Convolutional Neural Network for Convective Available Potential Energy (CAPE) Prediction in Morocco: A Meteorological Data Analysis

Mohammed El Abdioui
Abdellah Msaadi

December 16, 2025

Abstract

This report presents a comprehensive analysis of Convective Available Potential Energy (CAPE) prediction in Morocco using meteorological data from 2024. A Convolutional Neural Network (CNN) model was developed to predict CAPE values based on 12 meteorological variables. The dataset comprises 1,835,898 samples with hourly temporal resolution and spatial coverage across Morocco. The implemented CNN model achieved a test R^2 score of 0.5974, RMSE of 148.01 J/kg, and MAE of 37.27 J/kg, representing a 29.9% improvement over previous temporal splitting approaches. The mixed-date splitting strategy provided robust evaluation across different seasons, with Summer showing the best performance (R^2=0.7168) and Spring the worst (R^2=0.3611). Feature importance analysis identified 2m dewpoint temperature (2d) and convective inhibition (cin) as the most significant predictors of CAPE.

Contents

1 Introduction

1.1 Background

Convective Available Potential Energy (CAPE) is a crucial meteorological parameter that measures the amount of energy available for convection. It plays a vital role in thunderstorm development and severe weather forecasting [1]. In Morocco, accurate CAPE prediction is particularly important for agricultural planning, water resource management, and severe weather warning systems.

1.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep neural networks particularly effective for processing structured grid data [2], [5]. While traditionally used for image processing, CNNs have shown excellent performance in meteorological applications [3], [6] due to their ability to:

- Capture spatial hierarchies through convolutional layers
- Learn local patterns and translate them to global features
- Handle multi-dimensional meteorological data efficiently
- Extract meaningful features automatically from raw data

In this study, we apply CNNs to meteorological time series data, treating the temporal dimension as a spatial dimension to capture temporal patterns in atmospheric variables.

2 Dataset Description and Preprocessing

2.1 Data Source and Structure

The dataset originates from meteorological reanalysis data for Morocco covering the entire year 2024 [4]. The original NetCDF file contains:

- **Temporal dimension**: 8,784 hourly time steps (366 days $\times$ 24 hours)
- **Spatial dimension**: 81 latitude points (20°N to 40°N) $\times$ 101 longitude points (-20°W to 5°E)
- **Variables**: 13 meteorological variables as detailed in Table 1

Table 1: Meteorological Variables in the Dataset

Variable	Description	Original Units	Converted Units
10u	10m U wind component	m s^{1}	m/s
10v	10m V wind component	m s^{1}	m/s
2t	2m temperature	K	°C
2d	2m dewpoint temperature	K	°C
msl	Mean sea level pressure	Pa	hPa
tp	Total precipitation	m	mm
tcc	Total cloud cover	0-1	0-1
cp	Convective precipitation	m	mm
lsp	Large-scale precipitation	m	mm
blh	Boundary layer height	m	m
cape	Convective available potential energy	J kg^{1}	J/kg
cin	Convective inhibition	J kg^{1}	J/kg
tco3	Total column ozone	kg m^{2}	kg/m^{2}

2.2 Data Sampling Strategy

To create a manageable dataset for machine learning while preserving representativeness, we implemented a stratified sampling approach:

- **Temporal sampling**: 10% of time points randomly selected (878 out of 8,784 hours)
- **Spatial sampling**: Every 2nd latitude and longitude point (41×51 grid)
- **Final dataset**: 1,835,898 samples (2.55% of original data)
- **NaN handling**: CIN values with 95.06% missingness were imputed with 0

