Machine Learning for Human Intelligence
2026, Vol. 4, Issue 1

Research Article

Next-Gen Supply Chains: Harnessing Artificial Intelligence for Predictive Demand and Agile Operations

Brian Jeffry[1⋆], Zheng Wan[1], and Shengjun Meng[1]

[1]Faculty of Engineering and Management, Universidad Complutense de Madrid (UCM), Madrid, Spain; Email: brianjeffry@ucm.es; zhengwan20@ucm.es; meng@ucm.es
⋆Corresponding author: Brian Jeffry (brianjeffry@ucm.es)

Article History

Academic Editor:
Dr. Muhammad Sajid
Submitted: October 19, 2025
Revised: January 12, 2026
Accepted: March 1, 2026

Keywords:
Artificial intelligence, Demand forecasting, Machine learning, Supply chain optimization, Predictive analytics, Deep Learning

Abstract

The emergence of artificial intelligence (AI) has ushered in a new era of efficiency and precision across multiple sectors, particularly in next-generation supply chains, demand forecasting, and inventory management. Conventional inventory management methods, typically reliant on historical data and basic statistical models, are inadequate for addressing the dynamic and intricate nature of modern marketplaces. Artificial intelligence, through its advanced algorithms and machine learning capabilities, enables more predictive and agile supply chain operations. This study investigates the application of AI to enhance inventory management and demand forecasting by proposing a hybrid CNN-LSTM model that integrates Convolutional Neural Networks (CNN) for local pattern extraction and Long Short-Term Memory (LSTM) networks for temporal sequence modeling. Using an extensive dataset from the fourth Kaggle competition, the proposed approach captures both short-term variations and long-term dependencies in multi-dimensional time-series data. The model formulates demand prediction as a time-series regression task and demonstrates superior performance compared to traditional statistical methods, standalone machine learning models, and individual deep learning approaches. The results indicate significant improvements in prediction accuracy and robustness, supporting more effective predictive demand management and agile supply chain operations, ultimately contributing to enhanced inventory performance and overall supply chain efficiency.

1 Introduction

Demand forecasting is a vital function in contemporary supply chain management, directly impacting inventory optimization, production planning, and customer satisfaction [1, 7]. Conventional forecasting techniques, although historically successful, encounter growing difficulties in the current unpredictable and intricate market landscape marked by swift changes in consumer behavior, global supply chain interruptions, and unparalleled data volumes [8, 2]. The advent of artificial intelligence and machine learning technology has generated novel potential to improve forecasting precision and responsiveness. The use of AI in demand forecasting signifies a transition from rule-based statistical models to adaptive, learning-driven systems that can discern intricate patterns in multidimensional datasets [9]. This change is especially pertinent as firms contend with escalating demand volatility, abbreviated product lifecycles, and the necessity for real-time decision-making capabilities [10, 5]. In contrast to conventional time-series studies that predominantly depend on previous internal sales data, AI models can include a wide range of external variables [11, 3].

This encompasses rival price, social media sentiment, local weather patterns, macroeconomic data, and geopolitical events, resulting in a more comprehensive and nuanced demand signal [12, 13]. This improved analytical capacity enables AI systems to transcend mere extrapolation of historical trends and instead produce genuinely predicted insights [14, 15]. Through ongoing analysis of new data, these models can identify nuanced, emergent connections, such as the impact of a viral social media post on regional product demand or the effect of a predicted heatwave on sales of particular seasonal commodities [16, 17]. This facilitates a transition from reactive supply chain modifications to a proactive and prescriptive approach, allowing organizations to simulate diverse market scenarios and optimize their inventory and logistics networks to seize opportunities and mitigate risks prior to their full emergence [18, 4]. This study will examine the effectiveness of diverse AI techniques, including Long Short-Term Memory (LSTM) networks for capturing temporal dependencies in sequential data, convolutional neural networks (CNNs) for detecting spatial patterns in regional sales data, and transformer-based models that assess the significance of various predictive factors [19]. In the constantly changing realm of international trade, prominent retailers significantly influence market dynamics and consumer expectations. This study examines the operations of a leading corporate entity recognized as a large retailer with a strong global presence and a varied product selection [20]. At the core of its business is a unique pricing methodology that deliberately lowers grocery costs in its operational markets. The company's market dominance transcends retail, encompassing supplementary services and reinforcing its position as a formidable entity in the worldwide market. The retail giant's activities provide the extensive dataset that centers this analysis. This dataset is fundamental for the creation and assessment of an advanced hybrid model, designed for real-time sales data analysis and significant improvement in forecasting precision [23]. This study examines the complex relationship between the company's market strategy, product diversification, and its evolving position in the worldwide marketplace by utilizing the extensive information contained in this dataset.

Additionally, the utilization of ensemble approaches, including random forests and gradient boosting machines (XGBoost), will be analyzed for their potential to enhance resilience and accuracy by amalgamating several weaker models into a more potent, composite predictor [21]. The tangible advantages of effectively executing these AI-driven systems surpass mere enhancements in accuracy metrics. Organizations can significantly decrease stock-out occurrences and surplus inventory, resulting in optimized working capital and diminished warehousing expenses [22]. This improved forecasting accuracy facilitates a more adaptable and robust supply chain, capable of proactive rather than reactive modifications. For example, early identification of a demand increase facilitates strategic procurement and production planning, alleviating the effects of shortages and delays [6]. Nonetheless, the journey toward AI adoption is laden with considerable obstacles. Examining critical difficulties involves the need for high-quality, granular, and cohesive data, often scattered across numerous organizational systems [24].

The opaque nature of intricate deep learning models can pose challenges to interpretability, hindering supply chain managers' ability to trust and respond to projections without a comprehensive understanding of the underlying elements [25]. Moreover, a significant skills gap exists, necessitating investment in both technology and human capital to develop and sustain these sophisticated analytical capabilities. The principal aim of this research is to deliver an exhaustive analysis of AI-based demand forecasting methodologies, scrutinizing their theoretical underpinnings, practical utilization, and implementation obstacles. This study seeks to connect academic research with practical application by discussing best practices, emerging trends, and future research paths in this swiftly expanding domain. We improve the supply chain optimization model by employing machine learning techniques, specifically Multilayer Perceptron (MLP), to derive precise and dependable client demand values. This method substitutes the forecasting of future parameters derived from historical data, as utilized in conventional models. This paper outlines a four-step procedure. 1) Data collection, 2) Input and objective formulation, 3) Customer demand prediction. The fourth phase in assessing the performance of the optimal portfolio, based on the forecasted data, entails utilizing all three datasets: historical, anticipated, and validation data. The findings are compared between the historical and validation case solutions, as well as between the predicted and validation cases. This article asserts that AI-driven demand forecasting is not just a minor enhancement but a fundamental component for establishing competitive, efficient, and resilient supply chains essential for success in the 21st century.

