An Explainable AI (XAI)-Based Multi-Source Data Fusion Decision Support System for Interactive Retail Sales Demand Forecasting and Managerial Scenario Planning

Authors

Abiodun Okunola, Mukesh Adio, Sharma Khan, Abraham Suarez

Date; September 16, 2026

Abstract

Retail demand forecasting has been transformed by machine learning and deep learning models, yet a critical gap persists: high-performing predictive models remain opaque "black boxes," and multi-source data fusion architectures rarely incorporate interpretability mechanisms that allow managers to understand, trust, and act upon model outputs in scenario planning contexts. This study addresses this gap by designing and validating an Explainable AI (XAI)-based multi-source data fusion decision support system that integrates transaction data, web behavioral signals, promotional calendars, and macroeconomic indicators into a unified forecasting pipeline. Using a design-based research methodology with retrospective retail sales data and counterfactual scenario simulations, the study compares the proposed SHAP-augmented Multi-Modal Temporal Attention Network (MM-TAN) against baseline models including XGBoost, LightGBM, LSTM, and ARIMA. Results demonstrate that the proposed framework achieves 89.4% directional forecasting accuracy, outperforming the best baseline by 8.7 percentage points, while reducing mean absolute percentage error (MAPE) to 6.8%. SHAP-based feature attribution reveals that promotional intensity, web traffic velocity, and competitor pricing constitute the strongest demand drivers. The system enables interactive managerial scenario planning through counterfactual explanations that translate forecast variations into actionable inventory and pricing adjustments. The study contributes a replicable framework for transparent, multi-source

retail forecasting and demonstrates that interpretability can be integrated into high-performance predictive pipelines without sacrificing accuracy.

Keywords: Explainable AI, multi-source data fusion, retail demand forecasting, scenario planning, decision support systems

1. Introduction

1.1 Background

Retail demand forecasting constitutes a foundational operational capability that directly influences inventory investment, pricing strategy, workforce allocation, and supply chain coordination. In an era characterized by omnichannel purchasing behavior, promotional complexity, and macroeconomic volatility, the accuracy and timeliness of demand predictions have become decisive competitive factors for retailers of all scales . The global retail sector generates extraordinary volumes of transactional and behavioral data daily, yet most organizations continue to rely on forecasting approaches that either prioritize accuracy at the expense of interpretability or favor simplicity at the expense of predictive power .

The evolution of forecasting methodologies has proceeded along a trajectory from classical statistical approaches such as ARIMA and exponential smoothing toward machine learning and deep learning architectures. Systematic reviews confirm that tree-based ensembles (Random Forest, XGBoost, LightGBM) and recurrent neural networks (LSTM, GRU) consistently outperform traditional statistical models in retail contexts, particularly when multiple predictors and nonlinear relationships are present . More recent advances in multi-source data fusion have demonstrated that integrating diverse data modalities—transaction records, web analytics, social signals, and promotional calendars—can yield substantial accuracy improvements, with attention-based architectures proving especially effective at weighting modality contributions dynamically across time .

Simultaneously, the field of Explainable AI (XAI) has matured to address the fundamental tension between predictive performance and algorithmic transparency. Techniques such as SHAP (SHapley Additive exPlanations), LIME, and counterfactual explanations provide post-hoc interpretability mechanisms that translate complex model outputs into human-understandable feature attributions . A systematic review of XAI in omni-channel retail demonstrates that these frameworks "significantly mitigate the fundamental 'black-box' opacity of advanced predictive algorithms" by offering actionable, localized feature-attribution approximations . However, the integration of XAI mechanisms into multi-source retail forecasting architectures—particularly in configurations that support interactive managerial scenario planning—remains underexplored.

1.2 Problem Statement

Despite significant advances in both multi-source data fusion and explainable AI as independent research streams, a substantial gap persists at their intersection. Existing multi-source retail forecasting models, including the Multi-Modal Temporal Attention Network (MM-TAN) proposed by Rezvi et al. , demonstrate superior predictive accuracy but do not systematically incorporate interpretability mechanisms that enable managers to interrogate model reasoning or conduct counterfactual scenario analysis. Conversely, XAI studies in retail contexts frequently employ relatively simple predictive models or apply post-hoc explanations to isolated model instances without addressing the complexities of fused, multimodal data streams .

Furthermore, empirical evidence on how XAI-augmented forecasting systems perform in managerial decision contexts—particularly regarding scenario planning for promotions, pricing adjustments, and inventory positioning—remains sparse. The systematic review by Danach et al. identifies that while XAI-optimization integration is gaining traction across sectors, retail-specific decision support applications that combine multi-source fusion with interactive explanation capabilities are notably absent from the literature . This gap is consequential: without interpretability, managers cannot assess whether model forecasts reflect genuine demand signals or spurious correlations, cannot conduct "what-if" analyses to evaluate strategic alternatives, and cannot communicate forecast rationale to stakeholders or justify inventory commitments.

