AgriVoIT: A Multi-Modal Temporal Vision Transformer for Climate-Informed Commodity Price Forecasting

Sharanya Roy, Krisha Agarwal, Sahir Gupta, Anshul Patil,
Ahan M R, Sunishchal Dev, Vasu Sharma, Kevin Zhu

NEURAL INFORMATION PROCESSING SYSTEMS

ALGOVERSE

Motivation

- Climate extremes increasingly disrupt agricultural production, creating **volatility in staple commodity markets** and threatening food security.

- At 2°C warming, **10–31%** of current crop production moves outside safe climate zones.

- **733 million** people face food insecurity worldwide (FAO, 2024).

- Food price inflation and volatility are major causes of hunger.

Research Objective

We aim to forecast **state-level marketing-year prices** for **corn, wheat, and soybeans (2015–2023)** by modeling temporal and multi-modal dependencies between:

- Climate reanalysis data
- Satellite imagery (Sentinel-2)
- USDA production statistics
- Historical market prices

AgriVolt Framework

- Integrates **climate reanalysis**, **satellite imagery (Sentinel-2)**, **production data (USDA)**, and **historical market prices**.
- Employs **cross-modal attention** and **temporal encodings** to capture links between weather events and market dynamics.
- Features a **price-focused prediction head**, directly modeling economic outcomes rather than just crop yields.

Datasets

Dataset Type	Source	Examples of Variables
Climate Data	NOAA HRRR	Temperature, precipitation, radiation, wind, humidity
Remote Sensing	Sentinel-2	Vegetation indices, droughts, phenology
Production Data	USDA NASS	County-level yield, harvested acres
Market Data	USDA	Daily prices (corn, wheat, soybean)

Each data source is aligned by date and county, aggregated to **(state, marketing-year)** samples.

Model Architecture

Tokenization: One token per Sentinel-2 tile per date; one per weather record; one USDA per county.

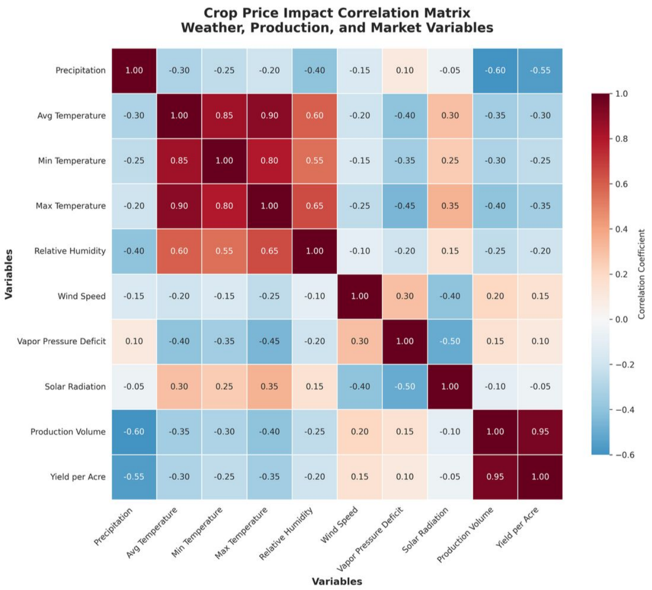
Image Encoder: 4-stage Pyramid Vision Transformer (PVT).

Per-Date Fusion: Cross-attention aligns imagery with concurrent weather & USDA data.

Spatial Transformer: Aggregates tile-level signals per county.

Temporal Transformer: Captures seasonal and lag effects.

Output: MLP head generates state-level price weighted by harvested area.



