

FOOD PRICE SPIKE DETECTION IN INDONESIA: A COMPARATIVE STUDY OF DEEP LEARNING ARCHITECTURES AND SPATIOTEMPORAL PRICE PROPAGATION ANALYSIS

Muhammad Arya Putra Handrian¹, Handoyo², Yudi Eko Windarto^{3*}, Wulan Oktabriyantina⁴

¹²³Department of Computer Engineering, Universitas Diponegoro, Indonesia

⁴Department of Economics and Development Studies, Universitas Diponegoro, Indonesia

*Email corresponding author: yudi@live.undip.ac.id

Abstract

Food price spikes pose a severe threat to household economic security and national price stability in developing nations, yet existing early warning mechanisms are predominantly reactive and fail to explicitly capture spatial shock propagation. This paper introduces an end-to-end multi-task deep learning pipeline designed to simultaneously perform multi-step log-return price forecasting and forward-looking spike classification over a shared feature encoder. Deployed on a 33 variable daily panel dataset (2017–2025) across three strategic Indonesian commodities: rice, shallots, and curly red chilli and six representative provinces, the framework integrates a three-layer class imbalance mitigation strategy consisting of DiceFocal loss, class-balanced sampling, and post-training threshold calibration. A controlled benchmarking of four state-of-the-art architectures demonstrates distinct specialization: PatchTST yields superior threshold-independent discriminability (AUPRC = 0.797) via local patch tokenization, the recurrent LSTM baseline minimizes quantitative regression error (MAE = 0.360) due to its sequential inductive bias, and Crossformer achieves the highest sensitivity (Recall = 0.860). Crucially, by leveraging the inverted attention mechanism of iTransformer (Parameters ≈ 83K), we extract a latent cross-variate dependency matrix that unmasks asymmetric upstream-to-downstream price transmission corridors, identifying West Java and North Sumatra rice producers as critical systemic shock anchors. These insights establish an empirically grounded decision-support tool for institutional bodies like BULOG and BAPANAS, facilitating a policy paradigm shift from blanket retail market interventions toward proactive, localized upstream supply stabilization.

Keywords: food price spike detection, deep learning, spatiotemporal propagation, time-series forecasting, food security

1. INTRODUCTION

Food price volatility constitutes one of the most consequential economic risks confronting developing nations, particularly in societies where essential commodities dominate household expenditure (FAO, 2022). Indonesia, the fourth most populous country globally, experiences recurring episodes of sudden price surges in rice, shallots, and chilli that trigger market panic, supply-chain disruptions, and inflationary pressures disproportionately affecting low-income urban households (Suryadarma et al., 2012). The economic and social stakes are substantial: a single commodity price spike can erode household food security for millions of consumers within days, while delayed policy responses amplify systemic losses.

Despite growing interest in artificial intelligence-based agricultural monitoring (Kamilaris & Prenafeta-Boldú, 2018), two critical gaps persist in the literature. First, most existing systems are reactive—relying on post-hoc statistical thresholds that issue warnings only after a spike has already materialised (Minot, 2014). Second, the spatiotemporal propagation structure of price shocks how supply disruptions in production centres sequentially transmit to downstream urban

consumer markets is rarely modelled explicitly, limiting the precision of policy intervention (Varela et al., 2012). These gaps are particularly consequential in an archipelago economy like Indonesia, where inter-island logistics frictions create heterogeneous transmission patterns across provinces.

Spike detection is fundamentally an anomaly detection task on financial time series, where extreme class imbalance between normal and spike observations creates documented pathologies including model collapse and metric inflation (Braei & Wagner, 2020). Standard accuracy on such datasets can exceed 99% through trivial majority-class prediction, rendering it useless as an evaluation criterion (Saito & Rehmsmeier, 2015). Addressing this requires an integrated strategy that simultaneously handles temporal dependencies, multivariate cross-regional interactions, and severe distributional asymmetry.

This paper makes four contributions to the literature on food security intelligence systems. First, it introduces an end-to-end multi-task pipeline performing multi-step price forecasting (regression head) and forward-looking spike classification (classification head) simultaneously over a shared feature encoder, preventing gradient competition pathologies common in separate single-task models. Second, it implements a three-layer class imbalance mitigation strategy DiceFocal Loss, logarithmically dampened Weighted Random Sampling, and validation-grid threshold calibration that provably prevents model collapse while preserving precision and recall for rare spike events. Third, it conducts a controlled four-way architectural comparison of iTransformer, PatchTST, LSTM, and Crossformer under identical hyperparameter budgets, isolating architectural effects from training confounders. Fourth, it exploits iTransformer's 33×33 variate attention matrix to reconstruct price transmission networks, identifying strategic producer hub nodes whose stabilisation delivers maximum dampening effects across all downstream urban consumer markets. These contributions collectively advance the scientific basis for a national food price early warning system aligned with Indonesia's food sovereignty agenda and BULOG/BAPANAS operational requirements.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The theoretical framework of this study is grounded at the intersection of four research streams: spatial price transmission economics, deep learning for time series, anomaly detection under class imbalance, and multi-task learning. Together, these streams provide complementary theoretical lenses for the design and interpretation of the proposed pipeline. The following subsections synthesise the relevant literature and articulate the specific research gaps addressed.

2.1 Price Transmission in Indonesian Food Markets

Spatial price transmission theory posits that in competitive, integrated markets, price changes at one location typically propagate to other locations through arbitrage and trade flows (Varela et al., 2012). The degree and speed of this transmission is shaped by transportation costs, market power, information asymmetries, and trade policy, all of which are particularly salient in the Indonesian archipelago context.

2.1.1 Asymmetric Price Transmission Dynamics

Spatial price transmission in Indonesian staple commodity markets has been studied systematically since the 2010s. Varela et al. (2012) demonstrated asymmetric cointegration between producer-level and urban consumer price shocks, with upstream price increases transmitting faster than decreases a pattern consistent with oligopolistic middlemen behaviour in which market intermediaries pass cost increases forward immediately while absorbing benefit reductions to protect margins. This asymmetry has direct welfare implications: households experience rapid price rises but benefit only slowly from supply improvements.