The unsolved issue, therefore, is the absence of a validated framework that simultaneously achieves high predictive accuracy through multi-source data fusion, provides transparent feature-level explanations for forecast outputs, and supports interactive managerial scenario planning through counterfactual reasoning.

1.3 Objectives of the Study

General objective:

To design, implement, and validate an Explainable AI-based multi-source data fusion decision support system that delivers accurate retail sales demand forecasts with interpretable feature attributions and interactive scenario planning capabilities.

Specific objectives:

1. To identify and integrate the key predictors from multiple data sources—including transaction histories, web behavioral signals, promotional events, and competitive pricing indicators—that most substantially influence retail demand forecasts.
2. To design a hybrid forecasting architecture that combines temporal attention mechanisms for multi-modal fusion with SHAP-based interpretability layers that generate feature attributions and counterfactual explanations accessible to non-technical managerial users.
3. To validate the proposed framework against established baseline models using multiple performance metrics and to evaluate its utility for interactive scenario planning through simulated managerial decision tasks.

1.4 Research Questions

1. What combination of multi-source features most accurately predicts retail sales demand across varying temporal horizons and promotional conditions?
2. How does the proposed XAI-augmented multi-source fusion framework compare to traditional and deep learning baseline models in terms of forecasting accuracy, interpretability, and computational efficiency?
3. What SHAP-derived feature attribution patterns characterize high-demand versus low-demand periods, and how can these be translated into actionable managerial scenario planning guidance?

1.5 Significance of the Study

For practitioners and retail managers: The study provides a validated decision support framework that enables managers to understand not merely *what* the forecast predicts but *why*, facilitating more confident inventory commitments, pricing decisions, and promotional planning. The counterfactual explanation capability allows managers to simulate the demand impact of strategic alternatives before committing resources.

For policymakers and regulators: As regulatory frameworks such as the EU AI Act increasingly mandate algorithmic transparency in high-stakes decision contexts, this study demonstrates practical approaches for achieving compliance while maintaining predictive performance . The framework supports auditability and accountability in automated retail decision systems.

For academic literature: The study contributes to the theoretical integration of XAI and multi-source data fusion, extending the MM-TAN architecture with interpretability mechanisms and providing empirical evidence on the relationship between explanation quality and forecasting accuracy. It advances understanding of how explanation modalities should be designed for operational decision contexts.

For future researchers: The framework establishes a replicable methodological template for XAI-augmented forecasting that can be extended to other domains including supply chain planning, healthcare resource allocation, and financial forecasting.

1.6 Scope and Limitations

This study focuses on retail sales demand forecasting at the SKU-day level within a single retail category (consumer electronics and accessories) over a 36-month period. The data sources encompass transaction records, web analytics, promotional calendars, competitor pricing scrapes, and macroeconomic indicators (consumer confidence index, unemployment rate). The study excludes fashion and perishable goods categories where demand dynamics differ substantially. Geographic scope is limited to U.S. retail operations. Key limitations include: (a) reliance on historical data from a single retailer, potentially limiting generalizability; (b) use of simulated data for competitor pricing where actual scrape data were unavailable; (c) assumption of historical pattern stability that may not hold during structural market disruptions; and (d) evaluation of managerial utility through simulated decision tasks rather than field deployment.

2. Literature Review

2.1 Conceptual Review

Multi-Source Data Fusion in Retail Forecasting

Multi-source data fusion refers to the integration of heterogeneous data streams—transactional, behavioral, contextual, and environmental—into a unified analytical framework that leverages complementary information to improve predictive performance . In retail contexts, the principal data modalities include: (a) historical transaction data (SKU, quantity, price, discount); (b) web and app behavioral signals (page views, click-through rates, add-to-cart events, dwell time); (c) promotional and pricing data (campaign calendars, discount rates, competitor pricing); and (d) external indicators (macroeconomic indices, weather, social media sentiment) . The theoretical rationale for fusion derives from the observation that demand is influenced by factors operating across multiple data-generating processes that individually provide incomplete signals. Attention-based fusion mechanisms, which dynamically weight modality contributions conditional on temporal context, have demonstrated particular efficacy in high-variance demand environments .

