MERCON 2026 · MORATUWA ENGINEERING RESEARCH CONFERENCE

When Prices Double in a Week

Supply-Chain-Aware Forecasting of Agricultural Price Volatility in Import-Isolated Markets

A Unified vs. Seasonal Ensemble Analysis of Sri Lankan Vegetable Markets

Department of Computer Science & Engineering, University of Moratuwa

THE PROBLEM

A market with nowhere to hide

Sri Lanka virtually never imports commercial vegetables. There is no external buffer -every drought, flood, or fuel shock passes straight through to the dinner table.

That structural isolation makes Sri Lankan vegetable prices one of the most natively volatile food markets in South Asia -and, until now, one of the least modelled.

2× – 3×

price swing possible in a single week

0%

import buffer for commercial vegetables

2

monsoon-driven cultivation seasons every model before this one has ignored

THE STRUCTURAL BLIND SPOT

One island, two agricultural clocks

Every prior forecasting study treats the year as one continuous series. But Sri Lanka's vegetable supply runs on two monsoon-governed seasons -with different crops, different rainfall regimes, and different supply corridors.

M

MAHA SEASON

October – April

- Northeast monsoon
- Main cultivation season -highest volume
- Highland (upcountry production strong
- Greater volatility: lower seasonal R^2 (0.9210)

Y

YALA SEASON

May – September

- Southwest monsoon
- Highland production contracts
- Lowland (dry/wet zone) output expands
- Most predictable regime: highest R^2 (0.9420)

POSITIONING THE CONTRIBUTION

Three gaps. Five questions.

WHAT THE LITERATURE MISSED

1. No integrated dataset addressing 26%+ missingness or the urban-weather / rural-origin spatial mismatch
2. No seasonal decomposition treating Yala and Maha as structurally distinct series
3. No distinction between genuine missing records and structural market absences

RESEARCH QUESTIONS

RQ1 Are price shocks structurally predictable, or just a random walk?

RQ2 What governs the farmer-gate-to-retail spread across seasons?

RQ3 Do seasonal models beat a unified model -or does split blindness cancel the gain?

RQ4 Do macroeconomic shocks outweigh weather in predictive importance?

RQ5 Can a 2013-2019 model survive an unseen 2024 hyperinflation regime?

CONTRIBUTION #1

The first integrated Sri Lankan vegetable price dataset

12	14	7	61,152
vegetable varieties	market centres	years (2013–2019)	cleaned records

FOUR DATA SOURCES, FUSED

- HARTI retail & farmer-gate weekly prices
- Open-Meteo origin-zone rainfall & temperature
- Ceylon Petroleum diesel prices
- Central Bank USD/LKR exchange rate

CONTRIBUTION #1

Missingness, crushed

Targeted Data Scoping (Strategy 1)

- Focused on the 12 most consistent crop varieties and 14 primary market centers from 2013 - 2019.

- Reduced raw retail missingness from 26.3% down to 5.4% without heavy synthetic imputation.

Dual-Origin Imputation & Interpolation (Strategy 2)

- Supply-Chain Mapping: Reconstructed non-producing urban market prices using the mean of their top 2 supplying regional hubs (dropped missingness from 33.8% to 7.6%).

- Bounded Linear Interpolation: Filled remaining short temporal gaps (<4 consecutive weeks) to push final missingness down to 3.9%.



HOW WE GOT THERE

Four preprocessing strategies -zero fabricated data

1 Data Reduction

Scoped 26 varieties → 12 varieties
37 markets → 14 markets. 26.3% → 5.37% missing, before any imputation.

2 Dual-Origin Farmer Imputation

Reverse-engineered supply chains: each missing urban farmer-gate price imputed from its top-2 historical supplying markets.

3 Origin-Averaged Weather

Weather pulled from actual cultivation districts, not retail-market cities -lagged 1, 4, 8 weeks for phenological delay.

4 Macroeconomic + Calendar

Diesel price, USD/LKR rate, and public-holiday count integrated at 100% completeness.

ENCODING MARKET MECHANICS

Features that know how a supply chain behaves

Lag features	1, 2, 3, 4, and 8-week offsets for retail price, farmer-gate price, rainfall, and temperature variables.
Rolling stats	4 and 8-week rolling mean and standard deviation of farmer prices to capture shifting volatility regimes.
Momentum & spread	1-week lag to 4-week rolling mean ratio for price momentum, plus lagged farmer-to-retail spread to encode supply-chain markup dynamics.
Cyclical encoding	Sine and cosine transformations of week numbers (sin(2 * PI * w / 52), cos(2 * PI * w / 52)) to preserve 52-week circular continuity.
Diesel × Season	Interaction term capturing how identical fuel costs create differing transport cost impacts in Yala vs. Maha seasons.