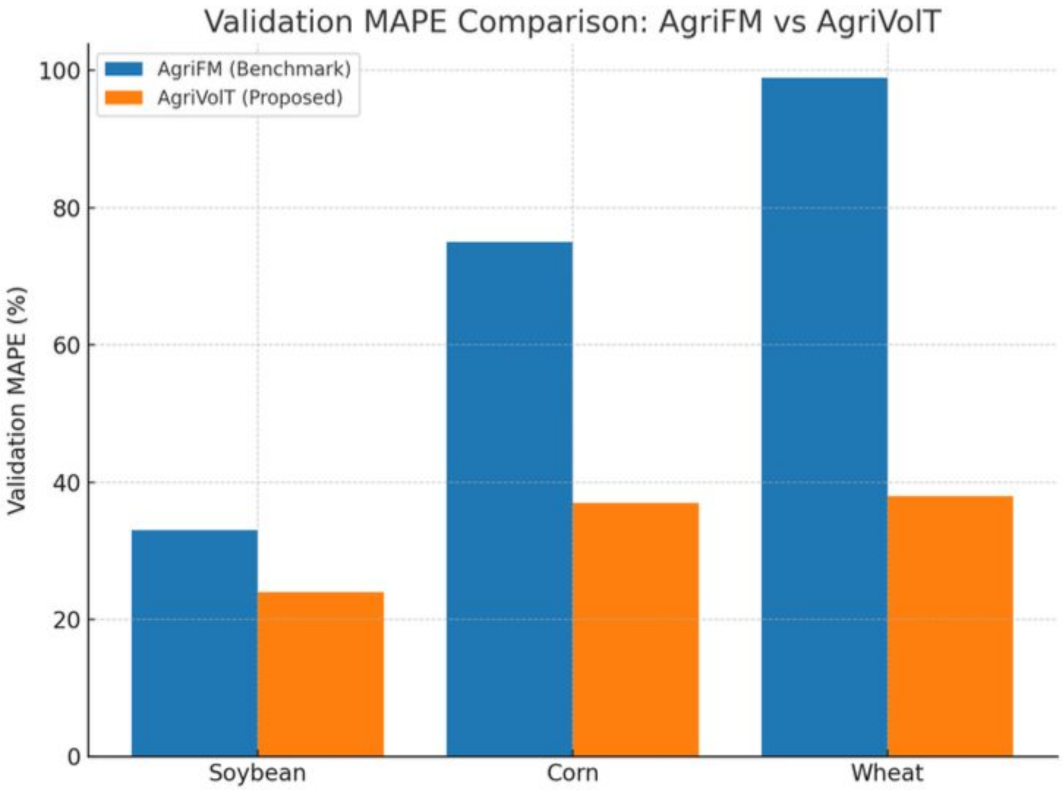
Experimental Setup

- **Crops:** Corn, Wheat, Soybean (U.S., 2015–2023).
- **Evaluation Metrics:** MAE, RMSE, MAPE, SMAPE.
- **Training:** AdamW optimizer, cosine schedule, early stopping.
- **Baselines:**
  - ARIMA, Prophet (classical)
  - Naïve (previous-year price)
  - XGBoost (nonlinear ML)
  - AgriFM (multi-modal baseline)

Quantitative Results

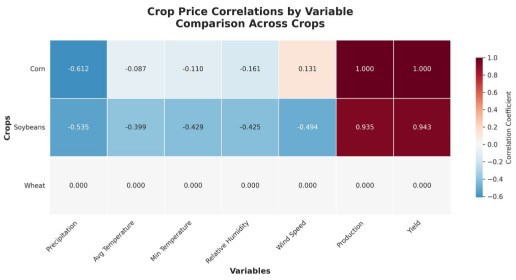
Model	Crop	MAE (USD/bu)	RMSE (USD/bu)	MAPE	SMAPE
ARIMA	Soybean	4.53	5.40	43.56%	35.0%
XGBoost	Soybean	4.10	4.74	32.5%	27.0%
AgriVoIT	Soybean	**3.44**	**4.65**	**24.39%**	**19.62%**
ARIMA	Corn	1.92	2.19	44.2%	36.0%
AgriVoIT	Corn	**2.36**	**2.90**	**37.12%**	**27.06%**
ARIMA	Wheat	2.47	2.89	44.36%	36.0%
AgriVoIT	Wheat	**2.85**	**3.59**	**38.21%**	**27.62%**

Result Comparison



Insights and Interpretability

- Cross-modal attention maps reveal strong coupling between **precipitation, vegetation indices, and yield**.
- Temporal encodings enable the model to capture **lagged effects** (e.g., drought → yield drop → price rise).
- County-level spatial attention preserves local heterogeneity.
- Ablation studies show each modality (imagery, weather, USDA) improves predictive power.



Conclusion and Future Work

Conclusion:

- AgriVoIT provides an end-to-end framework for **climate-informed commodity price forecasting**.
- Outperforms traditional econometric, ML, and existing multi-modal models.

Limitations:

- Does not yet include real-time economic variables or causal inference.
- Focused on three major U.S. crops.

Future Directions:

- Incorporate macroeconomic indicators, policy and trade data.
- Extend to high-frequency and global datasets.
- Develop causal interpretability for market-environment interactions.