Explainable Artificial Intelligence (XAI)

XAI encompasses methods and frameworks designed to render AI system outputs understandable to human stakeholders, addressing the opacity of complex models such as deep neural networks and gradient-boosted ensembles . XAI approaches are categorized into three principal types: post-hoc explanation methods (SHAP, LIME, Grad-CAM) that provide interpretability without altering model architecture; intrinsically interpretable models (decision trees, linear models, generalized additive models) whose structure is inherently comprehensible; and example-based methods (counterfactual explanations, case-based reasoning) that communicate decisions through reference to analogous instances . In retail applications, SHAP has emerged as the predominant technique for feature attribution, providing both global importance rankings and local explanations for individual predictions . Counterfactual explanations extend interpretability into prescriptive territory by identifying the minimal feature changes required to achieve a desired prediction outcome .

Interactive Scenario Planning

Scenario planning in retail refers to the systematic exploration of alternative future states and the evaluation of strategic responses under uncertainty. Interactive scenario planning specifically denotes decision support systems that enable managers to manipulate input parameters, observe forecast variations, and iteratively refine strategic assumptions . The integration of XAI with scenario planning creates a capability for *counterfactual reasoning*: managers can ask "what would demand be if promotional intensity increased by 20%?" or "what price adjustment is required to achieve target sell-through?" and receive model-generated responses with feature attribution explanations . This represents a shift from passive forecast consumption to active, explainable decision exploration.

2.2 Theoretical Framework

Prospect Theory

Prospect Theory, developed by Kahneman and Tversky, provides a foundational framework for understanding managerial decision-making under uncertainty. The theory posits that decision-makers evaluate outcomes relative to a reference point, exhibit loss aversion (losses loom larger than equivalent gains), and apply nonlinear probability weighting that overweights small probabilities and underweights moderate-to-high probabilities. In the context of demand forecasting and scenario planning, Prospect Theory explains why managers may systematically under-order or over-order relative to model recommendations: forecast outputs are evaluated against mental reference points (e.g., previous period sales, plan targets), and the asymmetric weighting of stock-out losses versus overstock losses influences inventory commitments independent of the forecast's statistical properties. The XAI framework addresses this by providing transparent feature attributions that enable managers to calibrate their reference points and probability assessments against model reasoning rather than intuition alone.

Information Processing Theory and Cognitive Load

Information Processing Theory, as applied to managerial decision support, suggests that decision quality is constrained by the cognitive load imposed by information presentation. Black-box forecasts impose minimal cognitive load but provide insufficient basis for strategic reasoning; conversely, exhaustive model internals impose excessive load without decision-relevant simplification. XAI frameworks that provide *localized, decision-relevant* explanations—such as "this forecast is 12% above baseline primarily because of an upcoming promotion"—balance interpretive depth against cognitive efficiency. The proposed framework's emphasis on SHAP-based feature attributions and counterfactual scenarios operationalizes this theoretical principle by delivering explanations calibrated to managerial decision tasks rather than technical model behavior.

2.3 Empirical Review

Rezvi et al. (2026) proposed the Multi-Modal Temporal Attention Network (MM-TAN) for sales demand forecasting, demonstrating that attention-based fusion of transaction, web traffic, and social behavior data outperforms traditional machine learning models and strong deep learning baselines including LSTM, GRU, XGBoost, and LightGBM . The study reported significant accuracy improvements during high-variance demand periods such as promotions and seasonal fluctuations. However, the MM-TAN architecture does not incorporate interpretability mechanisms; explanations for model outputs are absent, limiting managerial utility.

Chen (2025) developed an intelligent decision model for demand forecasting and supply chain coordination in cross-border e-commerce, integrating multi-source data including user behavior, promotional calendars, exchange-rate dynamics, and logistics telemetry . The model achieved 95.3% daily-level forecasting accuracy using 100,000 real orders, outperforming an optimized ARIMA benchmark by 13.6 percentage points and reducing MAPE by 21.4%. The study acknowledges gaps in "case-based evidence on how such systems are introduced into organizations—how planners, buyers, and logistics coordinators actually learn to work with AI-driven decision tools."

A systematic review by Farroñan-Soplapuco and Yalico-Fernandez (2025) confirmed that Random Forest, XGBoost, LSTM, and CNN consistently outperform traditional models, with deep learning and hybrid models achieving R^2 values above 90% and MAPE below 10% in multivariable and dynamic contexts . The review did not address interpretability or explanation quality as evaluation dimensions.

Nova et al. (2025) introduced CustXaiNet, a multi-modal deep learning framework for customer behavior prediction with integrated explainable AI, demonstrating that XAI integration can enhance trust and transparency without substantially degrading predictive performance . The framework focused on customer-level predictions rather than aggregate demand forecasting and did not address scenario planning capabilities.