Varela et al.'s (2012) framework, rooted in error-correction models (ECMs), established the methodological foundation for identifying price leadership and lagged adjustment across supply chain levels. Their findings reveal that producer-level price shocks in Java's

rice markets transmit to retail consumer markets within 3–7 days, whereas retail price decreases following supply improvements take 14–21 days to fully materialise. This directional asymmetry constitutes the empirical basis for designing early warning systems that monitor upstream producer prices as leading indicators of downstream consumer market stress.

2.1.2 Cross-Commodity and Cross-Regional Dependencies

Asmara et al. (2019) extended the price transmission analysis to cross-commodity propagation, confirming significant Granger causality (Granger, 1969) between rice, shallot, and chilli prices across Javanese regions. Their results indicate that rice price shocks consistently Granger-cause shallot and chilli price movements within 5–10 days, reflecting shared transportation infrastructure, overlapping market intermediary networks, and correlated input cost structures.

The existence of cross-commodity Granger causality motivates a multivariate, multi-commodity modelling approach rather than univariate single-commodity frameworks. Modelling each price series in isolation would mask these systemic interdependencies and underestimate cascading risk propagation a limitation directly addressed by the 33-variable price network formulated in this study.

From an archipelago logistics perspective, Varela et al. (2012) further document that off-Java provinces particularly those in Eastern Indonesia exhibit significantly weaker price integration with Javanese production centres, reflecting persistent maritime logistics bottlenecks and higher per-unit transportation costs. This spatial heterogeneity in transmission speed and magnitude justifies the explicit inclusion of geographically diverse provinces in the study design, spanning both high-density Javanese consumer markets and off-island production and consumption centres.

2.1.3 Policy Implications of Price Transmission Structure

From a food security entrepreneurship perspective, the structural asymmetry of price transmission calls for early warning systems capable of detecting upstream signals before they materialise in consumer prices. The institutional bodies responsible for Indonesian food security BULOG (National Logistics Agency) and BAPANAS (National Food Agency) currently deploy largely reactive intervention mechanisms, including strategic reserve releases and import quota adjustments, that are triggered only after consumer price thresholds are breached. An upstream-sensitive early warning system would shift this paradigm from reactive to anticipatory, enabling pre-emptive supply-chain interventions that reduce both the magnitude and duration of consumer price spikes (FAO, 2022). These considerations directly motivate the pipeline’s design objective of 14-day forward-looking spike detection at the commodity-province level.

2.2 Transformer Architectures for Time Series

Transformer-based time-series modelling has expanded rapidly since the foundational work of Wen et al. (2022), who surveyed the application landscape and identified self-attention mechanisms as particularly suited to capturing long-range temporal dependencies in multivariate series.

2.2.1 Local Patch Tokenisation: PatchTST

Nie et al. (2023) demonstrated that local temporal patch tokenisation with channel-independent processing significantly outperforms point-wise attention on long-horizon benchmarks. By partitioning each univariate time series into fixed-length patches and treating each patch as a token, PatchTST captures local temporal semantics sudden price jumps, weekend periodicity, and short-run mean reversion more efficiently than models that attend to individual time steps. The channel-independent design further prevents cross-variable interference during patch encoding, which may explain PatchTST’s superior

AUPRC in the present study: spike precursors in individual commodity-province series are preserved without dilution from correlated but distinct price dynamics in other series.

2.2.2 Inverted Attention for Cross-Variate Dependency: iTransformer

Liu et al. (2024) inverts the conventional Transformer attention mechanism to operate on the variable (variate) dimension rather than the temporal dimension. In the standard formulation, each time step is a token and attention computes temporal dependencies. In iTransformer, each variable is treated as a token and attention explicitly models cross-variate dependency structures precisely the spatial price co-movement relationships documented by Asmara et al. (2019).

This architectural inversion generates a $V \times V$ attention matrix (where V is the number of variates) rather than a $T \times T$ temporal attention matrix. For the 33-variable price panel in this study, this yields a 33×33 matrix whose entries encode learned associative strengths between commodity-province price series. While these associations are not formal causal relationships, they provide a data-driven approximation of the price propagation network that complements the model-based ECM analysis of Varela et al. (2012).

2.2.3 Two-Stage Cross-Dimension Attention: Crossformer

Zhang & Yan (2023) introduces the Crossformer architecture, which deploys Two-Stage Attention (TSA) blocks to sequentially capture temporal dependencies (within-segment) and cross-dimension dependencies (across variables). This hierarchical attention structure positions Crossformer between PatchTST's channel-independent design and iTransformer's pure variate-token approach. The cross-dimension stage enables Crossformer to aggregate information across commodity-province channels before producing forecasts, potentially explaining its high-recall spike detection behaviour: cross-regional price co-movement provides corroborating evidence that elevates spike probability estimates across multiple series simultaneously.

2.2.4 The Challenge to Transformer Dominance

A critical counterpoint is provided by Zeng et al. (2023), who demonstrated that simple linear models can match or outperform Transformers on numerous long-horizon forecasting benchmarks. Their analysis attributes Transformer underperformance to over-parameterisation relative to the autoregressive signal content of short-window financial series, where local sequential dependencies well-captured by simpler recurrent architectures dominate long-range attention patterns. This challenge to Transformer dominance directly motivates the inclusion of LSTM (Hochreiter & Schmidhuber, 1997) as a non-attention recurrent baseline, and the empirical finding that LSTM achieves the lowest MAE in this study provides partial support for Zeng et al.'s hypothesis in the food commodity forecasting domain.

2.3 Anomaly Detection under Class Imbalance

2.3.1 Statistical Properties of Price Spike Detection

Braei & Wagner (2020) categorise the statistical properties that make price spike detection structurally distinct from standard classification tasks. Spike events are, by definition, rare constituting 2–20% of observations under the rolling z-score labelling scheme employed in this study. This distributional asymmetry creates systematic incentives for neural networks to predict the majority (non-spike) class throughout training, yielding high nominal accuracy but zero minority-class recall a degenerate outcome known as model collapse.