2 Related Work

Demand forecasting is an essential element of supply chain management, directly impacting a company's operational efficiency, cost control, and customer satisfaction. Historically, these operations have depended on past data and basic statistical models to forecast demand and oversee inventories [26]. Although useful to some degree, these conventional methods frequently fail to encapsulate the intricacies and fluidity of contemporary marketplaces. The variability in customer behavior, swift technical progress, and the globalization of supply chains require increasingly advanced and flexible strategies. Demand forecasting entails the projection of future consumer demand for a product or service. This forecast is derived from the analysis of historical sales data, prevailing market trends, and additional pertinent factors [27]. Precise demand forecasting is essential for enterprises to make educated decisions on production planning, inventory management, supply chain optimization, and resource allocation. By comprehending future demand, organizations may prevent stockouts, reduce surplus inventory, and guarantee the availability of the appropriate quantity of product at the optimal time [28]. This results in enhanced customer satisfaction, diminished expenses, and augmented profitability. Demand forecasting can be conducted using a variety of ways, from basic qualitative techniques reliant on human judgment to intricate quantitative models employing statistical algorithms and machine learning.

2.1 Traditional Demand Forecasting Methods

Historical demand forecasting has predominantly depended on statistical time series models, including ARIMA (AutoRegressive Integrated Moving Average), exponential smoothing, and seasonal decomposition methods [29]. These strategies are inherently linear, based on the premise that future demand can be forecasted by extrapolating historical patterns and trends. They have demonstrated adequate accuracy for stable demand patterns characterized by distinct seasonality and trends, although they encounter difficulties with non-linear relationships, numerous seasonality patterns, and the incorporation of exogenous factors [30]. A fundamental limitation of these models is their intrinsic retrospective viewpoint. They are highly adept at recognizing past occurrences but are inadequately prepared to predict future events, particularly when demand is affected by unprecedented circumstances or disruptive elements absent from previous data [31]. For example, they are unable to inherently integrate the effects of a new marketing campaign, a competitor's product introduction, abrupt supply chain disruptions, or changes in consumer mood derived from social media. The integration of these external variables necessitates manual, frequently ad hoc, modifications by forecast analysts, thereby increasing subjectivity and the possibility for inaccuracy [32].

Chen et al. [33] (2019) found that conventional forecasting approaches generally attain accuracy rates of 65-75% in steady market conditions but suffer considerable decline during volatile periods or structural market shifts, with error rates frequently rising by 30-50%. Moreover, these models necessitate considerable user involvement for upkeep, encompassing parameter adjustment and model re-selection when data attributes change [34]. The constraints of these methodologies, notably their data dependency, inflexibility, and incapacity to manage high-dimensional data, have become increasingly evident as global markets evolve to be more dynamic, linked, and diversified in data sources. The expanding performance disparity has generated an urgent demand for more agile, intelligent, and automated forecasting systems adept at managing contemporary market complexity.

2.2 Machine Learning in Demand Forecasting

The utilization of machine learning in demand forecasting commenced with basic methods, including linear regression, decision trees, and support vector machines. Huang and Zhang (2018) [35] demonstrated that these methods could enhance forecasting accuracy by 8%–12% compared to conventional statistical techniques by more effectively managing nonlinear connections and various input variables. Advanced ensemble approaches, such as Random Forest and Gradient Boosting Machines, have exhibited enhanced efficacy in managing intricate interactions among demand factors. Petropoulos et al. (2020) demonstrated that ensemble approaches can enhance forecasting accuracy by 15-20% and offer superior uncertainty quantification [36]. Although these technologies marked a substantial advancement, they remained limited when dealing with highly granular, high-dimensional data. Their performance frequently stagnated when utilized for forecasting at the individual SKU-store level across extensive product portfolios or when modeling complex temporal connections in lengthy time series. This constraint became especially apparent with the emergence of e-commerce, where demand signals are shaped by an intricate network of digital variables like click-through rates, online reviews, and micro-trends [37]. The necessity to model intricate, non-linear patterns at scale facilitated the adoption of deep learning architectures, which are particularly adept at autonomously extracting pertinent features from extensive, unstructured datasets and identifying highly complex relationships that shallower models may overlook. Machine learning methodologies provide improved adaptability and precision in forecasting demand for e-commerce. Tugay et al. propose an innovative methodology that accounts for market dynamics, including various sellers presenting identical products at varying prices [38]. They utilize multiple regression algorithms and a stacking generalization (ensemble learning) method, demonstrating that the latter produces superior outcomes.

2.3 Techniques for Forecasting Demand

Demand prediction within the realm of e-commerce constitutes a relatively recent field of study. Conventional Statistical Techniques. Initial endeavors in demand forecasting for online advertising primarily utilized conventional statistical techniques. Levis et al. propose a systematic optimization-based methodology for forecasting consumer demand with support vector regression (SVR) [39]. Jain et al. investigate Support Vector Regression (SVR) for demand forecasting, employing a comparable three-step technique that incorporates nonlinear and linear programming, along with a recursive component to adjust to past sales data for precise forecasts [40]. These conventional statistical models have been extensively utilized owing to their simplicity and interoperability. Deep learning models, especially neural networks, have remarkable proficiency in analyzing extensive datasets and recognizing intricate patterns. Bandara et al. employ Long Short-Term Memory (LSTM) networks to leverage nonlinear demand interactions inside an e-commerce product hierarchy [41]. Zhang et al. consolidate research on artificial neural networks (ANNs) in predictive applications outside online advertising, offering insights into modeling challenges and prospective avenues for exploration [42]. Kuo et al. contrast neural networks with conventional models, emphasizing their superiority in managing non-linear connections [43]. Azzouni et al. provide an LSTM framework for predicting network traffic, showcasing its efficacy on empirical data [44].

