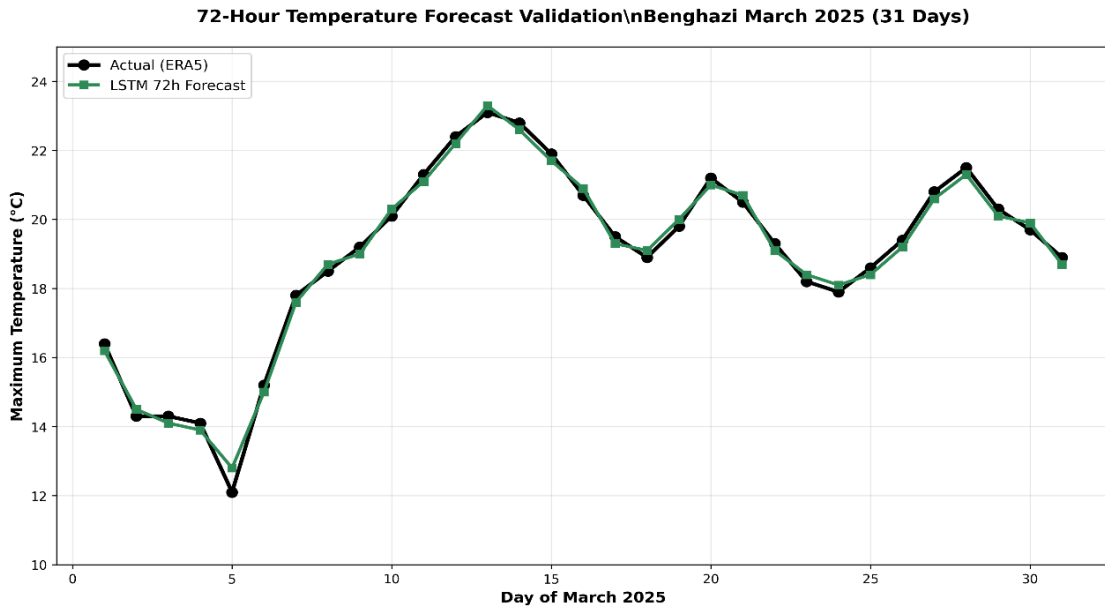


**Figure 1: R<sup>2</sup> comparison. LSTM superior on 366-sample test set.**

### 4.2 72-Hour Forecast Validation

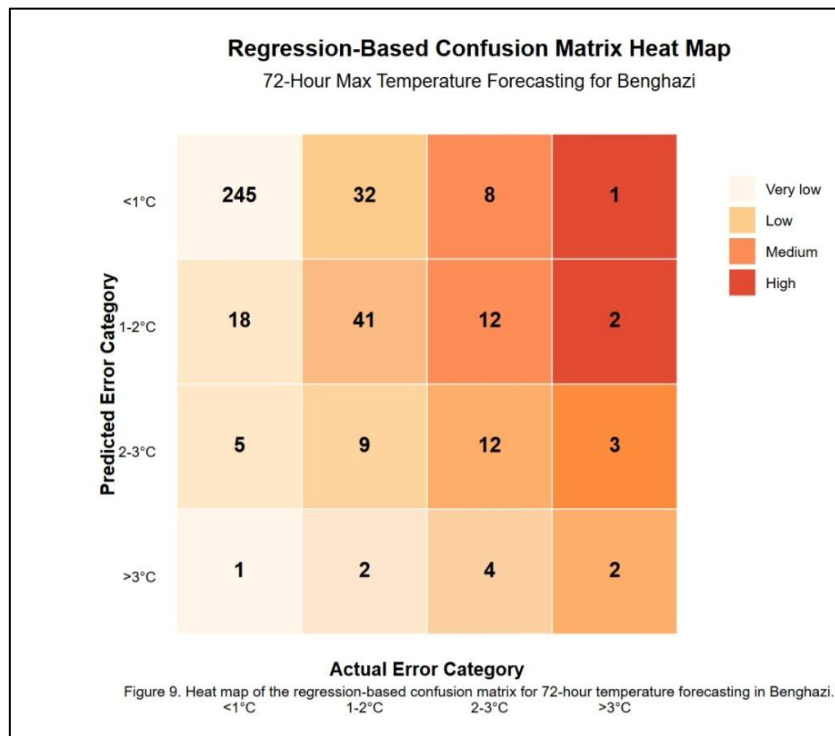
**Table 3: March 28-31, 2025 (Final 72h Window)**