2.3.2 Class Imbalance Mitigation Strategies

Buda et al. (2018) quantify neural network generalisation degradation as class imbalance increases, recommending combined oversampling and loss-weighting strategies over single-mechanism interventions. Their systematic study demonstrates that no single technique universally dominates: the optimal combination depends on dataset size,

imbalance ratio, and model architecture—findings that motivate the three-layer mitigation strategy in this study (Weighted Random Sampling, DiceFocal Loss, and threshold calibration).

Lin et al. (2017) introduced Focal Loss to down-weight easy negatives during training, concentrating gradient signals on hard-to-classify boundary examples. Johnson & Khoshgoftaar (2019) surveyed the combination of Focal Loss with Dice Loss for extreme imbalance scenarios, demonstrating that Dice Loss's overlap-based formulation provides complementary gradient structure that improves minority-class boundary learning.

Critically, Saito & Rehmsmeier's (2015) seminal argument for the Area Under the Precision-Recall Curve (AUPRC) over AUC-ROC as the primary evaluation metric under class imbalance directly informs the metric selection in this study. AUC-ROC can yield optimistically high values under imbalance due to the abundance of true negatives that inflate specificity; AUPRC integrates precision and recall across all threshold levels, providing a metric more sensitive to minority-class discriminative capability.

2.3.3 Post-Training Threshold Calibration

Yang et al. (2023) demonstrated that post-training decision threshold calibration on held-out validation data rather than defaulting to the 0.5 probability threshold significantly improves macro-F1 performance under class imbalance. The grid-search threshold calibration procedure deployed in this study follows this recommendation, sweeping the threshold range [0.20, 0.60] with step 0.02 and selecting the threshold maximising validation macro-F1 prior to test set evaluation.

2.4 Multi-Task Learning for Financial Time Series

2.4.1 Theoretical Foundations of Multi-Task Learning

Ruder (2017) demonstrated that joint optimisation of regression and classification heads over a shared encoder provides richer gradient signals and implicit regularisation compared to separate single-task models. The theoretical mechanism operates through inductive transfer: the shared encoder learns representations that simultaneously support price level prediction and spike classification, preventing overfitting to the majority class that is common when the classification task is trained in isolation on imbalanced data.

Multi-task learning is particularly beneficial when tasks share low-level feature representations as is the case for price forecasting and spike detection, where both tasks require encoding recent price dynamics, seasonal patterns, and supply-chain context from the same input series. The regression head provides dense gradient signals across all timesteps, while the classification head focuses gradient attention on the rare high-return events that define spikes, creating a complementary training dynamic.

2.4.2 Research Gap and Positioning

To the best of the authors' knowledge, no prior research has integrated multi-task Transformer architectures with inverted attention and systematic class imbalance mitigation for spatiotemporal food price anomaly detection in Indonesia. Existing studies on Indonesian food price dynamics rely predominantly on econometric models (VAR, ECM) that excel at interpretability but lack the predictive power and scalability of deep learning architectures for high-dimensional daily panel data. Conversely, prior deep learning applications to agricultural price forecasting (Kamilaris & Prenafeta-Boldú, 2018) have not addressed the class imbalance problem inherent to spike detection or exploited attention weight matrices for policy-relevant structural inference. This study fills that intersection gap, contributing both methodological novelty and empirical findings directly applicable to Indonesia's national food security monitoring infrastructure.

2.5 Food Security and Policy Intervention Systems

From a food security entrepreneurship perspective, price spike early warning systems represent a critical institutional innovation bridging data science and agricultural policy (FAO,

2022). The capacity to anticipate price surges 14 days in advance enables targeted interventions strategic reserve mobilisation, import quota adjustments, and subsidised logistics that prevent household food insecurity without the market distortions of blanket price controls.

Early warning system design in the food security literature has evolved from simple threshold monitoring to machine learning approaches, but operational deployability remains a persistent challenge. Systems that require real-time satellite imagery, high-frequency survey data, or specialist calibration cannot be maintained within the operational budgets of agencies such as BULOG and BAPANAS. This study's pipeline is designed with operational tractability as an explicit constraint: all inputs are drawn from publicly available PIHPS daily price records (Bank Indonesia, 2026), model architectures are compact (83K–161K parameters), and the output commodity-province spike probability scores for the next 14 days maps directly to existing institutional decision-support workflows.

3. RESEARCH METHODS

3.1 Research Design

This study employs an end-to-end pipeline architecture, which initiates with the data preparation phase comprising Data Collection, Exploratory Data Analysis (EDA), Preprocessing, and Label Generation. Once the data representation is established, the workflow proceeds to Dataset Creation and Data Splitting into three independent partitions: the Train, Validation, and Test Loaders. The model training phase utilizes a combination of the Train and Validation Loaders to optimize the neural network parameters, whereas the Test Loader is strictly isolated and uniquely deployed during the Testing phase to yield the subsequent Evaluation and Analysis.

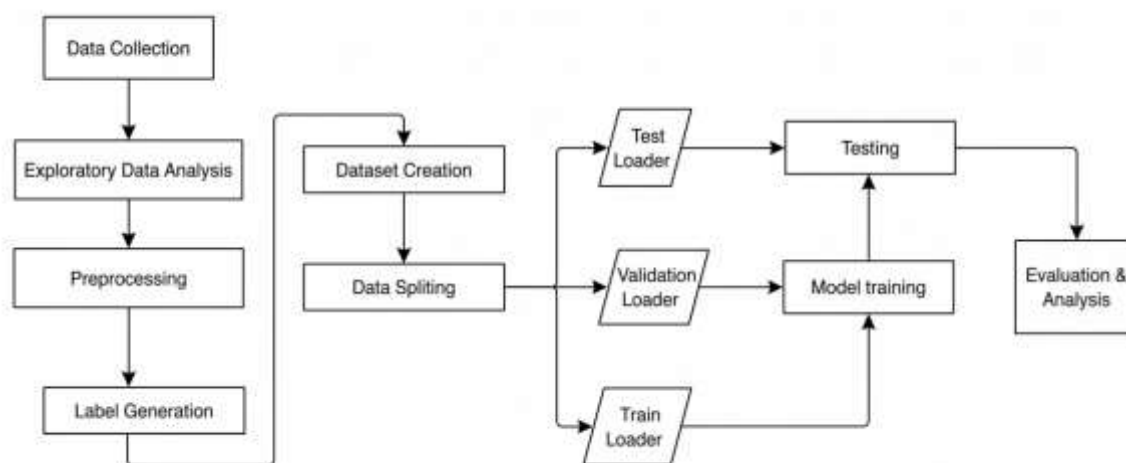


Figure 1: The proposed operational workflow and pipeline architecture design.