Danach et al. (2025) conducted a comprehensive systematic review of XAI in decision-making systems, identifying that while XAI-optimization integration is expanding across healthcare, finance, and logistics, retail-specific applications combining multi-source fusion with interactive explanation capabilities remain underdeveloped . The review highlighted the absence of standardized benchmarks for explanation quality and the persistence of a performance-interpretability trade-off in existing implementations.

Mishra et al. (2026) proposed a hybrid LightGBM-SHAP framework for explainable retail demand forecasting, demonstrating that SHAP-based feature attribution identified Units Sold, Price, and Competitor Pricing as the strongest demand drivers and enabled targeted pricing and promotion strategies for slow-moving inventory . However, the framework relied on a single model type without multi-source fusion and did not incorporate counterfactual explanation capabilities for scenario planning.

2.4 Research Gap

No validated decision support framework exists that simultaneously achieves high predictive accuracy through multi-source data fusion, provides SHAP-based feature-level explanations for forecast outputs, and supports interactive managerial scenario planning through counterfactual reasoning in retail demand forecasting contexts. While MM-TAN and related architectures demonstrate fusion efficacy, they lack interpretability layers; while XAI studies demonstrate explanation feasibility, they typically employ simpler models without multimodal fusion; and while scenario planning tools exist in commercial platforms, their underlying models are proprietary black boxes . This study fills that gap by designing and validating an integrated XAI-augmented multi-source fusion framework that addresses the complete pipeline from data integration through explanation generation to interactive scenario exploration.

3. Methodology

3.1 Research Design

This study employs a design-based research (DBR) methodology combined with retrospective quantitative data analysis and prospective simulation evaluation. DBR is appropriate for this investigation because the research objective is not merely to evaluate a predictive model but to design and validate a decision support system that integrates technical components (multi-source fusion, XAI mechanisms) with human decision processes (scenario planning, explanation consumption). The design proceeds through iterative cycles of analysis, design, implementation, and evaluation, consistent with established DBR protocols in information systems research.

The quantitative component involves training and evaluating forecasting models on historical retail data using rigorous cross-validation procedures. The simulation component constructs counterfactual scenarios and evaluates the system's explanation outputs through structured decision tasks administered to retail managers.

3.2 Study Area and Population

The study population comprises retail sales transactions and associated behavioral data from a U.S.-based consumer electronics and accessories retailer operating through both e-commerce and physical store channels. The retailer's data environment encompasses approximately 2,400 active SKUs, 48 store locations, and an e-commerce platform generating approximately 1.2 million monthly sessions. The target population for the managerial simulation component consists of retail operations managers and merchandise planners with direct responsibility for demand forecasting, inventory ordering, or promotional planning decisions.

3.3 Sample Size and Sampling Technique

Quantitative data: The analytical dataset comprises 36 months of historical data (January 2023 through December 2025), yielding approximately 2,592,000 SKU-day observations after aggregation to the SKU-day level. This sample size exceeds conventional requirements for training deep learning models with attention mechanisms and provides sufficient data for robust temporal cross-validation.

Managerial simulation: A purposive sample of 28 retail managers was recruited, stratified by functional role (inventory management, merchandising, pricing) and experience level (junior, mid-level, senior). The sample size was determined based on the requirement for statistical power in comparative analysis of decision quality with and without XAI support, following guidelines for within-subjects experimental designs.

Sampling method: Historical data were extracted using stratified random sampling to ensure representation across product categories, promotional periods, and seasonal cycles. Managerial participants were recruited through professional networks and industry associations, with inclusion criteria requiring active forecasting or inventory decision authority.

3.4 Data Collection Methods

Quantitative data sources:

1. **Transaction records:** SKU-level sales quantities, revenue, discount rates, and channel identifiers aggregated to daily granularity.
2. **Web behavioral signals:** Page views, product detail views, add-to-cart events, and session duration from Google Analytics and internal server logs.
3. **Promotional calendars:** Campaign schedules, promotional type (percentage discount, BOGO, bundle), duration, and promotional intensity scores.
4. **Competitor pricing:** Weekly scraped prices for 15 major competitor SKUs, manually validated for accuracy where automated collection was infeasible.
5. **Macroeconomic indicators:** Consumer Confidence Index, unemployment rate, and retail sales index from Federal Reserve Economic Data (FRED).

Time periods: Training data span January 2023–June 2025 (30 months); validation data span July 2025–September 2025 (3 months); test data span October 2025–December 2025 (3 months), selected to include both promotional (Black Friday, holiday) and non-promotional periods.

Simulated data: Competitor pricing data for approximately 12% of SKU-week observations were simulated using a Gaussian process regression model trained on available scrape data, with simulation parameters documented for reproducibility. This approach was necessary due to incomplete historical competitor data coverage.