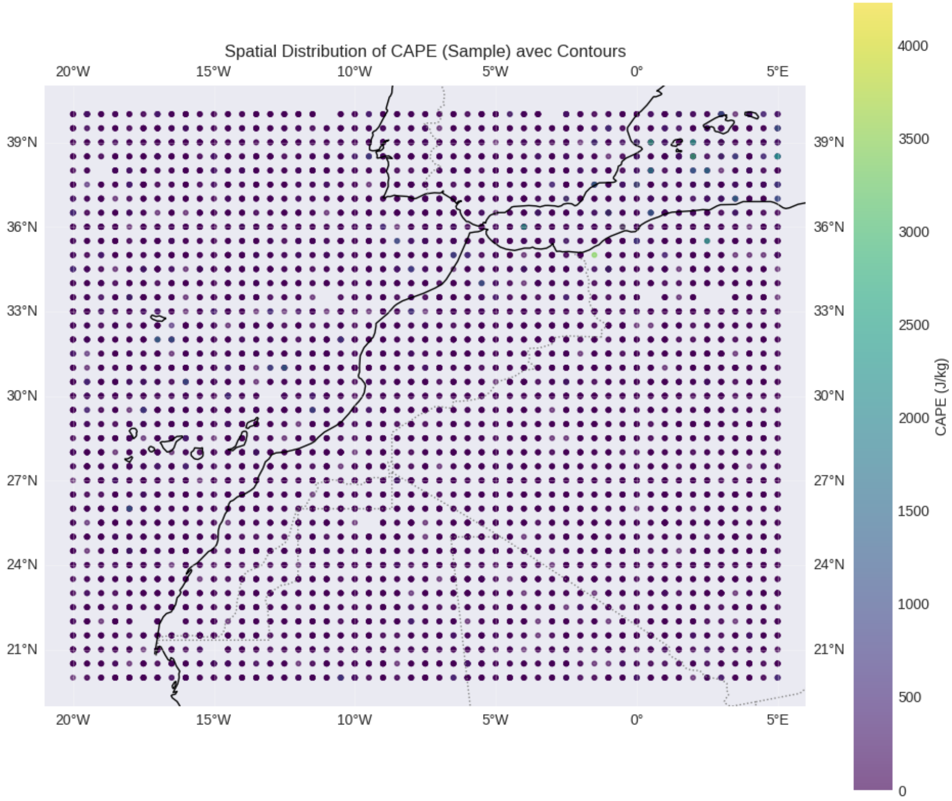
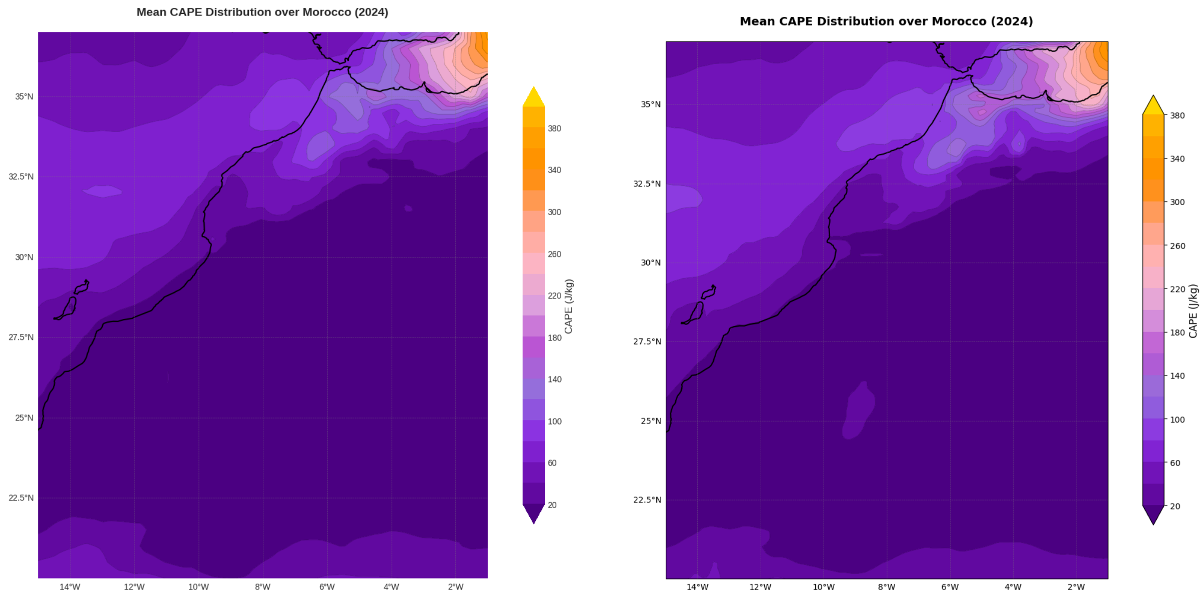


Figure 1: Spatial distribution of sampled data points over Morocco

2.3 Spatial Analysis of CAPE - Cohérence des Jeux de Données

Pour évaluer la cohérence de notre sous-échantillon (ou nouveau jeu de données) avec le fichier NetCDF initial, nous comparons la distribution spatiale moyenne du CAPE dérivée de chaque source. La similarité des motifs et des échelles de couleur sur les deux figures atteste que l'échantillon extrait est représentatif de l'ensemble de données initial.