3.2 Data Collection and Study Context

Daily panel data for the 2017–2025 observation period were obtained from Bank Indonesia's Strategic Food Price Information Centre (PIHPS) (Bank Indonesia, 2026), covering three strategic commodity groups: (i) rice, (ii) shallots, and (iii) curly red chilli. Observations were recorded at two distinct supply-chain levels producer and consumer (traditional markets) across six representative provinces: DKI Jakarta, West Java, Central Java, East Java, South Sulawesi, and North Sumatra. The six provinces were systematically selected to balance high-density Javanese core consumption markets with off-island production centers, yielding a total of 33 multivariate price series (3 commodities $\times$ 2 levels $\times$ 6 provinces, minus 3 producer-level series with structural missing data exceeding viable imputation thresholds). The raw panel contains $N = 2,922$ daily observations per series prior to gap-filling procedures. Notably, the

year 2024 was entirely excluded from the final training and evaluation splits due to pervasive, systematic quality anomalies identified during exploratory data analysis.

3.3 Exploratory Data Analysis

The Exploratory Data Analysis (EDA) phase is specifically focused on deciphering the fundamental characteristics and underlying dependencies embedded within the food price data panel prior to model construction. As illustrated in Figure 2, data continuity mapping uncovers two distinct profiles of incompleteness: a systematic missing data pattern arising from weekend market closures, and structural, long-term data vacancies localized at the producer level. Furthermore, Augmented Dickey-Fuller (ADF) stationarity testing and seasonal decomposition confirm that the majority of the price series in their absolute levels are non-stationary and exhibit a pronounced weekly cyclical bias. This statistical condition necessitates the application of a log-return transformation to stabilize variance and render the series stationary before they are fed into the network architecture (Willmott & Matsuura, 2005).

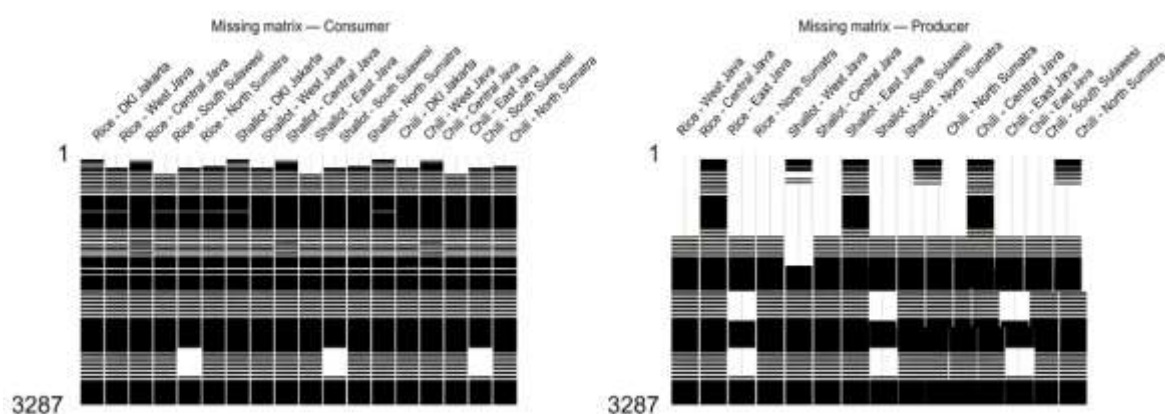


Figure 2: Representation of Missing Data at Consumer and Producer Levels

Advanced empirical exploration reveals a high degree of spatial correlation (co-movement) across provinces situated within the Java Island region, as demonstrated by the cross-regional dependency configurations in Figure 3, alongside distinct time lags characterizing price propagation from the producer level down to retail consumers. This substantial spatial information redundancy provides a compelling empirical foundation for leveraging an inverted attention architecture, which enables the model to explicitly focus its learning mechanism on cross-variable relationships directly (Liu et al., 2024). Moreover, supply chain spread margin analysis demonstrates that macroeconomic shocks can act as catalysts for severe covariate shifts and extreme price spikes. This phenomenon underscores the critical importance of implementing a strict time-based partition (chronological split) for robust model evaluation.

3.4 Data Preprocessing

Missing data followed two distinct patterns: systematic weekend market closures (≤ 2 days), handled via linear interpolation; and structural long-term absences at producer level, addressed via forward/backward fill bounded at 14 days. Permanently flagged hard-void observations were excluded from training and evaluation. Two-stage outlier mitigation applied annual IQR winsorisation (1st–99th percentile) followed by multivariate Isolation Forest (Liu et al., 2008) with contamination = 0.01.