2.4 Deep Learning Applications

The emergence of deep learning systems has represented a substantial enhancement in demand forecasting abilities. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), have demonstrated remarkable efficacy in capturing temporal dependencies and long-term patterns in sequential demand data, adeptly modeling intricate seasonality and trends [45]. Smyl et al. (2021) revealed that LSTM-based models could exceed traditional approaches by 20-25% in intricate forecasting situations characterized by several seasonal patterns and external variables. Convolutional Neural Networks (CNNs) have been effectively utilized for demand forecasting, especially in contexts including spatial-temporal data patterns, such as detecting demand correlations across several locations in a retail network [46]. Recently, transformer-based models, acclaimed for their application in natural language processing, have been modified for time series forecasting. These models employ self-attention mechanisms to assess the significance of various time stages in a sequence, providing a robust alternative for capturing intricate, long-range relationships in demand data that LSTMs may overlook, particularly across an extensive portfolio of related products [47]. The capacity of these deep learning models to autonomously acquire hierarchical features from unprocessed data without significant manual feature engineering signifies a transformative change towards more adaptable and precise forecasting systems.

2.5 Transformer Models and Attention Mechanisms

Recent advancements in transformer architectures have created novel opportunities for demand forecasting [48]. The attention method enables models to dynamically concentrate on pertinent historical eras and external circumstances, enhancing both accuracy and interpretability by elucidating which prior data points significantly impacted the future. Li et al. (2022) [49] demonstrated that transformer-based models may attain state-of-the-art performance in multi-horizon forecasting tasks, notably excelling at capturing intricate global relationships over extensive time series that pose challenges for LSTMs. In contrast to recurrent networks that handle data sequentially, transformers process complete sequences simultaneously, facilitating more efficient training on large datasets [50]. This architecture is exceptionally adept in unified forecasting for numerous interrelated items (e.g., a comprehensive retail catalog), as it can proficiently discern shared patterns and relationships throughout the entire portfolio. Moreover, their capacity to manage variable-length inputs and integrate various non-temporal data (such as promotional calendars or textual information from news events) establishes them as a versatile foundation for developing the next generation of comprehensive demand-sensing platforms [51].

2.6 Case Studies: Walmart, Amazon, and Zara

Numerous case studies highlight the revolutionary effects of AI on inventory management and demand forecasting. Walmart has adopted AI-driven demand forecasting to optimize supply chain efficiency, decrease inventory costs, and improve product availability [52]. Amazon's AI-driven inventory management system accurately forecasts customer demand, allowing the corporation to sustain optimal stock levels and minimize delivery delays [53]. Zara employs AI to scrutinize sales data and client preferences, enabling real-time inventory adjustments and the reduction of unsold stock [54]. These examples underscore the significant advantages that AI may offer in inventory management and demand forecasting. Nonetheless, they also exemplify the difficulties associated with the implementation of these technologies and the necessity of resolving concerns around data quality, integration, and ethics.

3 Methodology

This work utilizes a mixed-methods approach that integrates a thorough literature review, a comparative analysis of AI forecasting methodologies, and an empirical evaluation with real-world datasets. The research methodology is designed to offer theoretical insights and practical validation of AI-driven forecasting methods.

3.1 Real-World Datasets

The Walmart Store Sales dataset [56]is an independent, historical dataset frequently utilized for sales forecasting and machine learning training. The Walmart dataset contains historical sales data from 45 Walmart shops, each encompassing many departments. This dataset aims to forecast sales for each department with historical markdown data from the Walmart dataset, which encompasses information from 45 Walmart shops. In contrast to M5's emphasis on individual product forecasting for 10 stores in 3 states, the Walmart Store Sales dataset prioritizes department-level sales predictions across a significantly larger network of stores, generally encompassing multiple years of weekly sales data, incorporating variables such as temperature, fuel prices, unemployment rates, and holiday indicators. The M5 Walmart dataset [55] exemplifies a significant retail forecasting challenge, originating from the fifth Makridakis forecasting competition. The dataset comprises unit sales of 3,049 products, categorized into 3 product categories (Hobbies, Foods, and Household) and 7 product departments, distributed over ten locations in three states (CA, TX, and WI). The dataset comprises 30,490 time series of daily sales data over a period of about 6 years from the M5 Competition Dataset, establishing a hierarchical framework that facilitates forecasting across various aggregation levels. M5 offers 30,490 distinct time series representing individual product sales at specific Walmart shops.

3.2 Data preparation and feature engineering

The collection and preparation of data are essential for the development of resilient AI models. The quality of the input data significantly influences the precision and dependability of AI-driven inventory management and demand forecasting. This is a crucial phase in which raw data is converted into a format appropriate for deep learning models, and pertinent features are generated to enhance forecasting precision. We submitted an application. This study derives features such as 'day_of_week,' 'month,' and 'day_of_year' from the 'date' column. It further generates a binary 'is_weekend' attribute. These attributes enable the models to discern daily, weekly, and annual seasonal patterns. Simulated external variables such as 'weather_index,' 'economic_indicator,' 'social_media_mentions,' and 'competitor_price' are used in the synthetic data production. In a practical context with real datasets, this would constitute external data streams incorporated into the dataset. These factors furnish the models with context on events or conditions that may affect demand. This research employs one-hot encoding for categorical variables such as 'color_preference,' 'size_preference,' and 'region.' This transforms categorical information into a numerical representation suitable for deep learning models. The StandardScaler is employed for features, whereas the MinMaxScaler is utilized for the goal variable ('demand') within the prepare_sequences method. Scaling is crucial for deep learning models, as it facilitates expedited convergence during training and mitigates the risk of features with bigger values overshadowing the learning process. The prepare_sequences method is explicitly intended to generate sequential data for the LSTM and CNN-LSTM models. It produces input sequences (X) of a specified sequence_length (e.g., 30 days of features) and matching target sequences (y) of a defined forecast_horizon (e.g., the subsequent 7 days of demand). This framework enables the models to derive insights from historical patterns to forecast future values.

3.3 Data Cleaning and Normalization

Data cleaning is eliminating inaccuracies, including absent values and duplicates, to preserve data integrity. Standardizing data to a single scale ensured consistency across diverse inputs. This process was essential for enhancing the efficacy of machine learning algorithms. Feature engineering generated additional attributes to augment the model's predictive capability, including the integration of sales data with promotional events and the development of lag features to account for temporal dependencies. To improve model efficiency and reduce statistical instabilities, we applied standardization to scale the input to the range [0–1].

3.4 Model selection and building

This phase entails selecting and configuring the deep learning architectures designated for forecasting. The system employs two categories of models.