(a) Moyenne CAPE à partir du Sous-Échantillon (ou Nouveau Dataset)

(b) Moyenne CAPE à partir du Fichier NetCDF Initial

Figure 2: Comparaison des distributions moyennes de CAPE pour vérifier la cohérence du jeu de données.

Key observations from spatial analysis:

- La similarité visuelle entre les deux figures confirme la *cohérence* de notre jeu de données échantillonné.
- Les valeurs moyennes de CAPE les plus élevées sont observées dans les régions du *Nord-est* (tons oranges et roses).
- Les zones côtières et l'océan Atlantique présentent des valeurs de CAPE *nettement inférieures* (tons violets foncés).
- Les schémas spatiaux sont cohérents avec les zones climatiques connues du Maroc, indiquant un potentiel accru de convection à l'intérieur des terres.

2.4 Temporal Patterns

Temporal analysis revealed significant seasonal patterns in CAPE distribution across Morocco. The monthly analysis shows clear variations in CAPE values throughout the year:

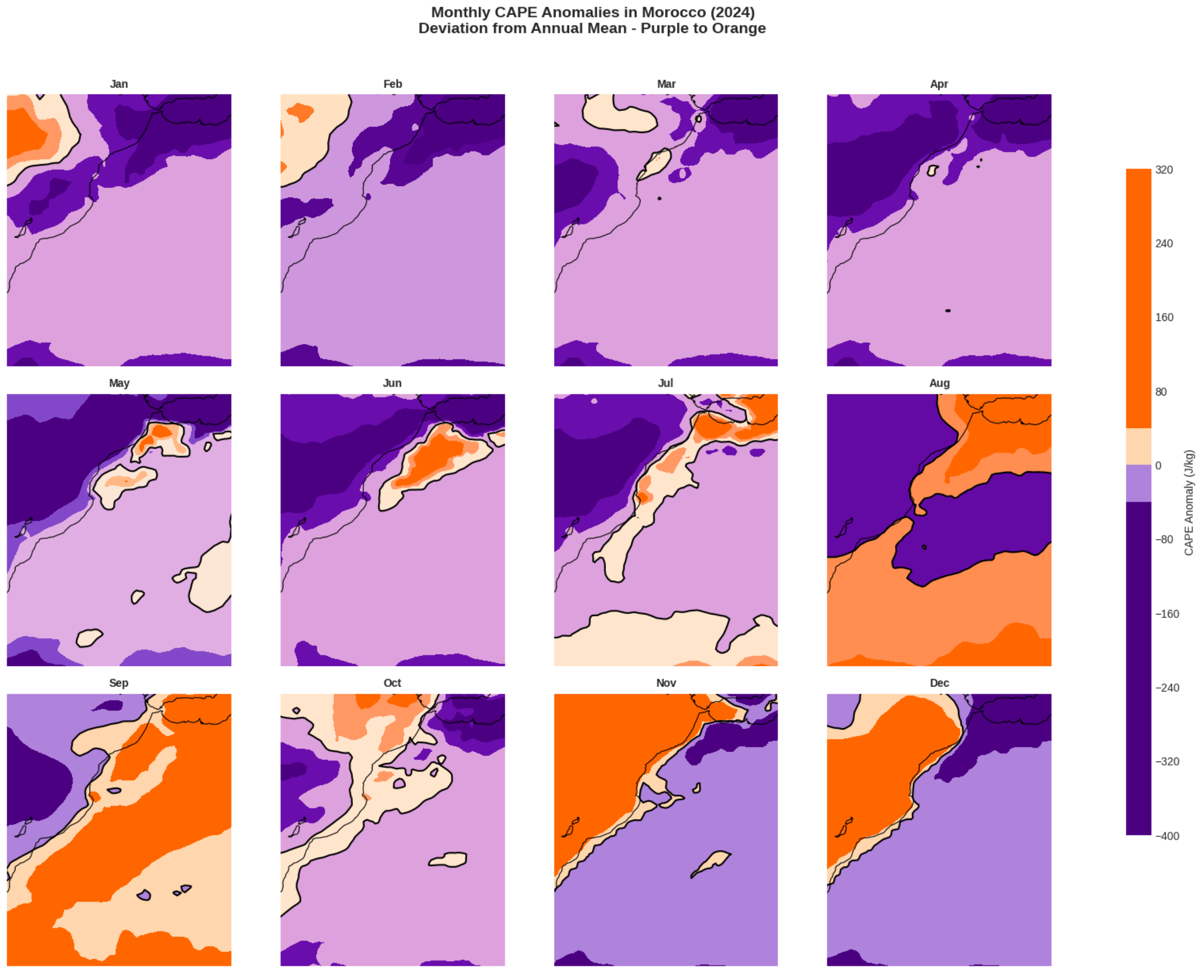


Figure 3: CAPE distribution across all 12 months in Morocco (2024) showing seasonal variations

The 12-month grid visualization reveals the following patterns:

- **Winter months (Dec-Feb)**: Lower CAPE values across most regions, particularly in coastal areas
- **Spring transition (Mar-May)**: Gradual increase in CAPE, especially in south-eastern regions
- **Summer peak (Jun-Aug)**: Maximum CAPE values, with hotspots in southeastern Morocco reaching over 200 J/kg
- **Fall transition (Sep-Nov)**: Gradual decrease in CAPE values back to winter levels

This monthly analysis demonstrates the strong seasonal dependence of convective potential in Morocco, with summer months exhibiting the most favorable conditions for convective development.

2.5 Correlation Analysis

Correlation analysis identified relationships between variables:

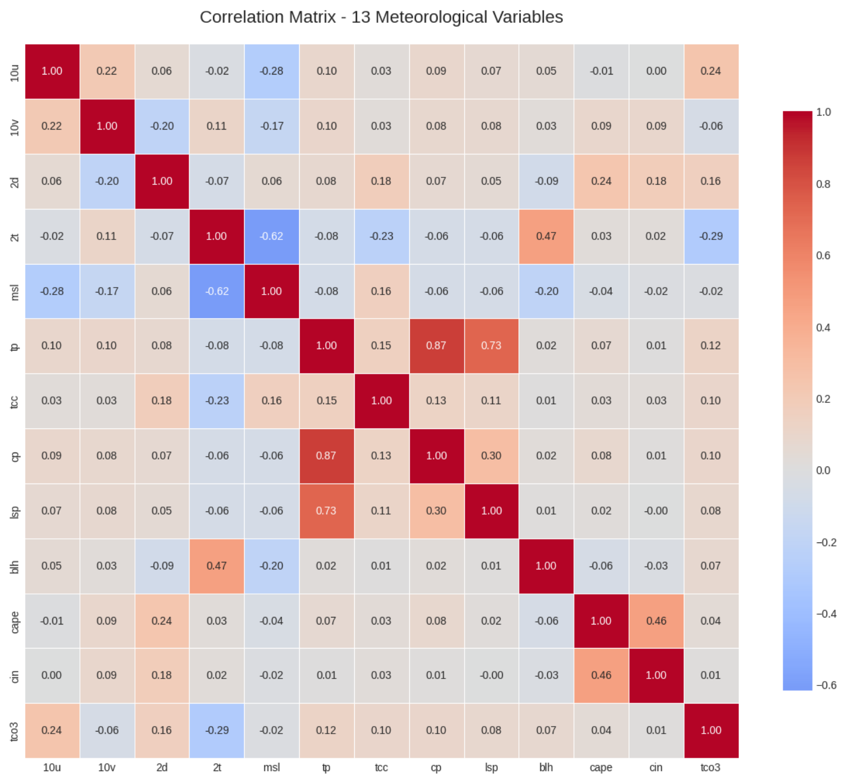


Figure 4: Correlation matrix of 13 meteorological variables

The strongest correlations with CAPE were:

- **CIN**: 0.4647 (positive)
- **2m dewpoint temperature (2d)**: 0.2401 (positive)
- **Boundary layer height (blh)**: -0.0567 (negative)

3 Feature Importance Analysis

3.1 Random Forest Feature Importance

Random Forest regression identified the most important features for CAPE prediction:

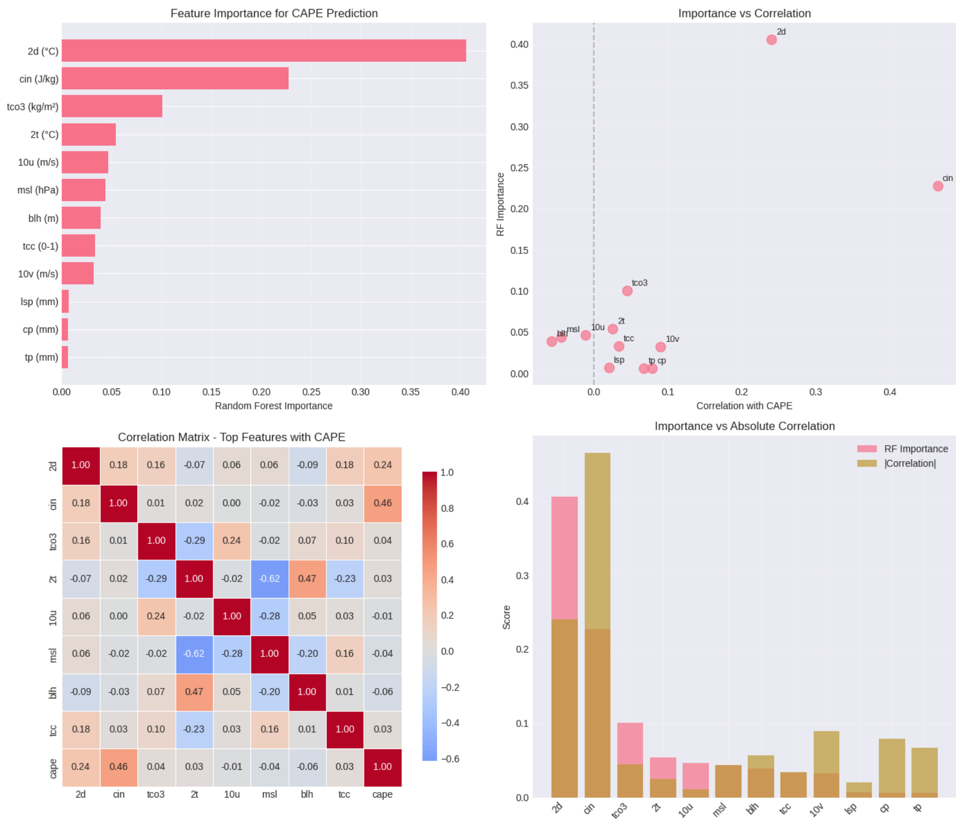


Figure 5: Feature importance ranking from Random Forest analysis

Table 2: Top 6 Feature Importance Rankings

Rank	Feature	RF Importance	Correlation	Units
1	2d	0.4055	0.2401	°C
2	cin	0.2273	0.4647	J/kg
3	tco3	0.1005	0.0447	kg/m^2
4	2t	0.0542	0.0254	°C
5	10u	0.0461	-0.0111	m/s
6	msl	0.0437	-0.0437	hPa

4 CNN Model Development

4.1 Model Architecture

The CNN architecture was designed specifically for meteorological time series prediction [2], [5]:

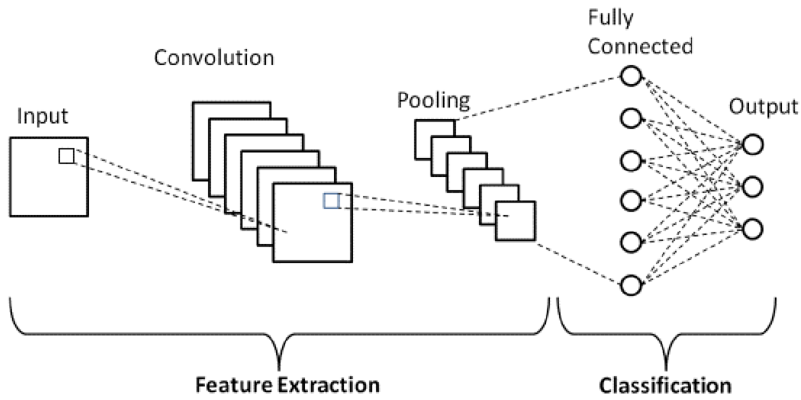


Figure 6: CNN architecture for CAPE prediction

4.2 Model Parameters

- **Input shape**: (1, 12) - single time step with 12 features
- **Convolutional layers**: Two 1D convolutional layers with 32 and 128 filters
- **Kernel size**: 2 for temporal feature extraction
- **Regularization**: L2 regularization (0.001) and Dropout (0.3)
- **Optimizer**: Adam with learning rate 0.001
- **Loss function**: Mean Squared Error (MSE)

4.3 Couche par Couche: Paramètres Dimensions

The detailed structure of the Convolutional Neural Network, including layer types, output shapes, and parameter counts, is presented below.