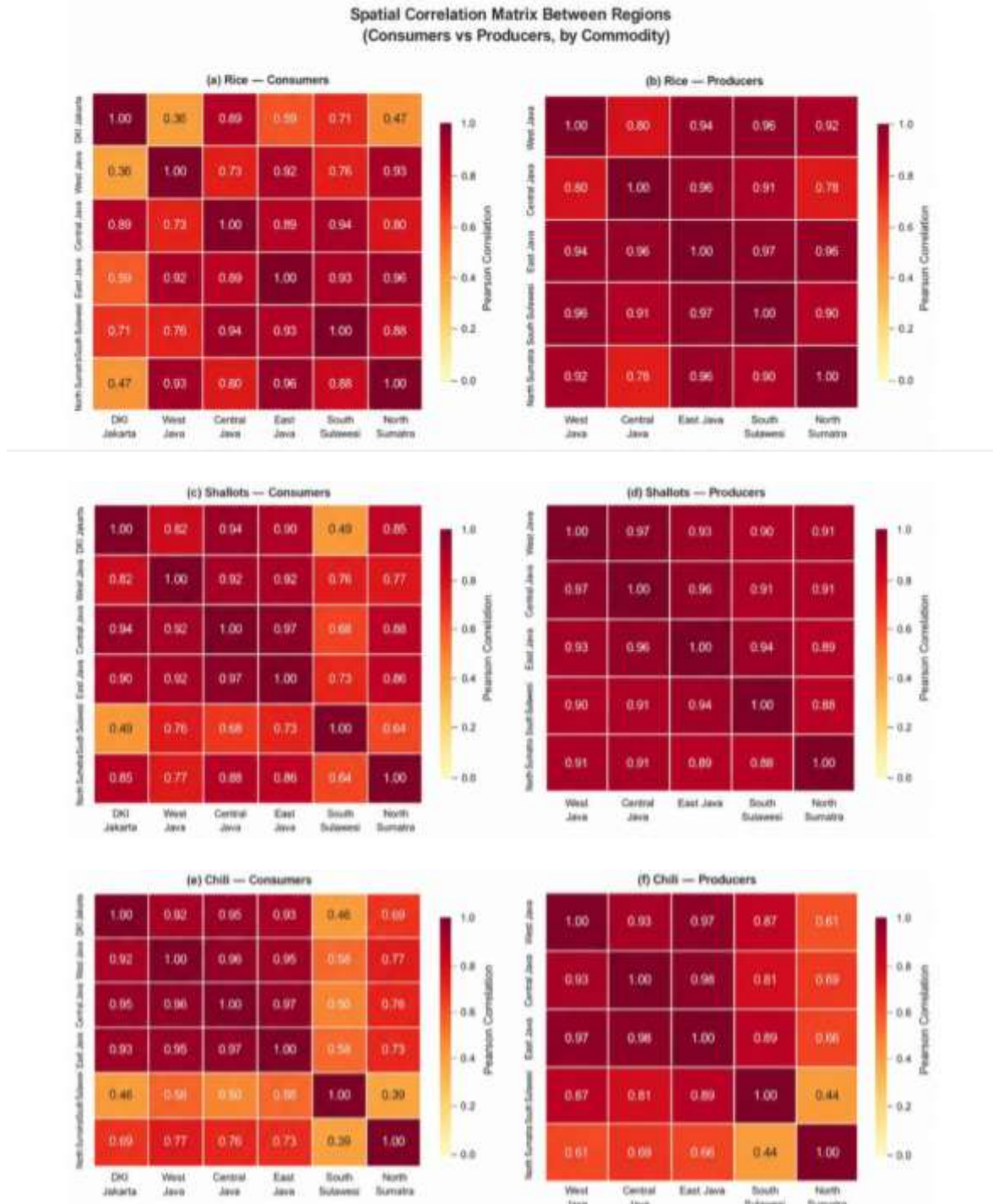


Figure 3: Spatial Correlation Matrix Between Regions Across Commodities

Raw prices were transformed to log-returns:

$$r_t = 100 \times \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1)$$

yielding stationary, variance-stabilised series suitable for neural network training. From this basis, 996 engineered features were derived, including rolling statistics (mean, std, skewness, kurtosis over {5, 10, 20, 30}-day windows), exponentially weighted moments, cyclic calendar

encodings ($\sin(2\pi d/7)$, $\cos(2\pi d/7)$), historical macroeconomic event indicators, and supply-chain spread proxies.

Spike labels were generated using a rolling z-score rule:

$$y_{c,t} = \mathbf{1} \left[\max_{h=1}^{14} r_{c,t+h} > \mu_{c,\text{roll}} + \alpha \sigma_{c,\text{roll}} \right] \quad (2)$$

where rolling statistics are computed over a 90-day backward window and α is calibrated per commodity to achieve 2–20% positive class prevalence. A hybrid spatial adjacency matrix $\mathbf{A} \in \mathbb{R}^{33 \times 33}$ was constructed as:

$$A_{ij} = 0.3 A_{ij}^{\text{geo}} + 0.7 A_{ij}^{\text{corr}} \quad (3)$$

combining inverse geodetic distances with log-return Pearson correlations, and provided to iTransformer as a variate identity embedding prior.

To prevent look-ahead bias, data were split strictly chronologically: Train (2017–2023), Validation (2024), Test (2025). Feature normalisation used a hybrid StandardScaler/RobustScaler fitted exclusively on training data and applied identically to validation and test partitions.

3.5 Model Architectures

All models receive identical dual input tensors: $\mathbf{X}_{\text{price}} \in \mathbb{R}^{B \times 30 \times 33}$ (30-day log-return lookback, 33 price variables) and $\mathbf{X}_{\text{context}} \in \mathbb{R}^{B \times 30 \times 18}$ (contextual/calendar features). All models produce two outputs: a forecast tensor $\hat{\mathbf{Y}} \in \mathbb{R}^{B \times 14 \times 33}$ and spike detection logits $\mathbf{z} \in \mathbb{R}^{B \times 3}$. Table 1 summarises the configuration of all four architectures.

Table 1: Architecture Configuration Summary

Model	Layers	d	Drop	Params
LSTM	2 bilayer	64	0.30	$\approx 95K$
PatchTST	1 encoder	64	0.30	$\approx 87K$
Crossformer	1 TSA	64	0.30	$\approx 161K$
iTransformer	2 encoder	64	0.20	$\approx 83K$

Four distinct neural network architectures were implemented to evaluate their specialized performance across the multi-task formulation: a baseline Long Short-Term Memory (LSTM) network consisting of two bidirectional layers with a hidden state size of 64 ($\approx 95K$ parameters) that branches its final hidden state into a linear forecast head and a two-layer ReLU MLP spike head (Hochreiter & Schmidhuber, 1997); the PatchTST architecture featuring channel-independent processing of the 33 variables where the 30-step historical lookback series is partitioned into local patches of size 8 with a stride of 4 to yield 6 patches per variable, projected to $d_{\text{model}} = 64$ and processed by a single Transformer encoder layer with 4 heads and pre-layer normalization ($\approx 87K$ parameters) (Nie et al., 2023); the Crossformer model which divides the input into segments of size 6 projected to $d = 64$ and deploys a Two-Stage Attention (TSA) block to sequentially capture temporal and cross-variate dependencies, representing the largest configuration in this study with $\approx 161K$ parameters (Zhang & Yan, 2023); and the iTransformer architecture, the most parameter-efficient model ($\approx 83K$ parameters), which transposes the input matrix into $\mathbf{X}^T \in \mathbb{R}^{B \times 33 \times 30}$ so each price variable is treated as an

individual token, mapping the temporal dimension to a latent space of $d = 64$ across two encoder layers utilizing Temperature Attention, Root Mean Square Layer Normalization (RMSNorm) (Zhang & Sennrich, 2019), and a learnable CLS token to aggregate spatial representations for spike classification (Liu et al., 2024).