3.5 LSTM Model

This study employed Long Short-Term Memory networks, a sophisticated advancement of classic recurrent neural networks, particularly engineered to mitigate the vanishing gradient problem that affected earlier RNN architectures. The LSTM cell comprises three essential gates: the forget gate, the input gate, and the output gate, each fulfilling a distinct role in regulating information flow. The forget gate ascertains which information from the preceding cell state should be eliminated, employing a sigmoid activation function to produce values ranging from 0 to 1, where 0 signifies "entirely forget" and 1 indicates "entirely retain." The input gate collaborates with a Tanh layer to ascertain whether new information should be retained in the cell state, thus regulating the influx of new data into the memory.

Table 1: Walmart Store Sales Dataset Statistics

Attribute	**Details**
Total Stores	45 Walmart stores
Aggregation Level	Store-Department level
Time Granularity	Weekly sales data
Time Period	∼3 years (2010-2012)
Total Records	∼6,435 records (45 stores × ∼143 weeks)
Departments per Store	Variable (typically 80+ departments)
Target Variable	Weekly_Sales (department-level)
Store Types	3 types (A, B, C based on size)
Data Files	3 main CSV files
Training Data	train.csv
Store Information	stores.csv
Feature Data	features.csv
External Factors	Temperature, Fuel_Price, CPI, Unemployment
Holiday Indicators	Super Bowl, Labor Day, Thanksgiving, Christmas
Markdown Events	MarkDown1-5 (promotional price reductions)
Store Size Range	34,875 - 219,622 sq ft
Challenge Objective	Department sales forecasting
Data Type	Weekly aggregated sales (continuous)
Forecast Horizon	Variable (typically next few weeks)
Geographic Coverage	Multiple states (unspecified)
Dataset Focus	Regression/Time series forecasting

Table 2: M5 Walmart Dataset Statistics

Attribute	**Details**
Total Time Series	30,490 individual series
Products	3,049 unique products
Stores	10 Walmart stores
States	3 (California, Texas, Wisconsin)
Product Categories	3 (Hobbies, Foods, Household)
Product Departments	7 departments
Time Period	∼6 years (2011-2016)
Total Days	1,913-1,969 daily observations
Hierarchical Levels	12 aggregation levels
Level 12 (Finest)	30,490 series (item-store level)
Level 11	9,147 series (item level)
Level 10	1,049 series (store-category level)
Data Files	4 main CSV files
Sales Training Data	sales_train_validation.csv
Price Data	sell_prices.csv
Calendar Data	calendar.csv (events, holidays, SNAP)
Evaluation Data	sales_train_evaluation.csv
Forecast Horizon	28 days (4 weeks)
Data Type	Daily unit sales (integers)
Challenge Objective	Hierarchical time series forecasting
External Factors	Prices, promotions, calendar events, SNAP benefits

The output gate regulates which components of the cell state are emitted as the hidden state, integrating the revised cell state using a sigmoid-controlled filter. This gating mechanism enables LSTMs to choose to retain significant information over extended sequences while discarding extraneous details, rendering them especially useful for demand forecasting, where both long-term seasonal trends and short-term variations impact future sales.

3.6 Bidirectional LSTM Processing

We employ Bidirectional LSTM architecture, which improves upon classic LSTM by simultaneously processing sequences in both forward and backward directions, thus equipping the model with comprehensive contextual information from both past and future time steps within the training sequence. In demand forecasting applications, bidirectional processing is essential as seasonal patterns, promotional events, and market trends frequently display dependencies that span both past and future timeframes. The forward LSTM analyzes the sequence from time step t=1 to t=T, elucidating the impact of past events on future demand, whereas the backward LSTM examines the sequence from t=T to t=1, discerning how anticipated events or seasonal trends may affect the interpretation of current demand. The ultimate hidden state representation amalgamates both directional outputs, usually via concatenation or element-wise addition, resulting in a more comprehensive feature representation that integrates temporal context from both directions. This methodology is especially advantageous for retail demand forecasting, as holiday seasons, back-to-school intervals, and scheduled promotional campaigns establish temporal dependencies that affect demand patterns weeks or months ahead, necessitating the model's comprehension of both historical trends and forthcoming market conditions.

Implementing dropout regularization in LSTM networks necessitates meticulous execution because of the recurrent characteristics of these designs, as indiscriminate application of dropout between time steps might interrupt the temporal information flow and impair model efficacy. The most efficacious method entails the selective application of dropout: input dropout influences the connections between input features and LSTM gates, recurrent dropout pertains to the hidden state connections across time steps, and output dropout is implemented on the final LSTM layer outputs prior to their integration into dense layers. Contemporary implementations frequently utilize variational dropouts, which preserve the same dropout mask over all time steps for a specific sequence, ensuring a consistent information flow while still offering regularization advantages. In demand forecasting, dropout rates often vary from 0.2 to 0.4, with elevated rates employed for smaller datasets or in instances of overfitting. The regularization effect enhances the model's ability to generalize to unfamiliar demand patterns by mitigating excessive dependence on particular temporal features, thereby ensuring consistent performance in the face of novel market conditions, seasonal fluctuations, or unforeseen events absent from the training data.

3.7 CNN-LSTM Hybrid Model

In this study, we integrate CNN and LSTM for demand forecasting, using CNN to extract significant features that we then input into LSTM for pattern recognition. In CNN-LSTM hybrid architectures, convolutional layers function as advanced feature extractors that detect local temporal patterns in time series data, differing fundamentally from their use in image processing. In demand forecasting, one-dimensional convolutional filters traverse the temporal dimension of input sequences, identifying repeating patterns such as weekly seasonality, promotional surges, or gradual trend alterations that manifest over designated time intervals. Each convolutional filter acquires the ability to identify a certain temporal pattern via its learned weights, while several filters operate concurrently to capture other pattern types simultaneously. The convolution process calculates dot products between filter weights and local temporal frames, subsequently applying nonlinear activation functions that enhance pattern recognition abilities. Pooling operations, such as max pooling or average pooling, diminish temporal dimensionality while preserving essential pattern information, resulting in translation-invariant feature representations. This hierarchical feature extraction process enables the network to construct progressively intricate pattern representations in deeper layers, commencing with basic local variations in initial layers and advancing to complex seasonal or cyclical patterns in subsequent layers, ultimately yielding rich temporal features that augment the LSTM's sequential modeling capabilities.