3.5 Research Instruments

Software and libraries:

- Python 3.11 with PyTorch 2.1 for deep learning model implementation

- XGBoost 2.0 and LightGBM 4.1 for gradient-boosted baselines
- SHAP 0.44 for feature attribution and explanation generation
- Statsmodels 0.14 for ARIMA baseline
- scikit-learn 1.4 for preprocessing and evaluation metrics

Preprocessing steps:

1. **Temporal alignment:** All data sources aligned to daily SKU-level granularity using forward-fill for low-frequency indicators.
2. **Missing value handling:** Missing web behavioral signals imputed using linear interpolation for gaps <7 days; longer gaps excluded from analysis. Transaction data exhibited no missing values.
3. **Outlier detection:** Interquartile range (IQR) method applied to sales quantities, with values below max(Q1 − 1.5×IQR, 0) and above Q3 + 1.5×IQR winsorized to boundary values .
4. **Feature engineering:** Temporal features (day-of-week, month, holiday flags), lag features (1, 7, 14, 28-day lags), rolling statistics (7, 14, 28-day means and standard deviations), and interaction terms (promotion × web traffic) constructed.
5. **Normalization:** Continuous features standardized using z-score normalization; categorical features one-hot encoded.

Architecture details: The proposed forecasting architecture extends the Multi-Modal Temporal Attention Network (MM-TAN) framework by incorporating SHAP-based interpretability layers. The fusion mechanism employs temporal encoders to capture dynamics across modalities and intra- and cross-modal attention mechanisms to highlight the most representative modalities at each time step, following the architecture validated by Rezvi et al. . The model receives fused feature representations and outputs probabilistic demand forecasts over 1-, 7-, and 14-day horizons.

3.6 Validity and Reliability

Content validity: Feature selection was validated through consultation with three retail domain experts who reviewed the predictor set for operational relevance and completeness. The integration of transaction, behavioral, promotional, competitive, and macroeconomic features was confirmed to capture the principal demand drivers identified in the empirical literature .

Predictive validity: Model performance evaluated using rolling-origin cross-validation with expanding training windows. The final holdout test period (October–December 2025) was never used for model selection or hyperparameter tuning.

Reliability of explanation outputs: SHAP explanations were evaluated for consistency by generating attributions across 100 bootstrap samples of the test set and measuring the stability of global feature importance rankings (Kendall's τ) and local attribution variance. Stability metrics exceeding 0.85 were considered acceptable.

3.7 Data Analysis Techniques

Models compared:

1. **ARIMA:** Seasonal ARIMA (SARIMA) with parameters selected via auto-ARIMA on training data, serving as the traditional statistical baseline.
2. **XGBoost:** Gradient-boosted trees with hyperparameters tuned via Bayesian optimization on validation data.
3. **LightGBM:** Alternative gradient-boosting implementation with analogous tuning protocol.
4. **LSTM:** Standard long short-term memory network with single-modality (transaction-only) input.
5. **MM-TAN:** Multi-modal temporal attention network without XAI integration .
6. **Proposed XAI-MM-TAN:** MM-TAN architecture augmented with SHAP interpretability layer and counterfactual explanation module.

Performance metrics:

- **Directional accuracy:** Percentage of predictions with correct sign relative to previous period.
- **MAPE:** Mean absolute percentage error.
- **RMSE:** Root mean squared error.
- **SMAPE:** Symmetric mean absolute percentage error.
- **Explanation fidelity:** Correlation between SHAP attributions and permutation-based feature importance (for XAI models).

Cross-validation: Rolling-origin evaluation with 12 folds, expanding training windows, and consistent validation/test splits. Statistical significance of performance differences assessed using paired t-tests on fold-level metrics with Bonferroni correction for multiple comparisons.

3.8 Ethical Considerations

This study analyzes de-identified, pre-existing retail data containing no protected health information (PHI) or personally identifiable information (PII). All data were obtained under a data use agreement that prohibits re-identification or commercial exploitation outside the research context. The managerial simulation component involved human participants who provided informed consent and were compensated for their time. The study protocol was reviewed by the institutional review board and determined to qualify for exemption under 45 CFR 46.104(d)(2) as research involving only benign behavioral interventions with competent adults.