CONTRIBUTION #2

A Two-Stage, Three-Way Modelling Framework

STAGE 1

Baseline

- **SARIMA** baseline (71.20% accuracy)
- Quantifies fundamental linear prediction boundaries
- Confirms necessity of capturing **non-linear market dynamics**

STAGE 2

Gradient Ensemble

- **XGBoost + LightGBM** heterogeneous blend
- Optuna hyperparameter optimization
  20 (XGBoost) + 20 (LightGBM) + 30 (weight blending) = 70 total

$$\hat{y} = w_{xgb} \cdot \hat{y}_{xgb} + w_{lgb} \cdot \hat{y}_{lgb}$$

Validation-tuned weights: **0.52** XGB / **0.48** LGB

Three Parallel Configurations (RQ3)

Unified	Yala-Only	Maha-Only
All 7 Years	May – Sep markup 38.4%	Oct – Apr markup 29.7%

RESULTS · RQ1

Isolated markets are not random walks

On the 2018–2019 held-out test set, no temporal leakage: Unified Model

0.9281

R^2 score

90.84%

accuracy (1 − MAPE)

9.16%

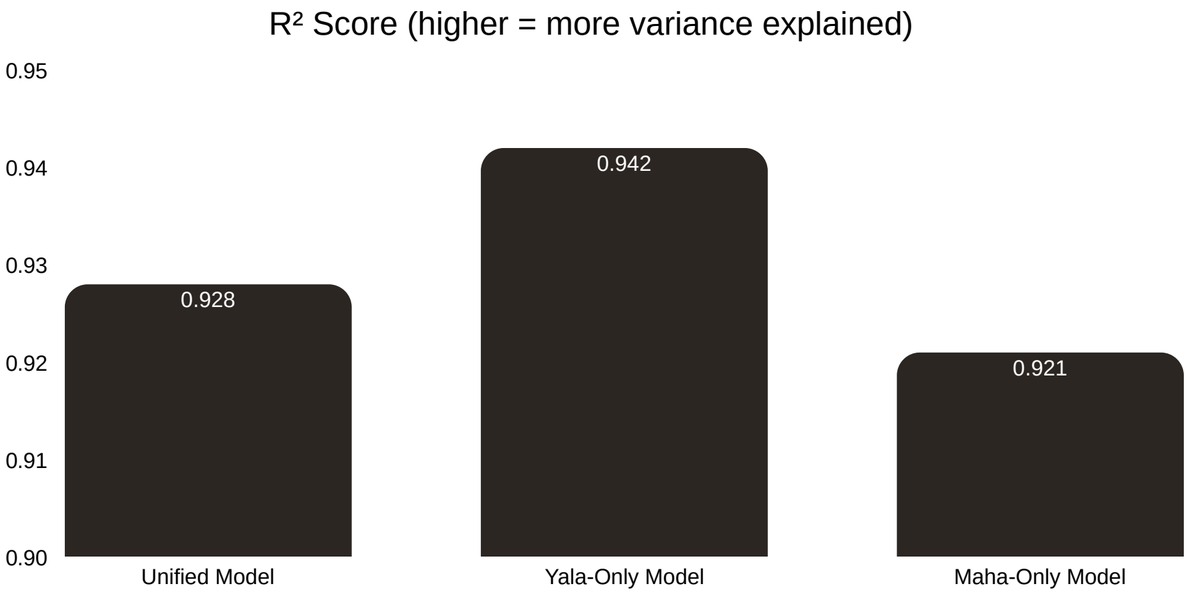
MAPE

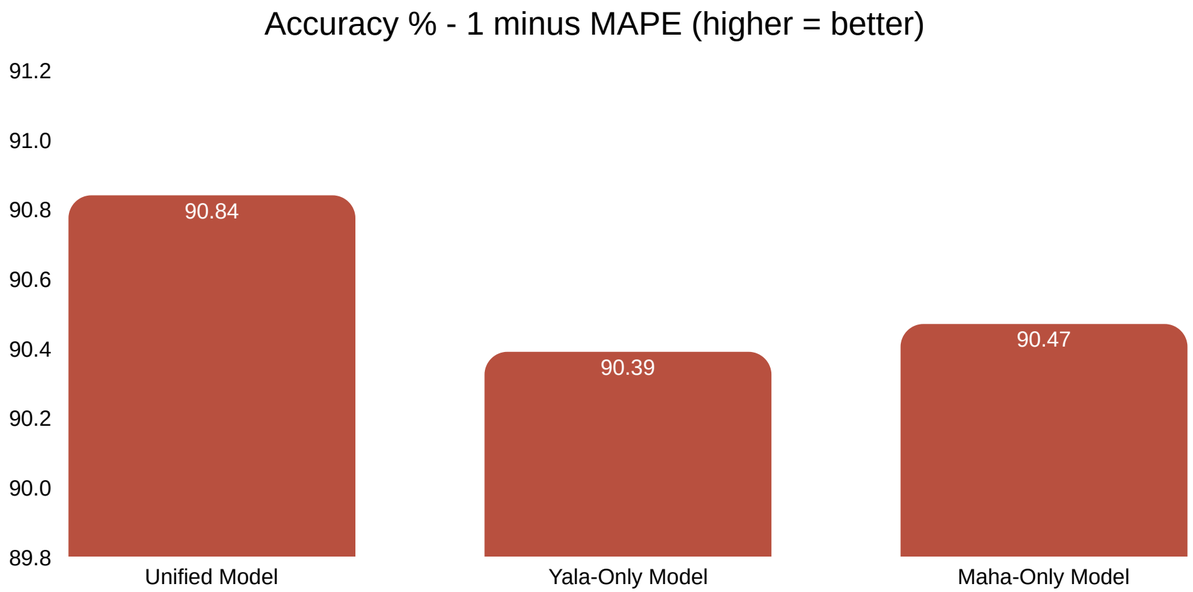
Predicted vs. Actual -tight diagonal clustering confirms low bias

Ensemble beats both base learners

XGBoost	**90.76%**
LightGBM	**90.72%**
Weighted Ensemble	**90.84%**

Precision vs. momentum: a trade-off no one had quantified





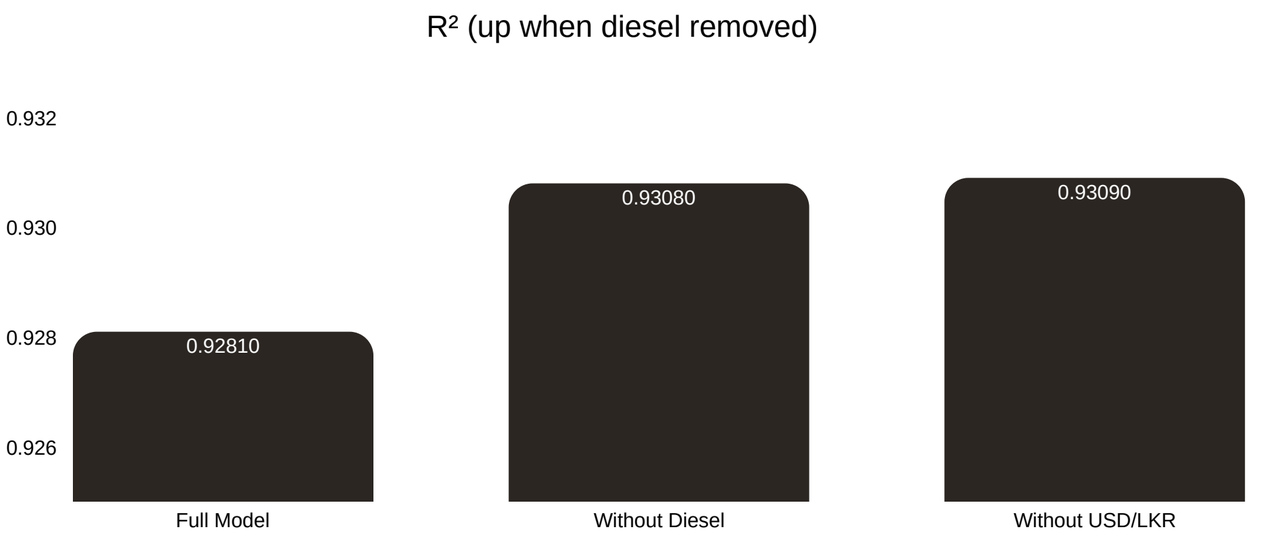
Yala-only wins on R^2 (+0.0139) -within-season training captures season-specific dynamics more precisely.