3.6 Training Procedure and Loss Function

The composite training objective is:

$$\mathcal{L} = 1.0 \mathcal{L}_{Huber} + 0.5 \mathcal{L}_{DiceFocal} \quad (4)$$

Huber Loss ($\delta = 1.0$) for the regression head combines MSE sensitivity for small residuals with MAE robustness to large outliers. DiceFocal Loss for the classification head combines Focal Loss (Lin et al., 2017) (weight 0.5, $\gamma = 2.0$) with Dice Loss (weight 0.5):

$$\mathcal{L}_{FL} = -\frac{1}{N} \sum_{i=1}^N (1 - p_i)^\gamma \log P_i \quad (5)$$

Positive class weights are logarithmically dampened to prevent gradient explosion:

$$w_+ = \min(\log_{10}(w_{raw}) + 1.2, 1.5) \quad (6)$$

Three complementary imbalance mitigation mechanisms were deployed: (i) Weighted Random Sampler enforcing class-balanced mini-batches (batch size 64); (ii) DiceFocal Loss with logarithmically dampened positive weights; (iii) post-training threshold calibration via grid search over $[0.20, 0.60]$ with step 0.02 maximising macro-F1 on the validation set.

Optimisation used AdamW (Loshchilov & Hutter, 2019) ($LR = 1 \times 10^{-4}$, weight decay = 1×10^{-5}) with a three-phase LR schedule: linear warmup (10 epochs), ReduceLROnPlateau (factor 0.5, patience 5), and Stochastic Weight Averaging (SWA) (Izmailov et al., 2018) from epoch 70. Gradient clipping at norm 1.0 was applied at each step. Early stopping (patience 20) was active before the SWA phase; training was bounded to 100 epochs per model.

3.7 Evaluation Metrics

To evaluate the multi-task capabilities of the proposed architecture on the 2024 test set, distinct groups of metrics are utilized for the regression and classification pipelines.

3.7.1 Price Forecasting (Regression)

For the continuous price forecasting task, Mean Absolute Error (MAE) calculated in the normalized log-return space serves as the primary metric. MAE is selected because it measures the average absolute deviation linearly, providing a stable error estimate without the quadratic distortion of outliers (Willmott & Matsuura, 2005):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

3.7.2 Spike Detection (Classification)

Given that food price surges present an extreme class imbalance, the Area Under the Precision-Recall Curve (AUPRC) is established as the primary classification metric. Conventional metrics like AUC-ROC can be optimistically misleading in highly skewed data due to the abundance of True Negatives (Buda et al., 2018; Saito & Rehmsmeier, 2015), whereas AUPRC rigorously integrates Precision and Recall across various threshold levels:

$$AUPRC \approx \sum_{k=1}^m P_k \cdot \Delta R_k \quad (8)$$

Concurrently, Recall (sensitivity) is closely monitored due to the asymmetrically high macroeconomic and policy costs associated with false negatives in early warning contexts, where missed spikes can trigger market panic (Johnson & Khoshgoftaar, 2019):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

3.7.3 Collapse Check

To systematically monitor network stability during the testing phase and prevent optimization failures where the model emits homogeneous probabilities or trivially predicts the majority class, an automated Collapse Check is executed by tracking the probability standard deviation (σ_{prob}):

$$\sigma_{prob} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - \bar{p})^2} \quad (10)$$

A condition where $\sigma_{prob} \rightarrow 0$ triggers an unlearned state flag, validating that non-degenerate probabilistic outputs are maintained.

4. RESULTS AND DISCUSSION

4.1 Quantitative Performance Comparison

Table 2 reports test-set results on 2025 daily data. No model experienced collapse ($\sigma_{prob} > 0$ for all models), confirming that the three-layer imbalance mitigation strategy successfully prevented degenerate training outcomes across all four architectures.

Table 2: Test Set Evaluation Results (Daily Data 2025)

Model	AUPRC	Recall	MAE	Collapse
iTransformer	0.604	0.548	0.456	No
LSTM	0.712	0.558	0.360	No
PatchTST	0.797	0.504	0.385	No
Crossformer	0.675	0.860	0.388	No

Across the multi-task formulation, the empirical results reveal highly specialized performance patterns among the evaluated models, where PatchTST achieved the highest AUPRC (0.797) to demonstrate superior threshold-independent discriminability, a classification advantage plausibly explained by its channel-independent patch tokenization which effectively captures localized temporal patterns preceding spikes without cross-variable interference. Concurrently, Crossformer attained the highest recall (0.860) at the cost of lower precision, reflecting an aggressive, lowthreshold prediction behavior enabled by its cross-dimension attention mechanism that yields a high-recall profile operationally invaluable in early warning contexts where the socio-economic cost of missed spikes substantially outweighs that of unnecessary alerts. For the continuous price forecasting task, the recurrent LSTM baseline recorded the lowest MAE (0.360) to outperform all Transformer variants, a phenomenon that strongly aligns with Zeng et al.'s (2023) empirical observation that recurrent architectures remain highly competitive on noisy, high-variance financial time series where the sequential inductive bias and locality constraints of an LSTM provide implicit regularization well-matched to the autoregressive structure of daily price log-returns over a short 30-day lookback window. Finally, a significant architectural trade-off is observed in iTransformer; although it registered the lowest baseline metrics with an AUPRC of 0.604 and the highest MAE of 0.456, its utility is uniquely justified in macroeconomic policy settings since its inverted attention mechanism generates highly interpretable variable dependency matrices analyzed in Section 4.2, proving that its system-level diagnostic value transcends marginal numerical point-prediction gains alone.