The LSTM component of CNN-LSTM hybrid architectures functions on the feature representations derived from convolutional layers, emphasizing the modeling of long-term sequential dependencies and temporal relationships that surpass the local patterns identified by convolutions. Subsequent to the extraction and pooling of temporal features by the CNN layers, the resultant feature maps are restructured and input into LSTM layers, which analyze these abstracted temporal sequences to comprehend the evolution and interrelation of the derived patterns across time. The memory mechanisms of LSTM are especially effective when utilizing features retrieved from CNN, as they engage with higher-level pattern representations instead of raw input values, facilitating more efficient learning of intricate temporal correlations. Stacking multiple LSTM layers facilitates hierarchical temporal modeling, wherein lower levels concentrate on short-term connections among extracted patterns, while upper layers include long-term strategic links and seasonal cycles. The concealed states of LSTM cells retain information regarding pattern development over prolonged durations, whereas the cell states safeguard essential long-term memory concerning seasonal trends, market cycles, and slowly evolving consumer behaviors that affect demand forecasting precision.

3.8 Training Strategies and Optimization Techniques

Sliding window data preparation is a crucial preprocessing method for time series forecasting that converts historical sequential data into a supervised learning format appropriate for neural network training. This method generates many input-output pairings by methodically sliding a fixed-size window across the time series, with each window comprising a sequence of historical observations as input features, combined with subsequent values as prediction targets. The window size, or lookback time, often spans from 30 to 120 days for retail demand forecasting, striking a balance between adequate historical knowledge and the challenges of computing complexity and overfitting risks. The stride parameter dictates the movement of the window between successive samples, with a stride of 1 resulting in maximal overlap and optimal data usage, whereas greater strides diminish the training data volume and may overlook significant transitional patterns. Advanced solutions utilize numerous window widths concurrently, generating multi-scale temporal characteristics that encapsulate both short-term variations and long-term trends. The technique addresses seasonality by ensuring that training data encompass entire seasonal cycles, preventing models from acquiring incomplete seasonal patterns that may result in inadequate generalization across varying seasonal periods.

Adam optimization integrates the advantages of adjustable learning rates with momentum-based gradient descent, rendering it especially efficient for training intricate neural architectures such as CNN-LSTM models using time series data. The approach employs distinct adaptive learning rates for each parameter, calculated using exponentially decaying averages of historical gradients and squared gradients, enabling it to effectively manage sparse gradients and noisy optimization landscapes typical in temporal data. The momentum component facilitates convergence in consistent gradient directions while mitigating oscillations in varying directions, which is especially beneficial when addressing seasonal data patterns that generate intricate loss landscapes. Learning rate scheduling improves Adam's efficacy by methodically modifying the base learning rate during training: warm-up schedules incrementally raise learning rates in the initial epochs to avert premature convergence to suboptimal local minima, cosine annealing establishes cyclical learning rate patterns that facilitate evasion of local optima, and step decay diminishes learning rates at specified intervals to refine convergence. In demand forecasting applications, learning rates generally commence at 0.001-0.01, with decay factors of 0.1-0.5 implemented after validation loss stabilizes, thereby facilitating ongoing model learning while preventing overshooting of optimal parameter values in subsequent training stages.

Early stopping functions as a regularization method and a means of enhancing computational efficiency, overseeing validation performance to terminate training when the model starts to overfit the training data, while preserving optimal generalization to novel patterns. The implementation monitors validation loss across successive epochs, employing patience parameters that dictate the allowable number of epochs without improvement before cessation, generally varying from 10 to 50 epochs based on model complexity and dataset dimensions. Time series cross-validation necessitates specialized methodologies that honor temporal sequencing, employing techniques such as walk-forward validation, in which the model is trained on historical data and validated on subsequent periods, or expanding window validation, where training sets incrementally increase while preserving the integrity of the temporal sequence. The validation method must consider seasonal cycles by ensuring that validation periods encompass representative seasonal patterns, and avoiding overly positive or pessimistic performance estimations that may result from verifying solely during particular seasonal intervals. Advanced implementations utilize multiple validation metrics concurrently, assessing both statistical accuracy indicators such as RMSE and MAE, as well as business-relevant metrics like inventory optimization performance or profit impact, thereby establishing comprehensive stopping criteria that reconcile model performance with practical deployment factors.

Gradient clipping mitigates the exploding gradient issue prevalent in deep recurrent networks, where gradients may increase exponentially during backpropagation across time, resulting in numerical instability and training divergence. The method observes gradient magnitudes during backpropagation and enforces clipping when gradients surpass established thresholds, commonly executed via gradient norm clipping, which proportionally scales down the entire gradient vector when its L2 norm exceeds the threshold, or value clipping, which restricts individual gradient components to maximum absolute values. In CNN-LSTM designs for demand forecasting, gradient clipping levels generally vary from 0.5 to 5.0, where lower values offer enhanced regularization but may result in slower convergence. Batch normalization enhances gradient clipping by standardizing layer inputs to provide uniform activation distributions throughout training, mitigating internal covariate shift and allowing for elevated learning rates without numerical instability. The integration of these techniques is crucial when training on-demand data characterized by extreme values, abrupt spikes, or extended sequences, as gradient accumulation over numerous time steps can pose numerical difficulties that conventional optimization algorithms are challenging to manage efficiently.

Multi-task learning in demand forecasting transcends single-point prediction by integrating auxiliary tasks that offer supplementary learning signals and enhance model robustness via shared representation learning. Auxiliary tasks encompass demand uncertainty quantification, wherein the model concurrently forecasts point estimates and confidence intervals; trend classification, which categorizes demand patterns as increasing, decreasing, or stable; and seasonality decomposition, which distinctly models seasonal, trend, and residual components. The auxiliary objectives generate supplementary gradient signals that direct the common CNN and LSTM layers to acquire more generalizable representations, as the model is required to extract features beneficial for several interconnected tasks instead of overfitting to a singular prediction purpose. The implementation generally comprises shared backbone networks with task-specific heads, wherein CNN-LSTM features are directed into multiple output branches, each utilizing suitable loss functions: regression losses for point predictions, classification losses for categorical auxiliary tasks, and specialized losses such as quantile loss for uncertainty estimation. The overall loss integrates all objectives through weighted summation, with weights modified according to task significance and convergence rates, typically commencing with uniform weights and subsequently adjusting based on validation performance to guarantee equitable learning across all objectives while preserving primary forecasting precision.

