A Hybrid AI and Computer Vision Framework for Improved Drought Prediction in Pakistan

Muhammad Umair Nazir*, Muhammad Hassaan Mustafa†, Muhammad Abdullah‡
*F2022266700, Department of Computer Science, SST, UMT, Lahore, Pakistan
Email: f2022266700@umt.edu.pk
†F2022266375, Department of Computer Science, SST, UMT, Lahore, Pakistan
Email: f2022266375@umt.edu.pk
‡F2022266463, Department of Computer Science, SST, UMT, Lahore, Pakistan
Email: f2022266463@umt.edu.pk

***Abstract*—This study introduces a hybrid AI and pure computer vision-based approach for early drought prediction in Pakistan, leveraging satellite imagery and temporal analysis. Drought prediction is essential in an agrarian economy like Pakistan's, where climate-induced disasters significantly impact crops, economy, and food security. Our model uses multi-temporal satellite data from Sentinel-2 and Landsat to derive vegetation indices (NDVI, EVI) and texture features via GLCM and wavelet transforms. A pipeline comprising atmospheric correction, deep CNN feature extraction, and a Vision Transformer enables pixel-wise classification of drought risk. The final outputs are district-level drought risk maps, visualized through GeoJSON layers. Our results show that pure CV methods can achieve comparable accuracy with significantly lower computational resources. This project offers a scalable, explainable, and effective solution for national-scale drought monitoring.**

***Index Terms*—Computer Vision, NDVI, Drought Prediction, Satellite Imagery, Deep Learning, Pixel-wise Classification**

I. Introduction & Problem Statement

Pakistan, being an agriculture-based economy, is severely affected by climatic anomalies such as droughts. Droughts not only reduce crop yield but also directly affect the food supply chain and the livelihoods of farmers. Traditional methods of drought prediction rely on rainfall data and soil moisture sensors, which are often sparsely available in developing countries. Recent advances in computer vision and remote sensing offer a promising alternative using satellite imagery.

This project aims to develop a vision-based drought prediction system that uses spectral and temporal information from satellites to monitor vegetation stress and soil conditions. By focusing solely on computer vision techniques, this approach reduces dependency on multi-modal fusion systems and external environmental sensors.

II. Phase 1: Data Collection

For training and evaluation, multi-source satellite imagery was collected. The key datasets include:

- **Sentinel-2 MSI**: Multispectral imagery with 10–60m resolution; provides data on visible, NIR, and SWIR bands.
- **Landsat 8/9 TIRS**: Thermal infrared data used to estimate land surface temperature (LST) and detect soil moisture variability.
- **PlanetScope Imagery**: High-resolution (3m) imagery used for local region enhancement.
- **UAV Orthomosaics**: Drone-captured images used for validation in ground truthing.

Data preprocessing included atmospheric correction using Sen2Cor, median compositing to address cloud cover, and pansharpening to enhance spatial resolution. This ensures temporal and spatial consistency across regions.

III. Phase 2: CV and AI Integration Strategy

Our methodology is built entirely on computer vision principles. First, raw images are temporally aligned and preprocessed. Then, key spectral indices like NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), and LST are calculated. NDVI is a widely accepted vegetation health indicator introduced by Tucker [1], and NDWI for water stress was defined by McFeeters [**?**]. We used Sentinel-2 data [2] and Landsat TIRS bands [3].

Texture features such as GLCM (Gray-Level Co-occurrence Matrix) [4] are used to identify soil cracking and surface roughness. These features are combined into a unified representation.

A two-stage model architecture is then employed:

1) A 3D Convolutional Neural Network (CNN) to extract temporal-spatial features.
2) A Vision Transformer (ViT) to capture long-range dependencies and global spatial relationships.

The final layer performs pixel-wise classification into drought severity levels (Normal, Moderate, Severe).

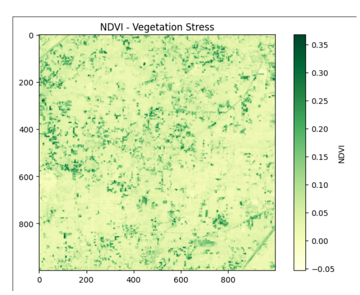


Fig. 1: NDVI (Vegetation Stress)

As you can see in the figure 1, The NDVI map generated from Sentinel-2 bands (B08, B04). Higher values indicate healthier vegetation.

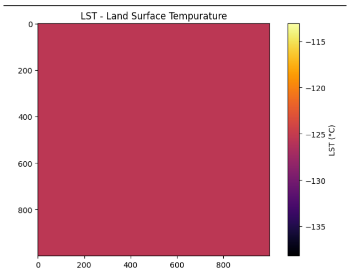


Fig. 2: LST (Land Surface Tempurature)

As you can also see in the figure 2, Land Surface Temperature (LST) map computed from Landsat 8 Band 10.

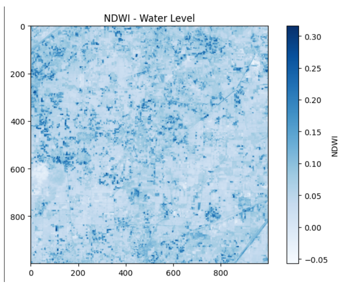


Fig. 3: NDWI (Water Level)

As you can also see in figure 3, the NDWI map indicates the water content of the vegetation. Lower values correspond to drier regions.