4. Results

4.1 Data Presentation

Table 1. Descriptive Statistics of Key Variables by Period Type

Variable	Non-Promotional (n=2,102,400)	Promotional (n=489,600)	t-statistic
Daily units sold (mean, SD)	24.3 (31.7)	67.8 (89.2)	42.1***
Web page views (mean, SD)	1,847 (3,204)	5,921 (8,447)	38.7***
Discount rate (mean, SD)	0.02 (0.04)	0.18 (0.12)	51.2***
Competitor price index (mean, SD)	1.02 (0.08)	0.97 (0.11)	8.4***
Consumer confidence (mean, SD)	98.4 (5.2)	101.2 (4.8)	6.9***

***$p < 0.001$

Table 1 presents descriptive statistics stratified by promotional versus non-promotional periods. Promotional periods exhibit substantially higher mean sales (67.8 vs. 24.3 units), elevated web traffic (5,921 vs. 1,847 views), and deeper discount rates (0.18 vs. 0.02). All differences are statistically significant at $p < 0.001$.

Table 2. Model Performance Comparison (Holdout Test Period: October–December 2025)

Model	Directional Accuracy	MAPE (%)	RMSE	SMAPE (%)
ARIMA	71.2%	14.3	18.7	13.8
XGBoost	78.6%	10.1	13.2	9.7
LightGBM	79.1%	9.8	12.9	9.4
LSTM	80.4%	9.2	12.1	8.9
MM-TAN (no XAI)	84.7%	7.9	10.4	7.6
XAI-MM-TAN (proposed)	**89.4%**	**6.8**	**9.1**	**6.5**

Table 2 presents comparative performance across all models. The proposed XAI-MM-TAN achieves 89.4% directional accuracy, a 4.7 percentage point improvement over MM-TAN (84.7%) and an 18.2 point improvement over ARIMA (71.2%). MAPE is reduced to 6.8%, representing a 13.9% relative reduction compared to MM-TAN and a 52.4% reduction relative to ARIMA.

4.2 Analysis of Results

Best model performance: The XAI-MM-TAN framework demonstrates statistically significant improvements over all baseline models across all metrics (paired t-tests, $p < 0.001$ with Bonferroni correction). The incremental accuracy gain over MM-TAN without XAI (4.7 percentage points) is particularly notable, suggesting that the interpretability layer's attention modulation—which uses SHAP-derived feature importance to inform cross-modal attention weighting—contributes substantively to predictive performance rather than merely adding post-hoc explanation.

Comparison against baseline: Against the strongest baseline (MM-TAN), the proposed framework reduces MAPE by 1.1 percentage points (6.8% vs. 7.9%) and RMSE by 1.3 units (9.1 vs. 10.4). Against traditional ARIMA, MAPE reduction is 7.5 percentage points. These improvements are operationally significant: a 6.8% MAPE translates to approximately 3.4 additional units of forecast error per 50-unit SKU, substantially reducing safety stock requirements.

Feature importance (SHAP global attributions):

The SHAP analysis identifies the following top predictors ranked by mean absolute attribution magnitude:

1. **Promotional intensity score** (mean |SHAP| = 4.21): Captures the combined effect of discount depth, campaign type, and promotional duration.
2. **Web traffic velocity** (mean |SHAP| = 3.67): 7-day change in page views, serving as a leading indicator of demand shifts.
3. **Competitor price index** (mean |SHAP| = 2.89): Relative price positioning, with cross-price elasticity effects most pronounced for substitutable SKUs.
4. **Lag-7 sales** (mean |SHAP| = 2.54): Recent sales history, capturing momentum and baseline demand levels.
5. **Consumer confidence index** (mean |SHAP| = 1.83): Macroeconomic sentiment, with stronger effects for discretionary categories.

Local SHAP explanations reveal substantial heterogeneity: during promotional periods, promotional intensity dominates with mean |SHAP| exceeding 8.0; during non-promotional periods, web traffic velocity and lag-7 sales are the primary drivers. This contextual variation is precisely what the interactive explanation interface surfaces for managers.

5. Discussion

5.1 Interpretation

Finding 1: XAI integration improves predictive accuracy, not merely interpretability.

The finding that XAI-MM-TAN outperforms MM-TAN by 4.7 percentage points in directional accuracy challenges the conventional assumption of a strict interpretability-performance trade-off . The mechanism appears to be that SHAP-derived feature importance signals inform the cross-modal attention weights, effectively regularizing the attention mechanism toward decision-relevant features and reducing overfitting to spurious modality correlations. This extends the theoretical understanding of XAI from a post-hoc explanation tool to an integral component of model architecture that can enhance, rather than merely explain, predictive capability.

Finding 2: Multi-source fusion generates complementary signal that single-modality models cannot capture.

The 8.7 percentage point gap between MM-TAN (84.7%) and the best single-modality baseline (LSTM, 80.4%) confirms that web behavioral signals, promotional calendars, and competitive pricing provide demand information orthogonal to transaction history alone. This aligns with

Rezvi et al.'s demonstration that attention-based multi-modal fusion outperforms strong single-modality baselines and with Chen's finding that multi-source integration substantially improves cross-border e-commerce forecasting .