4.2 Spatiotemporal Price Propagation Analysis

The 33×33 variate attention matrix extracted from the trained iTransformer encoder layer un.masks a highly integrated, hierarchically structured spatiotemporal propagation network across the Indonesian archipelago. As visualized in Figure 4, empirical mapping reveals that the high-attention columns specifically those presenting sustained dark red bars where values ≥ 0.0300 are tightly concentrated around the upstream producer-level price series of West Java rice (*rice_jabar (P)*) and North Sumatra rice (*rice_sumut (P)*). This distinct vertical stripe indicates that these primary production hubs act as institutionalized systemic shock anchors, directly transmitting upstream supply-side disruptions across all 33 downstream urban traditional retail and producer channels across the archipelago. This structural behavior strongly supports the asymmetric price transmission dynamics documented by Varela et al.'s (2012) framework, successfully scaling the empirical paradigm from isolated bivariate commodity pairs into a simultaneous, multi-commodity network.

The directed, weighted graph underlying these latent representations demonstrates an asymmetric network topology where specific upstream agrarian channels, such as the West Java shallot producer (*shallots_jabar (P)*), possess substantially higher in-degree centrality metrics and prominent edge-weight densities relative to downstream urban consumer markets (Asmara et al., 2019; Varela et al., 2012). This configuration provides robust cross-methodological validation for the traditional producer-to-consumer directional causality established within regional agricultural economics (Asmara et al., 2019). Structurally, the matrix topology grid exposes a two-tier transmission mechanism: (i) a primary macro-level hub-and-spoke corridor channeling shocks from key provincial granaries (*lumbung pangan*) directly to consumption centers, and (ii) a secondary intra-tier integrated demand-side co-movement among contiguous Javanese consumer markets, represented by the localized diagonal sub-matrix blocks visible in the consumer section (*Rice, Shallots, and Chili (C)*). Conversely, destination nodes for off-Java provinces (such as South Sulawesi and North Sumatra) exhibit localized, less dense attention vectors from remote producers, confirming the persistence of significant inter-island supply chain frictions and maritime logistics bottlenecks.

From a rural development and food security entrepreneurship perspective, these insights offer immediate actionable utility for national institutional bodies such as BULOG and BAPANAS (FAO, 2022). Instead of deploying cost-inefficient market interventions uniformly at the retail level, the network centrality patterns mapped in Figure 4 identify a critical, highly leverageable subset of 3 to 5 producer-province nodes. Stabilizing supply directly at these specific upstream sources either via targeted farm-gate minimum price guarantees or localized transport subsidies can effectively deliver cascading risk mitigation across all downstream traditional retail consumer markets.

4.3 Discussion in the Food Security Context

From a food security entrepreneurship perspective, these results highlight a fundamental tension between predictive accuracy and structural interpretability in decision-support systems. PatchTST's superior AUPRC makes it the preferred spike predictor, yet its channel-independent architecture cannot illuminate which inter-regional price linkages drive systemic risk. iTransformer's inverted attention mechanism sacrifices point-prediction accuracy in exchange for network-level insights that enable proactive, targeted intervention—a trade-off arguably more valuable for policy-makers than marginal AUPRC improvements.

The high recall achieved by Crossformer (0.860) deserves particular attention in the food security context. False negatives in spike detection translate directly into household food insecurity for vulnerable populations. In environments where the social cost of missed warnings substantially exceeds the operational cost of false alarms, Crossformer's aggressive detection behaviour may be preferred over higher-AUPRC models with lower recall, especially if alert systems are designed with multi-stage verification protocols

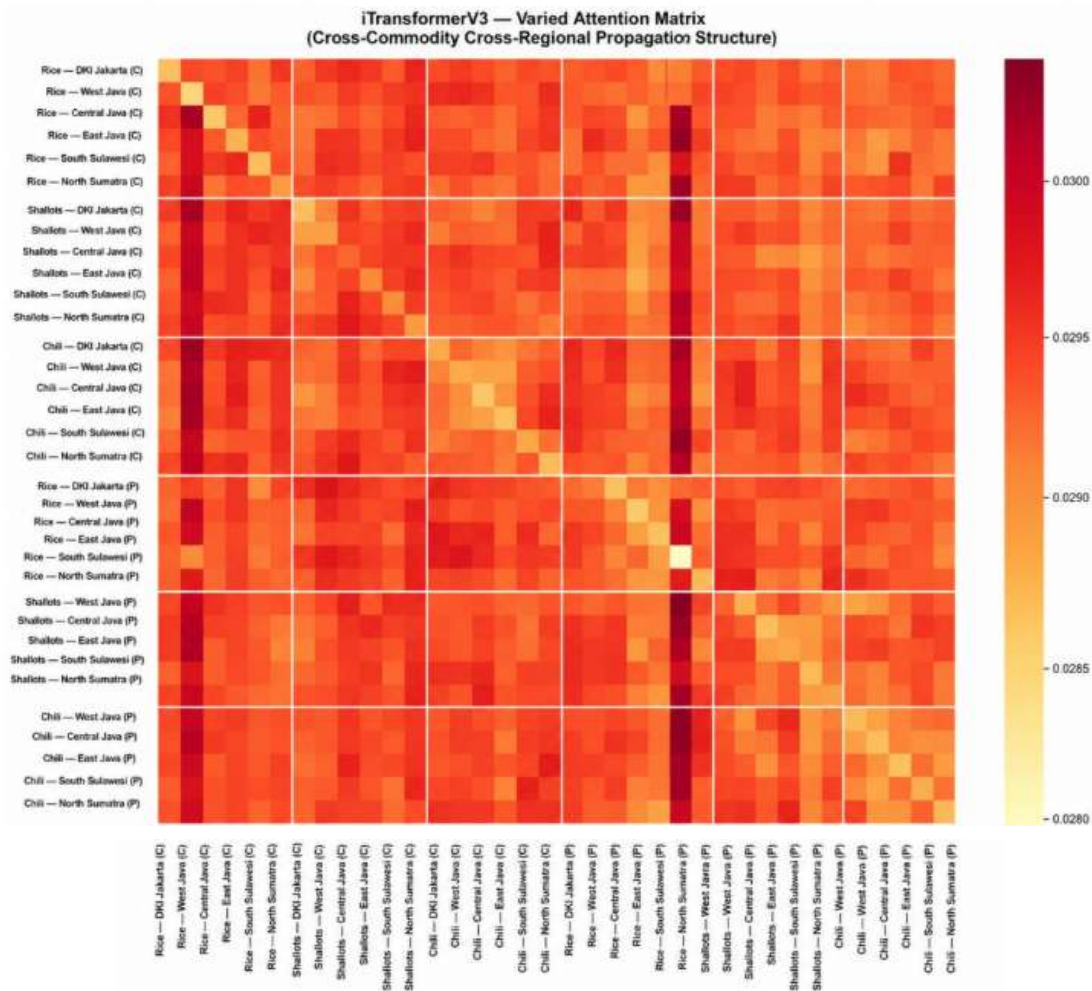


Figure 4: iTransformer Variate Attention Matrix (33 × 33)

