Deep Learning for Drought and Crop Yield Forecasting in Morocco: A Hybrid CNN-LSTM Approach

Abdelkarim Achaq*, Youness Lakhrissi* *Faculty of Sciences and Technologies of Fez, Sidi Mohamed Ben Abdellah University, Morocco
Email: abdelkarim.achaq@usmba.ac.ma

Abstract—Morocco faces increasing climate vulnerability, characterized by recurrent droughts and volatile agricultural yields. Traditional forecasting methods often fail to capture the complex, non-linear spatiotemporal dependencies of semi-arid climate systems. This study proposes a deep learning framework integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to forecast agricultural drought indices (SPI/SPEI) and cereal yields. Leveraging multi-source remote sensing data (CHIRPS precipitation, ERA5 temperature, MODIS/Sentinel-2 vegetation indices) from Google Earth Engine, we develop a low-latency forecasting system tailored to Moroccan agro-climatic zones (e.g., Saïss, Al Haouz). Our results demonstrate that the hybrid CNN-LSTM architecture outperforms traditional statistical baselines (ARIMA) and shallow machine learning models (Random Forest) in predicting seasonal drought stress. Furthermore, we analyze the model's transferability across different climatic regions, contributing to the development of resilient AI-driven early warning systems for North African agriculture.

Index Terms—Deep Learning, Drought Forecasting, Crop Yield, Morocco, CNN-LSTM, Remote Sensing, Climate Change Adaptation.

I. INTRODUCTION

A. Background

Agriculture remains the backbone of the Moroccan economy, contributing approximately 14% to the GDP and employing nearly 40% of the workforce. However, this sector is highly sensitive to climate variability. The "Plan Maroc Vert" and the subsequent "Generation Green 2020-2030" strategies emphasize the need for adaptation, yet decision-makers often lack precise, localized, and timely forecasts of seasonal conditions.

B. The Challenge

Meteorological drought in Morocco is driven by the complex North Atlantic Oscillation (NAO) and local topographic effects. Traditional physical models require extensive parameterization and ground data (e.g., soil properties, dense station networks) which are often scarce or outdated in the MENA region. This "data scarcity paradox" necessitates alternative modeling approaches.

C. AI in Climate Adaptation

Recent advances in Artificial Intelligence, particularly Deep Learning (DL), offer a promising solution. By ingesting vast amounts of earth observation data, DL models can learn latent representations of the climate system without explicit physical equations. While architectures like LSTMs have been successful in hydrological modeling globally (e.g., Namibia, USA), their application to the specific mixed-farming systems of Morocco remains underexplored.

D. Objectives

This paper aims to:

1) Develop a hybrid CNN-LSTM model that ingests time-series of satellite-derived indices (Precipitation, Temperature, NDVI) to forecast drought conditions 1-3 months ahead.
2) Validate the model against historical yield data for cereal crops in key Moroccan basins.
3) Interpret the model decisions to understand primary drivers of drought in different zones.

II. STUDY AREA AND DATA

A. Study Area

The study focuses on major agro-ecological zones in Morocco:

- **Fes-Meknes (Saïss Basin)**: A rainfed cereal and olive production hub.
- **Marrakech-Safi (Al Haouz)**: A semi-arid region heavily dependent on irrigation.
- **Casablanca-Settat (Chaouia)**: The historical "granary" of Morocco.

B. Data Sources

We utilize the Google Earth Engine (GEE) platform for data acquisition:

TABLE I
REMOTE SENSING DATASETS

Variable	Source	Resolution	Period
Precipitation	CHIRPS v2.0	0.05° (Daily)	2000-2024
Temperature	ERA5-Land	0.1° (Hourly)	2000-2024
Vegetation	MODIS (MOD13Q1)	250m (16-day)	2000-2024
	Sentinel-2	10m (5-day)	2015-2024
Soil Moisture	NASA-USDA SMAP	10km (3-day)	2015-2024

C. Ground Truth

Ground truth data consists of annual cereal yield statistics (Quintals/Hectare) obtained from the Ministry of Agriculture, along with calculated Standardized Precipitation Indices (SPI) derived from station data where available.

III. METHODOLOGY

A. Data Preprocessing

1) Temporal Alignment: All datasets are aggregated to a common temporal resolution (e.g., monthly or dekadal steps). Missing values are handled via linear interpolation.

2) Normalization: Features are normalized using Min-Max scaling to the range $[0, 1]$ to facilitate neural network convergence:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

B. Model Architecture: CNN-LSTM

We propose a hybrid architecture designed to capture both local feature patterns and long-term temporal dependencies.

Fig. 1. Proposed CNN-LSTM Architecture (Placeholder)

1) 1D-CNN Component: A 1D Convolutional layer slides over the input time window (e.g., 12 months) to extract high-level features such as sudden rainfall drops or heatwaves.

$$h_{cnn} = \sigma(W_c * x + b_c) \tag{2}$$

2) LSTM Component: The feature maps from the CNN are fed into a Long Short-Term Memory (LSTM) network, which maintains a memory state C_t to learn seasonal cycles and trends.

$$h_t = \text{LSTM}(h_{cnn}, h_{t-1}) \tag{3}$$

3) Output Layer: A fully connected dense layer maps the final LSTM hidden state to the target variable (Yield or SPI).

IV. EXPERIMENTAL DESIGN

A. Training Strategy

- **Split**: Training (2000-2018), Validation (2019-2021), Testing (2022-2024).
- **Loss Function**: Mean Squared Error (MSE) for regression.
- **Optimizer**: Adam with learning rate scheduling.

B. Baselines

To benchmark performance, we compare against:

- **ARIMA**: Auto-Regressive Integrated Moving Average.
- **SVR**: Support Vector Regression.
- **Random Forest**: Ensemble learning on tabular features.

V. PRELIMINARY RESULTS (EXPECTED)

[This section will be populated with experiments] We anticipate the CNN-LSTM model to achieve an RMSE reduction of 15-20% compared to baseline statistical models, particularly in capturing the onset of the 2022 drought.

VI. DISCUSSION

A. Model Interpretability

Using SHAP values, we expect to show that accumulated precipitation in the growing season (Nov-Feb) is the dominant predictor, but that temperature anomalies (heat stress) during the grain-filling stage (March-April) are critical for yield accuracy.

B. Implications for Policy

This system can support the "Generation Green" digital strategy by providing proactive alerts to regional agricultural offices (ORMVA), enabling better water allocation planning.

VII. CONCLUSION

This study demonstrates the potential of Deep Learning to bridge the data gap in Moroccan climate services. By integrating global satellite data with advanced sequence modeling, we provide a robust tool for drought and yield forecasting.