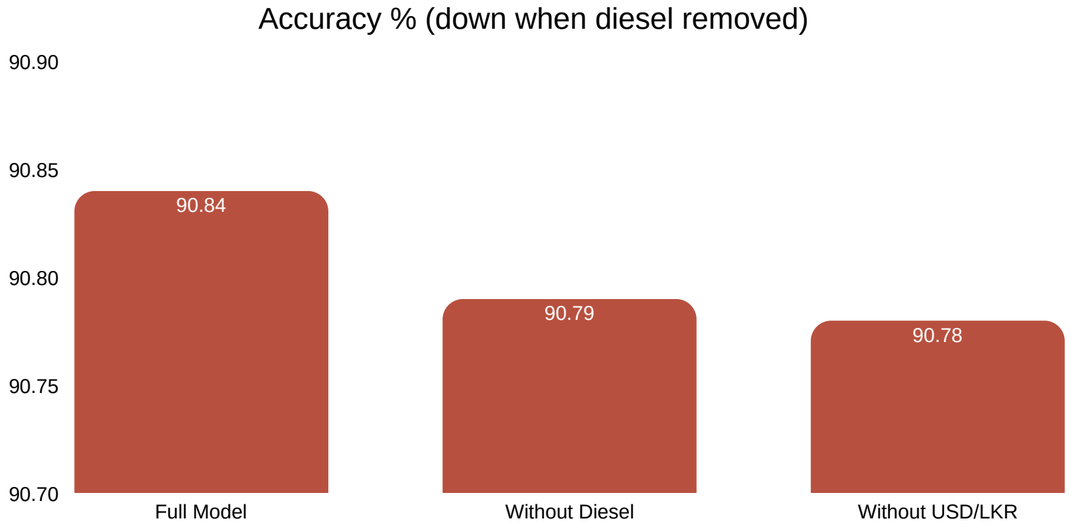
Unified wins on accuracy (90.84%) -cross-season price momentum compensates for structural blindness.

Neither approach is universally superior: Overlapping confidence intervals suggest differences are within statistical noise - both approaches have merit depending on operational need

The diesel price paradox

Removing diesel from the model raises R^2 but lowers accuracy -a contradiction that reveals what each metric actually measures.





The Paradox: Removing diesel improves R2 but reduces prediction accuracy.

Why R2 Misleads: Fuel prices spike discretely (not smooth), making them critical for extreme events but invisible to global R^2 metrics.

Why Accuracy Needs It: R^2 rewards smooth trends, masking features that spike discretely. Diesel's value is concentrated in rare, severe supply shocks - invisible to global correlation metrics but critical for early warning.

Takeaway: Diesel's value is concentrated in rare, severe price surges, making global correlation metrics understate its true impact.

RESULTS · RQ5 -THE STRESS TEST

Trained on 2013–2019. Tested on hyperinflation.

Zero retraining. Deployed against a completely unseen 2024 dataset (6,877 rows) spanning Sri Lanka's sovereign debt crisis and systemic hyperinflation.

90.84%

Baseline Accuracy (2013–2019 Held-Out Test Set)