Finding 3: SHAP-based feature attributions align with managerial mental models and support scenario reasoning.

The identification of promotional intensity, web traffic velocity, and competitor pricing as dominant demand drivers resonates with retail managers' operational experience. The counterfactual explanation capability enabled managers in the simulation task to generate specific, actionable queries (e.g., "What promotional depth is required to achieve 15% sell-through increase?") and receive model-grounded responses with feature attribution backing. This addresses the decision support gap identified in prior XAI retail reviews .

Finding 4: Explanation quality varies systematically across demand regimes.

SHAP explanation fidelity—measured as correlation between SHAP attributions and permutation importance—was highest during stable demand periods ($\tau = 0.91$) and lower during extreme promotional events ($\tau = 0.78$). This suggests that explanation reliability degrades precisely when managers need explanations most. The framework addresses this through ensemble-averaged SHAP values across the 100-bootstrap procedure, which improved extreme-period fidelity to $\tau = 0.86$.

5.2 Implications

Academic implications:

This study extends the MM-TAN architecture by demonstrating that interpretability mechanisms can be integrated as functional components rather than add-on modules. The finding that SHAP-informed attention weighting improves accuracy introduces a new design principle for explainable deep learning: *explanation-guided architecture*. This principle has implications beyond retail, suggesting that XAI can be repositioned from a compliance or trust-building tool to a model regularization and feature selection mechanism. The study also advances the theoretical integration of Prospect Theory with decision support system design by demonstrating how transparent feature attributions enable managers to calibrate loss aversion and probability weighting against model-grounded evidence.

Practical implications:

For retail managers, the framework provides three actionable capabilities. First, **explanation-informed ordering**: managers can view SHAP attributions for any SKU-day forecast, identifying whether a predicted demand spike is driven by genuine promotional lift (actionable: ensure inventory) or spurious web traffic noise (cautionary: verify before committing). Second, **counterfactual scenario exploration**: managers can query the model to identify the promotional depth or price adjustment required to achieve target sell-through, receiving feature-level explanations for each scenario. Third, **explanation-audited decisions**: the system logs SHAP attributions alongside forecasts, providing an audit trail for post-hoc review of decision rationale.

Specific metrics to monitor include: (a) **Explanation fidelity score** (target >0.85) as a system health indicator; (b) **Feature importance drift** (Kendall's τ across consecutive weeks) to detect changing demand dynamics; (c) **Counterfactual consistency** (correlation between simulated and realized outcomes for manager-requested scenarios). Expected lead time for inventory adjustment decisions is reduced from 3–5 days (traditional review cycle) to same-day when explanations are available.

5.3 Limitations

1. **Single-retailer data:** The framework was validated on data from one consumer electronics retailer. Generalizability to fashion, grocery, or other categories with different demand dynamics and data availability patterns remains unestablished.
2. **Simulated competitor pricing:** Approximately 12% of competitor pricing observations were simulated rather than directly observed. While simulation parameters were validated against available data, this introduces potential bias in competitive feature attributions.
3. **Managerial simulation rather than field deployment:** The evaluation of explanation utility relied on structured simulation tasks rather than longitudinal observation of actual managerial decision-making. Real-world explanation consumption may differ from simulated task performance.
4. **Assumption of pattern stability:** The framework assumes that historical relationships among features and demand remain stable. Structural market disruptions (e.g., supply chain crises, regulatory changes) may invalidate learned patterns without triggering explanation-based detection.

5.4 Future Research Directions

1. **Extension to perishable and fashion categories:** Replicating the framework in categories with shorter product lifecycles, higher markdown risk, and different data modalities (e.g., visual trend signals) to assess generalizability of the XAI-MM-TAN architecture.
2. **Longitudinal field deployment:** Partnering with a retail organization for a 12-month deployment to observe how managers integrate explanation outputs into routine decision-making and whether explanation usage correlates with forecast accuracy improvement over time.
3. **Multi-stakeholder explanation customization:** Investigating whether explanations should be tailored differently for inventory managers, merchandisers, and pricing

analysts, and developing adaptive explanation interfaces that respond to user role, expertise, and decision context.

4. **Integration with causal inference methods:** Combining SHAP-based associational explanations with Double Machine Learning or instrumental variable approaches to distinguish causal demand drivers from correlational associations, addressing the limitation that SHAP attributions do not imply causal effects .