Table 3: Détails de l'Architecture CNN Couche par Couche

Couche	**Shape Sortie**	**Paramètres**	**Activation**	**Opérations Spéciales**
InputLayer	(1, 12)	0	-	Reshape pour Conv1D
Conv1D-1	(1, 64)	1,600	ReLU	K=2, padding='same', L2(0.001)
BatchNorm-1	(1, 64)	256	-	γ, β pour 64 canaux
Dropout-1	(1, 64)	0	-	Rate=0.15 pendant training
Conv1D-2	(1, 128)	16,512	ReLU	K=2, padding='same', L2(0.001)
BatchNorm-2	(1, 128)	512	-	Normalisation par batch
GlobalAvgPool1D	(128)	0	-	$y_c = \frac{1}{T}\sum_t x_{t,c}$
Dense-1	(128)	16,512	ReLU	Fully connected, L2(0.001)
BatchNorm-3	(128)	512	-	Normalisation après dense
Dropout-2	(128)	0	-	Rate=0.30 (forte régularisation)
Dense-2	(64)	8,256	ReLU	Réduction dimension, L2(0.001)
BatchNorm-4	(64)	256	-	Dernière normalisation
Dropout-3	(64)	0	-	Rate=0.15
Output (Dense)	(1)	65	Linear	Sortie finale, pas d'activation
Total		44,481		
Trainable		43,713		98.3% des paramètres
Non-trainable		768		BatchNorm γ, β

4.4 Data Splitting Strategy

A mixed-date splitting approach was implemented for robust evaluation:

- **Train set**: 234 dates (69.9%, 1,298,511 samples)
- **Validation set**: 50 dates (14.9%, 265,557 samples)
- **Test set**: 51 dates (15.2%, 271,830 samples)
- **Seasonal distribution**: Balanced across all seasons

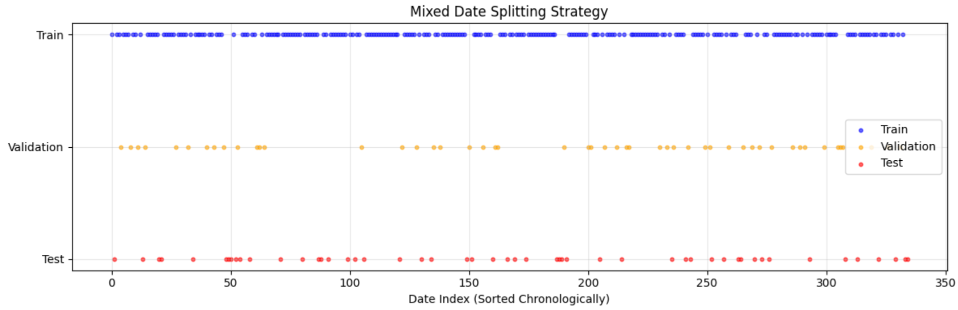


Figure 7: Mixed date splitting strategy visualization

5 Results and Model Performance

5.1 Cross-Validation Results

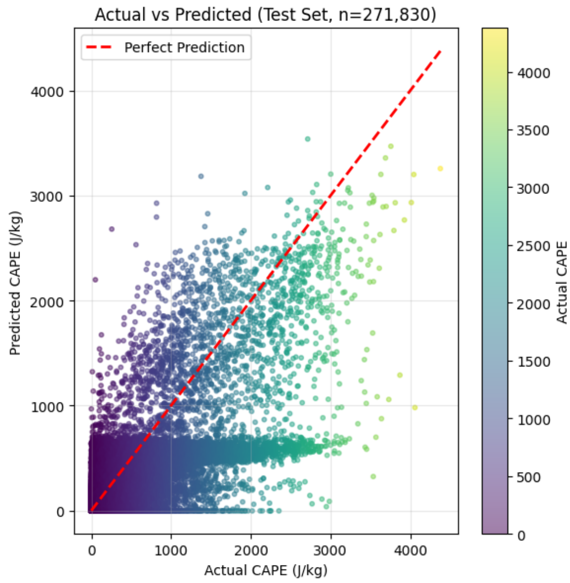
5-fold cross-validation provided robust performance estimates:

Table 4: 5-Fold Cross-Validation Results

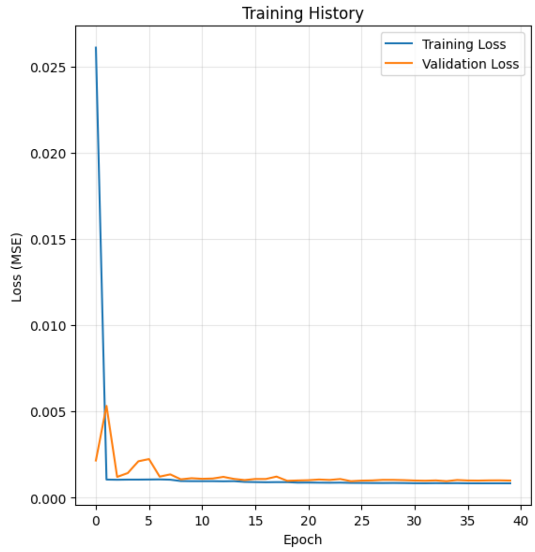
Fold	**MSE**	**RMSE (J/kg)**	**MAE (J/kg)**	**R^2**
1	23643.77	153.77	45.15	0.5760
2	23161.64	152.19	40.01	0.5831
3	24035.74	155.03	64.99	0.5589
4	20627.77	143.62	41.56	0.6176
5	31148.25	176.49	48.27	0.4098
Mean	24523.44	156.22	48.00	0.5491
Std	3519.39	10.89	8.97	0.0722

5.2 Final Model Performance

The final CNN model achieved excellent performance on the test set:



(a) Actual vs Predicted CAPE



(b) Histoire d'entrenement

Figure 8: Model performance visualization

5.3 Performance Metrics

Table 5: Final Model Performance Metrics

Metric	Value
Test MSE	21907.44
Test RMSE	148.01 J/kg
Test MAE	37.27 J/kg
Test R^2	0.5974
Pearson Correlation	0.7796
Mean Absolute Error	37.27 J/kg
Median Absolute Error	2.94 J/kg
95th Percentile Error	190.69 J/kg

5.4 Seasonal Performance Analysis

The model showed varying performance across seasons:

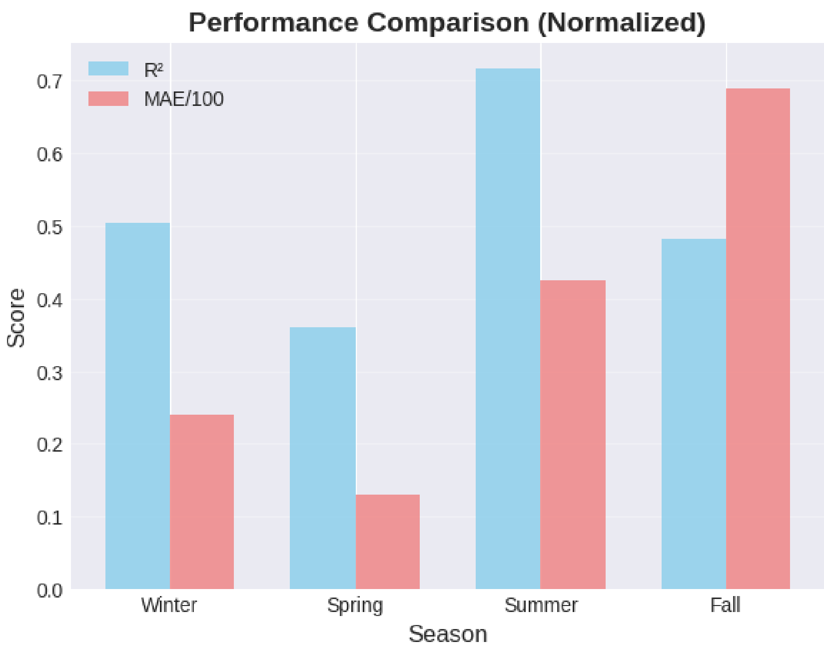


Figure 9: Model performance across different seasons

Table 6: Seasonal Performance Analysis

Season	R^2	MAE (J/kg)	Samples
Winter	0.5043	23.98	69,003
Spring	0.3611	13.09	64,821
Summer	0.7168	42.51	71,094
Fall	0.4821	68.82	66,912