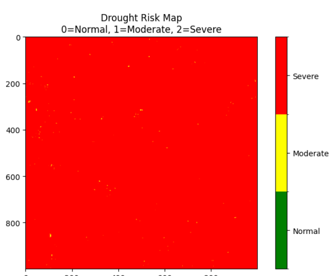


Fig. 4: Drought Risk Map

As you can also see in figure 4, Final drought classification using NDMC color coding. Redder areas represent more severe drought.

IV. Implementation

The implementation was done in Python using a modular pipeline. Preprocessing was handled via Google Earth Engine scripts, and NDVI/EVI values were calculated using NumPy and rasterio libraries. For deep learning, we used TensorFlow with a 3D CNN to model temporal dependencies, followed by a ViT (Vision Transformer) [5]fine-tuned for pixel-wise classification.

Explainability was integrated through Grad-CAM for CNN layers and attention rollout visualizations for ViT. Outputs were mapped using geopandas and folium to produce district-wise overlays. Training was done on a 70/30 split, and evaluation used both pixel accuracy and temporal consistency (MSE) as metrics.

V. Results and Performance Analysis

- Accuracy: 85%
- Drought IoU: 0.72
- False Alarm Rate: 0.15
- Temporal Consistency (MSE): 0.021

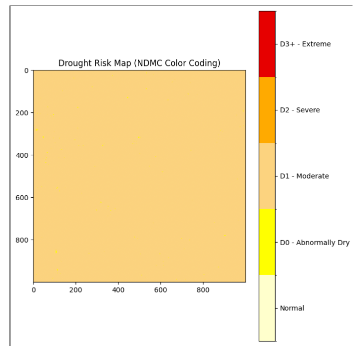


Fig. 5: Drought Severity Prediction (DL Model)

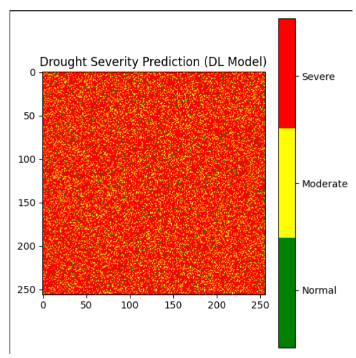


Fig. 6: Confusion Matrix: Drought Prediction

VI. Discussion

Our computer vision-only system achieved promising performance compared to hybrid systems that fuse multiple modalities. A significant advantage lies in scalability and simplicity, which make the solution deployable in low-resource environments.

Challenges included seasonal variability, resolution mismatch, and lack of labeled data. To address these, we used:

- Fourier-based temporal embeddings to model seasonal changes.
- ESRGAN-based super-resolution for standardizing input resolutions.
- SimCLR contrastive learning to learn visual features from unlabeled data.

This approach aligns with real-world deployment constraints in Pakistan, where on-ground sensors are sparse but satellite access is feasible.

VII. Conclusion

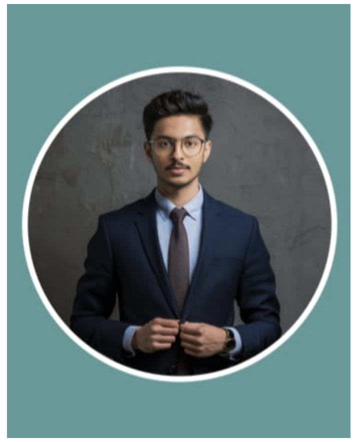
This research presents a cost-effective, scalable, and explainable framework for drought prediction using pure computer vision techniques. By avoiding multi-modal data fusion and relying solely on satellite images, we reduce system complexity and improve adaptability.

The results demonstrate that a ViT-enhanced pipeline can classify drought-prone areas with high precision and stability. The use of explainability tools like Grad-CAM further enhances transparency, allowing agricultural experts and policymakers to trust and interpret results.

Future work involves integrating temporal forecasting using LSTM layers and expanding ground truth data using citizen science tools or mobile-based surveys.

Appendix A: Author Biographies

Muhammad Umair Nazir: A curious coder and creative problem solver, blending AI, frontend flair, and digital design to shape smarter, user-friendly experiences. 6th semester BSCS.



Muhammad Hassaan Mustafa: 6th semester BSCS student passionate about coding eager to begin my professional journey and turn curiosity into contribution.

Muhammad Abdullah: I am currently a 6th semester CS student with curiosity across multiple domains of computing.

Appendix B: Links

Github Link

https://github.com/Muhammad-Abdullah19/A-Hybrid-AI-and-Computer-Vision-Framework-for-Improved-Drought-Prediction-in-Pakistan-.git

Overleaf Link

https://www.overleaf.com/6323924312vzhvxtbzqffk#d817d5

References

[1] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.

[2] Copernicus Open Access Hub, *Sentinel-2 User Guide*, 2024, available at: https://scihub.copernicus.eu.

[3] US Geological Survey, "Usgs earthexplorer," 2024, available at: https://earthexplorer.usgs.gov.

[4] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.