4 Results and Discussion

Table 3 delineates the exhaustive set of parameters used to train an LSTM (Long Short-Term Memory) neural network model. The design has two LSTM layers with 128-256 and 64-128 hidden units, respectively, succeeded by a dense layer of 50-100 neurons and a solitary linear output neuron. The model analyzes sequential data spanning 30 to 90 days and incorporates 5 to 15 input characteristics, employing dropout regularization (0.2-0.3 for LSTM layers, 0.2 for dense layers) and L2 regularization (0.001-0.01) to mitigate overfitting. Training employs the Adam optimizer with an initial learning rate ranging from 0.001 to 0.01, beta values set between 0.9 and 0.999, and utilizes batch sizes of 32 to 128 samples across 100 to 200 epochs. The configuration incorporates early stopping methods that observe validation loss with a patience of 10-20 epochs and a minimum delta of 0.001, in addition to learning rate reduction on plateau (factor 0.5, patience 5-10 epochs). The dataset is divided

Table 3: LSTM Only Training Configuration Parameters

Parameter Category	Parameter	Value/Range
Architecture	LSTM Hidden Units (Layer 1)	128-256
	LSTM Hidden Units (Layer 2)	64-128
	Dense Hidden Layer	50-100 neurons
	Output Layer	1 neuron (linear)
Input Configuration	Sequence Length	30-90 days
	Input Features	5-15 features
	Input Shape	(batch_size, seq_len, n_features)
Regularization	Dropout Rate (LSTM)	0.2-0.3
	Dropout Rate (Dense)	0.2
	L2 Regularization	0.001-0.01
	Batch Normalization	Applied between layers
Optimization	Optimizer	Adam
	Learning Rate (Initial)	0.001-0.01
	Beta 1	0.9
	Beta 2	0.999
	Epsilon	1e-8
Training Setup	Batch Size	32-128
	Epochs	100-200
	Loss Function	MSE/MAE
	Metrics	RMSE, MAE, MAPE

into 70% for training, 15% for validation, and 15% for testing, with gradient clipping set at 1.0 to ensure training stability. The model employs MSE and MAE as loss functions, with RMSE, MAE, and MAPE as evaluation metrics. Table 4 details the configuration parameters for a sophisticated CNN-LSTM hybrid neural network that combines convolutional and recurrent architectures for advanced sequence modeling. The CNN component features two convolutional layers with 64 and 128 filters respectively, using 3x5 and 3x5 kernel sizes, followed by 2x2 max pooling and ReLU activation with 0.25 dropout and same padding. The LSTM architecture consists of two layers with 100-200 and 50-100 hidden units, 0.3 dropout, and the option to return sequences, followed by a 50-neuron dense layer with 0.2 dropout and a single linear output neuron. The model processes sequences of 60-120 days with 8-20 input features and employs a sophisticated two-stage training approach: Stage 1 focuses on CNN pre-training for 50-100 epochs, while Stage 2 performs end-to-end fine-tuning for 100-200 epochs. Advanced techniques include cosine annealing learning rate scheduling, gradient accumulation over 2-4 steps, mixed precision training, label smoothing (0.1), spatial dropout (0.1-0.2), and Gaussian noise injection (0.01-0.05). The training uses Adam optimization with warm-up epochs (10-20), Huber/MSE loss functions, comprehensive regularization including L2 regularization and batch normalization after convolutional and LSTM layers, early stopping with validation loss monitoring, and maintains the standard 70-15-15% data split with gradient control mechanisms to ensure stable training. Table 5 illustrates that the CNN-LSTM hybrid model substantially surpasses the independent LSTM model in all assessment metrics. The CNN-LSTM attains enhanced predictive accuracy with a Mean Absolute Error (MAE) of 1.47, in contrast to the LSTM's 4.67, indicating a significant 2.62% enhancement. Correspondingly, the Mean Squared Error (MSE) significantly declines from 7.6 to 2.3 (a 5.3% enhancement), while the Root Mean Squared Error (RMSE) reduces from 8.91 to 3.68 (a 5.26% enhancement). The coefficient of determination (R^2) demonstrates significant augmentation from 0.76 to 0.89, signifying that the CNN-LSTM model accounts for 89% of the variance in the data, in contrast to 76% for the LSTM alone, reflecting a 0.13% improvement in model fit. Both models attain identical accuracy enhancements of 35.17%, indicating equivalent performance in classification accuracy; however, the CNN-LSTM exhibits superior performance in regression metrics, highlighting the advantages of integrating convolutional layers to capture spatial-temporal patterns before sequen-

Table 4: CNN-LSTM Combined Training Configuration Parameters

Parameter Category	Parameter	Value/Range
CNN Architecture	Conv1D Filters (Layer 1)	64
	Conv1D Kernel Size (Layer 1)	3-5
	Conv1D Filters (Layer 2)	128
	Conv1D Kernel Size (Layer 2)	3-5
	MaxPooling Size	2
	CNN Activation	ReLU
	CNN Dropout	0.25
	Padding	Same
LSTM Architecture	LSTM Hidden Units (Layer 1)	100-200
	LSTM Hidden Units (Layer 2)	50-100
	LSTM Dropout	0.3
	Return Sequences	True/False
Input Configuration	Sequence Length	60-120 days
	Input Features	8-20 features
	Input Shape	(batch_size, seq_len, n_features)
Early Stopping	Monitor	Validation Loss
	Patience	15-25 epochs
	Min Delta	0.0001
	Restore Best Weights	True
Regularization	Spatial Dropout	0.1-0.2
	Gaussian Noise	0.01-0.05
	L2 Regularization	0.0001-0.001
	Batch Normalization	After Conv/LSTM layers

Table 5: Performance Comparison of LSTM and CNN-LSTM Models

Metric	LSTM Model	CNN-LSTM Model	Improvement
MAE	4.47	1.85	2.62%
MSE	7.6	2.3	5.3%
RMSE	8.94	3.68	5.26%
R^2	0.76	0.89	0.13%
Accuracy Improvement	**20.6%**	**35.17%**	**14.57%**

tial processing via the LSTM layers. Figure 1 illustrates a time series graph of Demand spanning from early 2031 to mid-2033. The data reveal a distinct seasonal trend, with demand consistently reaching its zenith during a specific timeframe each year and declining to its nadir at other intervals. An observable upward trend in peak demand values over the three years indicates a progressive increase in total demand over time. Figure 2 illustrates the monthly average demand, offering a distinct analysis of the seasonal trends over the year. The data indicates that demand peaks throughout the spring months, particularly in March, April, and May, with the highest levels observed in March and April. Demand subsequently declines consistently throughout the summer and autumn, attaining its nadir in September and October. Subsequent to this decline, demand commences to increase once more in November and December, signifying the initiation of a new seasonal cycle. The graphic accurately illustrates the annual seasonality of the demand data. Figure 3 juxtaposes the training and validation performance of two deep learning models, an LSTM and a CNN-LSTM, throughout 23 epochs. The y-axis, denoting Loss (MSE), is presented on a logarithmic scale, signifying that the loss values diminish exponentially. The training loss for both models (solid lines) exhibits a consistent decrease, indicating that both are assimilating knowledge from the data. The validation loss (dashed lines) indicates a significant disparity: the LSTM's validation loss is inconsistent and typically exceeds its training loss, implying overfitting to the training data. The validation loss of the CNN-LSTM model

Figure 1: Historical demand pattern.