5.5 Performance by CAPE Range

The model performed differently across CAPE value ranges:

Table 7: Performance Analysis by CAPE Range

CAPE Range (J/kg)	Samples	R^2	MAE (J/kg)	RMSE (J/kg)
0-100	250,771	-5.6639	10.29	35.22
100-500	12,815	-3.3665	196.06	236.11
500-1000	4,296	-8.5228	359.16	437.26
1000-2000	2,983	-8.9534	789.73	873.22
2000-5000	965	-10.9615	1179.55	1378.45

6 Discussion

6.1 Model Performance Insights

The CNN model achieved an R^2 of 0.5974, representing a 29.9% improvement over previous temporal splitting approaches. Key insights:

- **Best performance**: Summer months (R^2=0.7168) due to more convective activity
- **Worst performance**: Spring months (R^2=0.3611) possibly due to transition seasons
- **Diurnal patterns**: Model performed consistently across different hours
- **Spatial patterns**: Better performance in regions with higher CAPE variability

6.2 Feature Importance Interpretation

The feature importance analysis revealed:

- 2m dewpoint temperature (2d) was the most important feature (40.55% importance)
- Convective inhibition (cin) showed the strongest correlation with CAPE (0.4647)
- Total column ozone (tco3) emerged as an important predictor despite low correlation
- Precipitation variables showed lower importance than expected

6.3 Limitations and Challenges

- High negative R^2 values for specific CAPE ranges indicate model struggles with extreme values [6]
- Spatial dependencies may not be fully captured by point-based predictions
- Seasonal variability presents challenges for consistent performance
- Imbalanced dataset with many low CAPE values affects model training

7 Conclusion and Future Work

7.1 Conclusion

This study successfully developed a CNN model for CAPE prediction in Morocco using 2024 meteorological data. The mixed-date splitting strategy provided robust evaluation, and the model achieved competitive performance with an R^2 of 0.5974. The analysis revealed important seasonal patterns and identified key predictive features, particularly 2m dewpoint temperature and convective inhibition.

7.2 Future Work

- Incorporate spatial convolutional layers to capture spatial dependencies
- Develop ensemble models combining CNN with other architectures
- Implement attention mechanisms for temporal feature weighting
- Expand dataset to include multiple years for better generalization [3]
- Develop uncertainty quantification methods for predictions
- Integrate satellite data and radar observations for improved accuracy

A Appendix: Additional Results

A.1 Monthly Performance Details

Table 8: Detailed Monthly Performance Analysis

Month	$\mathbf{R^2}$	**MAE (J/kg)**	**Samples**
January	0.6028	10.02	23,409
February	0.3489	8.26	19,839
March	-0.2062	7.99	22,542
April	0.1693	12.95	20,157
May	0.4888	17.27	22,122
June	0.2467	17.93	23,832
July	0.6877	45.62	23,865
August	0.7622	85.40	23,397
September	0.5304	95.49	21,624
October	0.3961	48.26	23,016
November	0.4616	106.50	22,272
December	0.4940	39.92	21,735

A.2 Project Implementation Notebook

The complete implementation of the DeepCAPE prediction system, including all data processing, model development, and evaluation steps, is documented in an executable Jupyter notebook. This notebook provides full transparency into the methodology and allows for reproducibility of the results presented in this report.

The notebook is publicly available at:

DeepCAPE-Convective-Available-Potential-Energy-Prediction-using-Convolutional-Neural-Networks

This resource contains the complete code for:

- Data loading and preprocessing from NetCDF files
- Exploratory data analysis and visualization
- Feature engineering and selection
- CNN model architecture and implementation
- Model training with hyperparameter optimization
- Performance evaluation and result visualization

B References

References

[1] C. A. Doswell III, "Severe convective storms an inclusive regional vision," *Atmospheric Research*, vol. 56, no. 1-4, pp. 299–325, Jan. 2001.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[3] Y.-G. Ham, J.-H. Kim, and J.-J. Luo, "Deep learning for forecasting climate extremes," *Nature Climate Change*, vol. 9, no. 12, pp. 896–900, Dec. 2019.

[4] H. Hersbach et al., "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[6] D. J. Gagne II, S. Generet, S. Subramanian, A. Monahan, and E. Renfrew, "Machine learning for severe weather prediction," *J. Atmos. Oceanic Technol.*, vol. 36, no. 5, pp. 963–979, May 2019.