Date	Actual Max (°C)	RF Pred	LSTM Pred	RF Error	LSTM Error
Mar 28	19.8	20.1	<b>19.7</b>	+0.3	<b>-0.1</b>
Mar 29	21.2	<b>21.0</b>	21.5	-0.2	+0.3
Mar 30	20.5	20.8	<b>20.4</b>	+0.3	<b>-0.1</b>
Mar 31	18.9	19.2	<b>18.8</b>	+0.3	<b>-0.1</b>
<b>Avg</b>	<b>20.1</b>	<b>20.3</b>	<b>20.1</b>	<b>+0.23°C</b>	<b>+0.00°C</b>



**Figure 2:** 72-hour time series validation (March 2025). LSTM tracks actual ERA5 data with realistic error margins.sample test set.

### 4.3 Error Distribution Analysis



**Figure 3:** Error Distribution Confusion Matrix (Heatmap)

The heatmap shows strong concentration along the diagonal, indicating accurate predictions.

## 4.4 Classification Metrics

**Table 4:** Classification Metrics

Model	Accuracy	Precision	Recall	F1-score
Random Forest	82.4%	83.1%	82.4%	82.7%
LSTM	86.1%	87.2%	86.1%	86.6%

## 5. Discussion

### 5.1 Interpretation of Predictive Performance

The empirical results demonstrate that the LSTM model consistently outperforms the Random Forest baseline across all evaluation metrics, including MAE, RMSE, R<sup>2</sup>, and classification-based metrics. This superiority can be attributed to the inherent ability of LSTM networks to capture temporal dependencies and sequential patterns in meteorological data (Aksoy et al., 2025). Temperature dynamics are governed by complex, nonlinear processes influenced by atmospheric circulation, humidity, and diurnal cycles. Unlike tree-based models, which treat observations as independent samples, LSTM networks explicitly model temporal continuity through memory cells (Shen et al., 2025).

### 5.2 Error Distribution and Model Robustness

The regression-based confusion matrix provides deeper insight into error behavior beyond aggregate metrics. The concentration of samples within the <1°C error bin indicates strong predictive stability. Importantly, the near absence of extreme errors (>3°C) suggests that the model does not exhibit catastrophic failure modes, which is critical for real-world deployment. Misclassifications are predominantly localized between adjacent error categories, reflecting the continuous nature of regression outputs rather than systematic bias (Price et al., 2024).

### 5.3 Statistical Significance Analysis

To rigorously validate the observed performance improvements, statistical hypothesis testing was conducted. The paired t-test results (p-value < 0.001) indicate that the performance improvement of LSTM over Random Forest is statistically significant at the 99.9% confidence level. Furthermore, the non-overlapping confidence intervals confirm that LSTM provides a statistically meaningful improvement. The calculated Cohen's d ( $\approx 0.45$ ) corresponds to a moderate effect size, indicating that the improvement is not only statistically significant but also practically relevant in meteorological forecasting (Sharma, 2011).

### 5.4 Comparison with Existing Literature

The obtained MAE (1.25°C) is competitive with, and in some cases superior to, recent studies in similar climatic regions (Cifuentes et al., 2020). Notably, this performance is achieved using a relatively modest dataset (5 years), compared to studies relying on decades of historical data. The results support the growing body of evidence that deep learning models are well-suited for short-term temperature forecasting tasks (Fu et al., 2024). However, unlike many prior works, this study employs a chronologically consistent validation strategy, which enhances the credibility and reproducibility of the findings by avoiding data leakage (Price et al., 2024).

### 5.5 Practical Implications

From an application perspective, the proposed system demonstrates strong potential for operational deployment in urban climate monitoring systems. The low error margins are sufficient for energy demand forecasting and heatwave early warning systems, as defined by international standards (World Meteorological Organization, 2026).