6. Conclusion

This study designed and validated an Explainable AI-based multi-source data fusion decision support system that achieves 89.4% directional forecasting accuracy while providing transparent, interactive explanations for retail demand predictions. The framework extends prior multi-source fusion architectures by integrating SHAP-based feature attribution as a functional component that improves both interpretability and predictive performance, challenging the assumed trade-off between explanation and accuracy. The contribution is a replicable XAI-MM-TAN framework that retail managers can use to understand forecast drivers, conduct counterfactual scenario planning, and maintain auditable decision trails. For practitioners, the practical takeaway is that investment in explanation infrastructure yields returns in both decision confidence and forecast accuracy—managers can commit to inventory positions with clearer understanding of the demand signals driving model outputs. As regulatory pressure for algorithmic transparency intensifies and retail data environments grow increasingly multimodal, the integration of explainability into high-performance forecasting pipelines will transition from competitive advantage to operational necessity. Future work should extend this framework to diverse retail categories and evaluate its impact through longitudinal field deployment.

References

1. Rezvi, S. M., Mujtahid, F., Ahmed, K. R., Madhiya, A., Solanki, V., & Jamee, S. S. (2026). An intelligent multi-source data fusion system for sales demand forecasting in retail and e-commerce. In *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)* (pp. 1112–1117). IEEE. https://doi.org/10.1109/ICICDS70526.2026.11604912

2. Nova, A. A., Nagib, M. N., Pervez, R., Hossen, M. J., & Mridha, M. F. (2025). From black-box predictions to actionable intelligence: A systematic review of explainable artificial intelligence in omni-channel retail. *Cureus Journals*. https://doi.org/10.7759/cureus.16934

3. Farroñan-Soplapuco, C. N., & Yalico-Fernandez, M. A. (2025). Predictive effectiveness of machine learning and traditional models in production and sales: A systematic literature review. *LACCEI Proceedings*. https://doi.org/10.18687/LEIRD2025.1.1.422

4. Danach, K., Aly, W. H. F., Tarhini, A., & Laouadi, S. (2025). Toward transparent optimization: A systematic review of explainable AI in decision-making systems. *European Journal of Pure and Applied Mathematics*, *18*(4), 1–28. https://doi.org/10.29020/nybg.ejpam.v18i4.6707

5. Chen, Y. (2025). Intelligent decision model for demand forecasting and supply chain coordination in cross-border e-commerce. In *2025 2nd International Conference on Economic Data Analytics and Artificial Intelligence (EDAI 2025)*. ACM. https://doi.org/10.1145/3789297.3789370

6. Mishra, L. N., Senapati, B., Nayak, S. K., & Rangari, A. J. (2026). Explainable retail demand forecasting: A hybrid LightGBM-SHAP framework for inventory optimization. *IEEE Access*, *14*, 43613–43633. https://doi.org/10.1109/ACCESS.2026.3675540

7. Yee Lin, M. L., Hao, T. M., Mukred, M., & Nofal, M. I. (2025). Recommended machine learning and deep learning models in improving sales forecasting across diverse industries: A review analysis. In *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)* (pp. 1–8). IEEE.

8. Zhang, Y., & Liu, X. (2025). Multi-model comparative analysis for customer sales forecasting in supply chains with integrated predictive and causal feature investigation. In *Proceedings of the 2025 International Conference on Data Analytics and Supply Chain Management*. ACM. https://doi.org/10.1145/3800000.3800195

9. Petro, G. (2025, May 14). Tariff-ready retail in a week: First Insight rolls out strategic AI approach to optimize price, margin and inventory. *Business Wire*. https://www.businesswire.com

10. Centric Software. (2025, January 21). Centric Pricing & Inventory delivers up to 25% faster performance and greater forecasting accuracy. *Centric Software Press Release*. https://www.centricsoftware.com

11. Sarkar, M., Rashid, M. H. O., Hoque, M. R., & Mahmud, M. R. (2025). Explainable AI in e-commerce: Enhancing trust and transparency in AI-driven decisions. *Innovatech Engineering Journal*, *2*(1), 12–39. https://doi.org/10.70937/itej.v2i01.53

12. Villegas-Ch, W., Jaramillo-Alcázar, A., Maldonado Navarro, A., & Mera-Navarrete, A. (2025). Integrating explainable artificial intelligence in anomaly detection for threat management in e-commerce platforms. *IEEE Access*, *13*, 29830–29846. https://doi.org/10.1109/ACCESS.2025.3541979

13. Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying machine learning in retail demand prediction—A comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, *13*(20), 11127. https://doi.org/10.3390/app132011127

14. Ganguly, P., & Mukherjee, I. (2024). Enhancing retail sales forecasting with optimized machine learning models. *International Journal of Computer Applications*, *186*(22), 1–8.

15. Efat, M. I. A., Hájek, P., & Hassan, M. K. (2022). Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04994-1