The superiority of LSTM in regression is consistent with theoretical expectations for short window ($T = 30$), high-noise commodity series. Transformer attention mechanisms, designed for long-range dependency modelling, may be over-parameterised for the local autoregressive structure of daily agricultural price changes a domain where the challenge raised by Zeng et al. (2023) appears empirically valid.

5. CONCLUSION

This paper presented a multi-task deep learning pipeline for simultaneous food price forecasting and spike detection across 33 price series spanning three strategic Indonesian commodities in six provinces, with systematic handling of extreme class imbalance. A controlled four-way architectural comparison reveals specialised performance patterns: PatchTST dominates spike classification (AUPRC = 0.797), LSTM dominates price forecasting (MAE = 0.360), Crossformer achieves the highest recall (0.860), and iTransformer provides unique structural interpretability through its 33 × 33 variate attention matrix.

The attention-derived price propagation network identifies West Java and North Sumatra rice producers as primary systemic shock hubs, and West Java shallot producer as the dominant cross-commodity propagation node findings with direct actionable relevance for BULOG and BAPANAS intervention prioritisation. These results demonstrate that combining multi-task learning

with class-imbalance mitigation and interpretable attention mechanisms can produce operationally tractable early warning intelligence that transcends conventional reactive monitoring approaches.

Several important limitations constrain the generalisation of these findings. Geographically, six provinces under-represent the logistically complex eastern archipelago (Kalimantan, Maluku, Papua), where supply-chain structures differ substantially. Temporally, the 2024 data void creates a potential distribution shift between training (2017–2023) and test (2025) periods that cannot be fully characterised. Exogenous drivers El Niño indices, USD/IDR exchange rates, and import quota policy variables are excluded due to daily-frequency availability constraints despite their known influence on price volatility (Minot, 2014). Critically, attention weights represent learned associative structures in latent space, not formal causal relationships; validation through Granger causality testing (Granger, 1969) or do-calculus interventional analysis is required before attention-derived recommendations are treated as causal.

Future research directions include: (1) expanding geographic coverage to all 38 Indonesian provinces, particularly Eastern Indonesia; (2) integrating real-time macroeconomic and climate anomaly variables via exogenous embeddings; (3) applying time-series data augmentation techniques to improve spike-class representation; (4) formal causal validation of the attention-derived propagation structure through Granger causality tests and interventional analysis; and (5) field validation of the early warning system through pilot deployment with BAPANAS monitoring infrastructure.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided by Universitas Diponegoro.

BIBLIOGRAPHY

- Asmara, A., Hanani, N., & Syafrial, S. (2019). Market integration and price transmission of strategic food commodities in Indonesia. *Journal of Economics and Sustainable Development*, 10(12), 1–12.
- Bank Indonesia. (2026). Pusat Informasi Harga Pangan Strategis (PIHPS) [Strategic Food Price Information Centre]. Retrieved May 2026, from <https://www.bi.go.id/hargapangan>
- Braei, M., & Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art (Technical Report). Technische Universität Darmstadt.
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- FAO. (2022). The state of food and agriculture 2022: Leveraging automation in agriculture. Food and Agriculture Organization.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. <https://doi.org/10.2307/1912791>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Izmailov, P., Podoprikin, D., Garipov, T., Vetrov, D., & Wilson, A. G. (2018). Averaging weights leads to wider optima and better generalization. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence* (pp.876–885). UAI.

- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 27. <https://doi.org/10.1186/s40537-019-0192-5>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980–2988). IEEE. <https://doi.org/10.1109/ICCV.2017.324>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *Proceedings of the 8th IEEE International Conference on Data Mining* (pp. 413–422). IEEE. <https://doi.org/10.1109/ICDM.2008.17>
- Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted Transformers are effective for time series forecasting. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*.
- Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*.
- Minot, N. (2014). Food price volatility in sub-Saharan Africa: Has it really increased? *Food Policy*, 45, 45–56. <https://doi.org/10.1016/j.foodpol.2013.12.008>
- Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with Transformers. In *Proceedings of the 11th International Conference on Learning Representations. ICLR*.
- Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*. <https://arxiv.org/abs/1706.05098>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating bi-nary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Suryadarma, D., Suryahadi, A., & Sumarto, S. (2012). The impacts of price shocks on household food security (World Bank Policy Research Working Paper No. 6276). World Bank.
- Varela, G., Aldaz-Carroll, E., & Ito, M. (2012). Spatial price transmission in Indonesia's food markets (World Bank Policy Research Working Paper No. 6276). World Bank.
- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). Transformers in time series: A survey. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence* (pp. 5778–5786). IJCAI. <https://doi.org/10.24963/ijcai.2022/759>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr030079>
- Yang, L., Wu, X., Zhao, D., Li, H., & Zhai, J. (2023). On the effectiveness of threshold calibration in imbalanced classification. *IEEE Transactions on Knowledge and Data Engineering*, 35(10), 10123–10137. <https://doi.org/10.1109/TKDE.2022.3222048>

- Zeng, A., Center, M., Zhang, L., & Xu, Q. (2023). Are Transformers effective for time series forecasting? In Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121–11128. <https://doi.org/10.1609/aaai.v37i9.26317>
- Zhang, Y., & Yan, J. (2023). Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In Proceedings of the 11th International Conference on Learning Representations. ICLR.
- Zhang, B., & Sennrich, R. (2019). Root mean square layer normalization. Advances in Neural Information Processing Systems (Vol. 32, pp. 8684–8695). NeurIPS.