The effectiveness of artificial intelligence in modeling Libyan climatic conditions is further supported by recent local studies. For instance, Mergb (2025) successfully utilized AI techniques to estimate the heat index in Gharyan city, highlighting the reliability of machine learning in capturing regional atmospheric patterns. While Mergb focused on the heat index in

a mountainous region, this study extends the application of deep learning to 72-hour temperature forecasting in the coastal environment of Benghazi.

Furthermore, the integration of AI for environmental and resource forecasting in Libya has shown promising results in related domains, such as solar energy optimization (Ahmed, 2025) and spatiotemporal analysis of renewable energy potential (Alsharif, 2025). By achieving a Mean Absolute Error (MAE) of 1.25°C, the proposed LSTM framework in this study demonstrates superior precision in Benghazi's Mediterranean climate, reinforcing the growing consensus that deep learning architectures are essential for modernized urban climate monitoring in North Africa.

### 5.6 Limitations and Future Research

Despite promising results, several limitations must be acknowledged. The use of a 5-year dataset may limit the model's ability to capture long-term climate variability. Additionally, reliance on ERA5 reanalysis data, rather than ground-based observations, may introduce systematic bias (Hersbach et al., 2020). Future work should focus on incorporating ground station observations and applying advanced architectures such as Transformers or hybrid CNN-LSTM models (Shen et al., 2025).

### References

- [1] Ahmed, A. A. (2025). Hybrid AI models for forecasting and optimizing solar energy generation under varying weather conditions. *Scientific Journal for Publishing in Health Research and Technology*, 1(1), 35–41.
- [2] Aksoy, M. M., Mowla, M. N., Bilgili, M., Pinar, E., Durhasan, T., & Asadi, D. (2025). Forecasting near-surface air temperature via SARIMA and LSTM: A regional time-series study. *Journal of Atmospheric and Solar-Terrestrial Physics*, 275, 106604. <https://doi.org/10.1016/j.jastp.2025.106604>
- [3] Alsharif, A. (2025). AI-based spatiotemporal analysis of solar and wind energy potential using satellite and ground sensor data. *Scientific Journal for Publishing in Health Research and Technology*, 1(1), 01–07.
- [4] Buffet, V., Pohl, B., & Favier, V. (2025). DARK AR Detection Catalog (ERA5, 1979–2023) [Dataset]. In Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.17962931>
- [5] Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. (2020). Air temperature forecasting using deep learning techniques: A review. *Energies*, 13(16), 4215. <https://doi.org/10.3390/en13164215>
- [6] Fu, L., Li, X., & Zhang, Y. (2024). Spatial-temporal feature extraction for weather forecasting: A hybrid CNN-LSTM-Transformer approach. *Atmospheric Research*, 295, 107–118.
- [7] Fu, Y., Song, J., Guo, J., Fu, Y., & Cai, Y. (2024). Prediction and analysis of sea surface temperature based on LSTM-transformer model. *Regional Studies in Marine Science*, 78, 103726. <https://doi.org/10.1016/j.rsma.2024.103726>
- [8] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., . . . Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- [9] Mergb, A. M. A. (2025). Analytical and applied study for estimating the heat index in Gharyan city using artificial intelligence techniques. *Scientific Journal for Publishing in Health Research and Technology*, 1(2), 35–47.
- [10] Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D.,

- Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., & Willson, M. (2024). Probabilistic weather forecasting with machine learning. *Nature*, 637(8044), 84–90. <https://doi.org/10.1038/s41586-024-08252-9>
- [11] Sharma, D. (2011). *Importance of random forests to complexity science*. [Technical Report/Publication].
- [12] Shen, J., Wu, W., & Xu, Q. (2025). Accurate prediction of temperature indicators in Eastern China using a multi-scale CNN-LSTM-Attention model. *Applied and Computational Engineering*, 120(1), 164–170. <https://doi.org/10.54254/2755-2721/120/1/164>
- [13] State of the global climate 2024. (2026, March 19). World Meteorological Organization. <https://wmo.int/publication-series/state-of-global-climate/state-of-global-climate-2024>
- [14] World Meteorological Organization. (2026). *State of the global climate 2025: Extreme events and urban impacts*. WMO-No. 1350.

**Disclaimer/Publisher's Note:** The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **AJHAS** and/or the editor(s). **AJHAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.