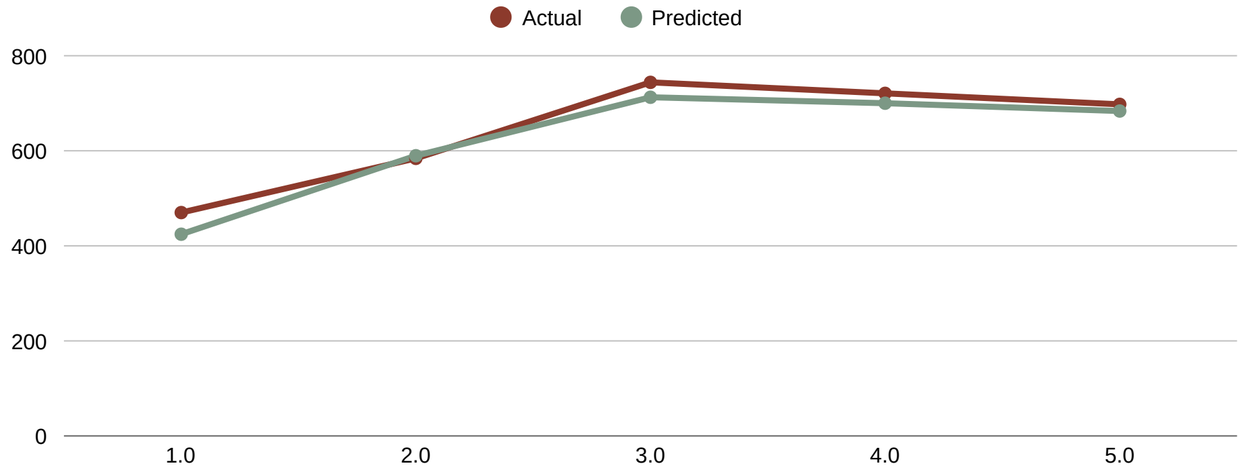
85.96%

Zero-Retraining Accuracy on Unseen 2024 Crisis

≈5 pp

Minimal Performance Drop Across a 5x Inflation Surge

MICRO STRESS-TEST -GREEN CHILLIES, KALUTHARA: TRACKING A 274 LKR SURGE



0.95%

Prediction Error at Peak Surge Inflection (Week 29)

WHY IT MATTERS

Four contributions that move the field

First integrated dataset

12 vegetables × 14 markets × 7 years, fused with weather, diesel & FX -retail missingness cut from 26.3% to 5.37%.

Origin-aware weather

Rainfall & temperature sourced from real cultivation zones, not retail cities -with vegetable-specific optimal lags.

Unified vs. seasonal trade-off, quantified

First empirical evidence that seasonal models win on precision while unified models win on momentum-driven accuracy.

Proven regime resilience

85.96% accuracy on an unseen 2024 hyperinflationary shock -with zero retraining.

INTELLECTUAL HONESTY

Where we push next

The tree extrapolation problem

Leaf nodes cannot predict values outside their training range -the source of the ~5pp accuracy drop in the 2024 regime. Relative supply-chain dynamics transfer; absolute price levels are harder.

THREE CONCRETE NEXT STEPS

1. Inflation-adjusted targets normalised by a diesel inde
2. Time-decay sample weighting to emphasise recent cost structures
3. Shock-detection composite features that dynamically rebalance ensemble weights during extreme events

“Neither approach is universally superior -the selection of unified versus seasonal model should be guided by whether transition-week accuracy or within-season precision is the operational requirement.”

-Conclusion, Section VII

This is a system that knows the edges of its own confidence -and tells you exactly when precision matters more than momentum.

References Slide 1

1. Champika, J. A., & Mugera, A. (2023). Analysis of price behavior in Sri Lankan vegetable market. J. Agribus. Market, 10(1), 4–29.

2. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37–45.

3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD Int. Conf. KDD, 785–794.

4. Ke, G., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NeurIPS, 30.

5. Madubhashini, E. L. N. D. (2023). Predicting vegetable prices in Sri Lanka using machine learning techniques. Master's thesis, Univ. of Colombo School of Computing.

6. Little, R. J. A., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. John Wiley & Sons.

References Slide 2

7. Komarek, A. M., De Pinto, A., & Smith, V. H. (2020). A review of types of risks in agriculture. Agricultural Systems, 178, 102738.

8. HARTI. (2024). Weekly average retail prices of all vegetable varieties. [harti.gov.lk] (https://www.harti.gov.lk/index.php/en/market-information/data-food-commodities-bulletin)

9. Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models. Journal of Economic Perspectives, 15(4), 157–168.

10. Purohit, S. K., et al. (2021). Time series forecasting of price of agricultural products using hybrid methods. Applied Artificial Intelligence, 35(15), 1388–1406.

11. Illankoon, I. M. G. L., & Kumara, B. T. G. S. (2020). Analyzing the influence of various factors for vegetable price using data mining. Proc. 13th Int. Res. Conf. KDU.

12. Akiba, T., et al. (2019). Optuna: A next-generation hyperparameter optimization framework. Proc. 25th ACM SIGKDD, 2623–2631.

Thank you

Structural governance in a volatile, import-isolated market -and a model that knows when to trust itself.

Authors: Ranuga Weerasekara, Heshan Nethmina, Manuja Ranathunga, Vinma Wettasinghe, Dinithi Navodya, Subavarshana Arumugam, Nirasha Munasinghe, Dr. Nisansa de Silva, Dr. Sandareka Wikramanayake

Department of Computer Science & Engineering, University of Moratuwa

Questions welcome.