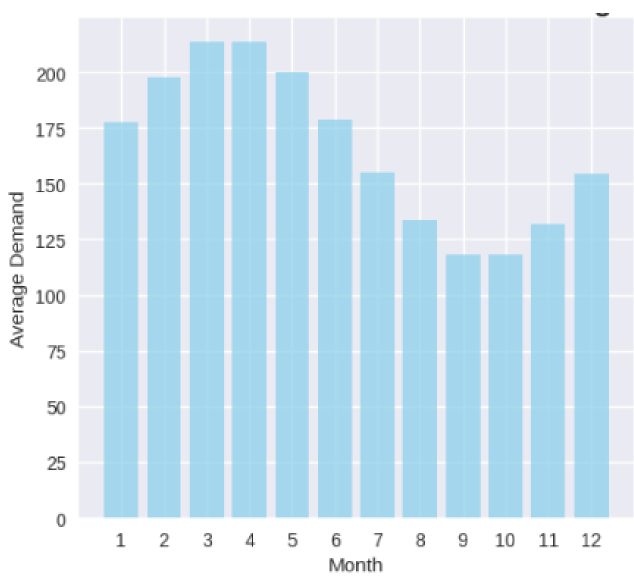

Figure 2: Monthly Demand Patterns.

initially declines before exhibiting some instability after approximately five epochs. The CNN-LSTM model demonstrates greater stability and potential for superior performance in this task, evidenced by its continuously lower validation loss compared to the LSTM model.

Figure 4 illustrates that the scatter plot, referred to as a Predicted vs. Actual plot, evaluates a model's efficacy. The x-axis denotes the actual demand values, while the y-axis illustrates the model's expected demand values. The dashed black line denotes an ideal model in which predictions precisely align with actual values (y=x). The dispersed orange dots signify distinct data points. The proximity of the dots to the dashed line indicates superior model performance. The plot illustrates a robust positive correlation between the projected and actual values, as the points cluster closely around the ideal prediction line. This signifies that the model is effectively forecasting the actual demand. Nonetheless, there is a degree of dispersion across the data points, especially at elevated demand values, indicating that the model is imperfect and exhibits certain predictive inaccuracies.

Figure 5 depicts the relationship between demand and four principal external factors that affect market dynamics. Marketing exhibits the most substantial positive association with demand at roughly 0.22, signifying that marketing initiatives exert the most influence on stimulating consumer

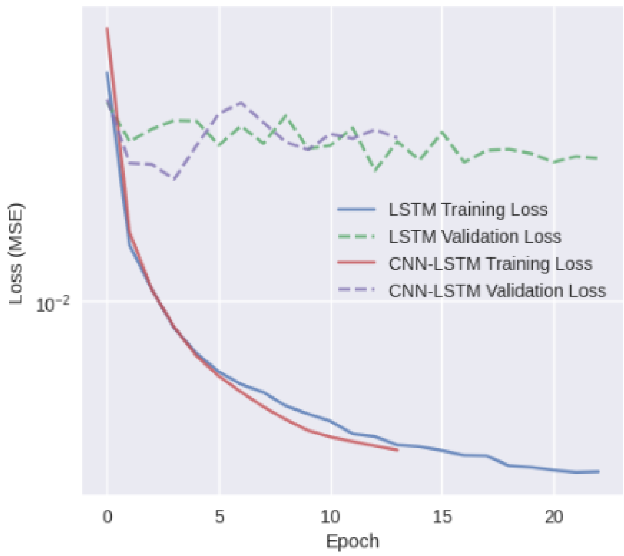

Figure 3: Model Training History.

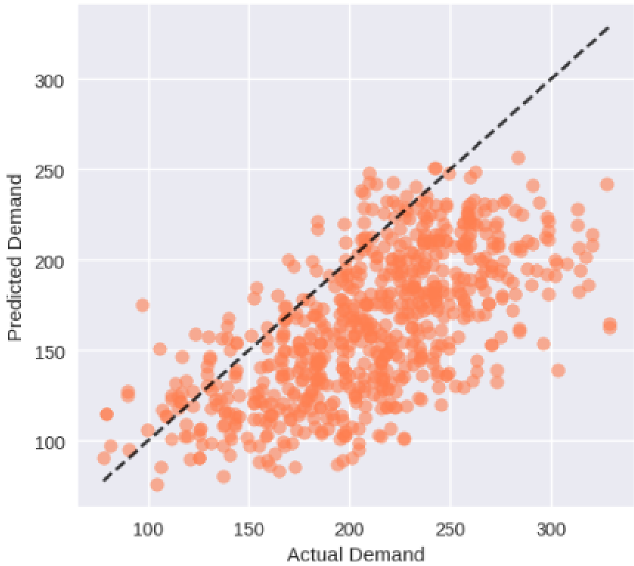

Figure 4: Actual Vs Predicted Demand.

demand. Economic factors exhibit a moderate positive correlation of approximately 0.09, indicating that overarching economic conditions also influence demand swings, albeit to a smaller degree than marketing. Weather circumstances demonstrate a minimal positive correlation of approximately 0.08, indicating a slight impact on demand patterns, potentially attributable to seasonal fluctuations or weather-influenced consumer behavior. Social media exhibits the lowest correlation at approximately 0.05, which may appear counterintuitive considering the significance of digital marketing; however, the results suggest that social media's influence on demand is either more intricate, indirect, or potentially eclipsed by traditional marketing channels. The graphic indicates that marketing is the principal driver of demand, whereas economic, weather, and social media elements serve as supportive but less significant influences on consumer demand patterns.

Figure 6 presents a horizontal bar chart that ranks various categorical combinations according to their corresponding demand. The x-axis denotes the Demand value, while the y-axis illustrates various categories, presumably encompassing qualities such as color, size, and region. The chart is organized in descending order of demand, facilitating the identification of combinations with the highest and lowest demand. The highest-performing combination is "South" in black and medium sizes, exhibiting a demand above 150. A massive green item from the "North" region has the lowest demand, slightly

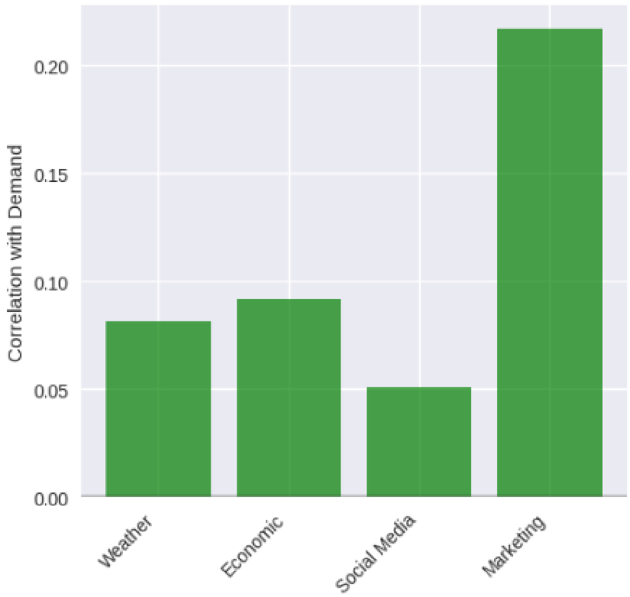

Figure 5: External Factors Impact on Demand.

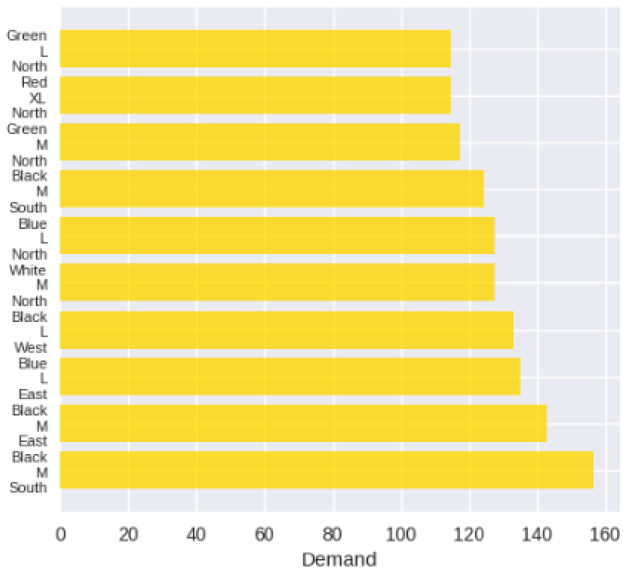

Figure 6: Top 10 Products Variants by Demand.

exceeding 100. This graphic effectively facilitates the rapid identification of both high-performing and low-performing categories.

Figure 7 illustrates the mean daily demand over the week, spanning from Monday to Sunday. The x-axis represents the days of the week, while the y-axis indicates the average demand. The figure demonstrates a distinct weekly pattern, with demand commencing at a modest level on Monday, reaching a zenith mid-week on Wednesday at over 190 units, and subsequently diminishing through Friday and Saturday to a nadir of just under 150 units on Saturday. Demand experiences a slight boost on Sunday, indicating a modest rise at the beginning of the new week. This image clearly illustrates the weekly seasonality and customer behavior in the demand data.

Figure 8 delineates the principal advantages and enhancements realized by the forecasting model. The y-axis denotes Improvement (%), with each bar illustrating a distinct business outcome. The results indicate substantial improvements universally. Customer Satisfaction experienced the most significant enhancement at 35%, followed by Inventory Optimization at 30%, suggesting that the model is effectively managing stock levels and addressing customer requirements. Forecast accuracy increased by 25%, attributable to a more precise model. Ultimately, the model facilitated a 20% decrease in costs,

Figure 7: Weekly Demand Patterns.

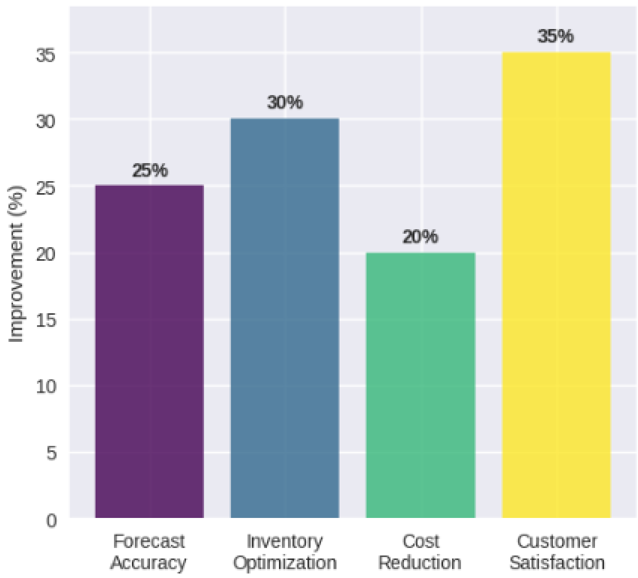

Figure 8: Supply Chain Integration Benefits.

attributable to enhanced planning and resource management. The graphic illustrates the measurable beneficial effect of the forecasting solution on multiple business indicators.

Figure 9 depicts the allocation of principal advantages obtained from an effective demand forecasting method. Demand forecasting is an essential business activity that utilizes previous sales data, market trends, and many external factors to anticipate future consumer demand. This approach enables firms to make more informed decisions regarding inventory management, production planning, and resource allocation. The data indicates that Reduced Stockouts is the most significant portion of the benefits at 30%, underscoring the model's efficacy in averting lost revenues resulting from inadequate inventory. Improved client retention is 26.7%, reflecting the constant fulfillment of client needs. Reduced inventory costs account for 23.3% of the advantages, whereas operational efficiency contributes 20%, illustrating how precise forecasting enhances company processes. The figure offers a clear and proportional representation of the diverse benefits of adopting a comprehensive demand forecasting strategy.

Figure 10 illustrates the efficacy of a predictive model (depicted in red) in comparison to real demand data (shown in blue) across around 1,400 time steps. The chart indicates that the forecast-

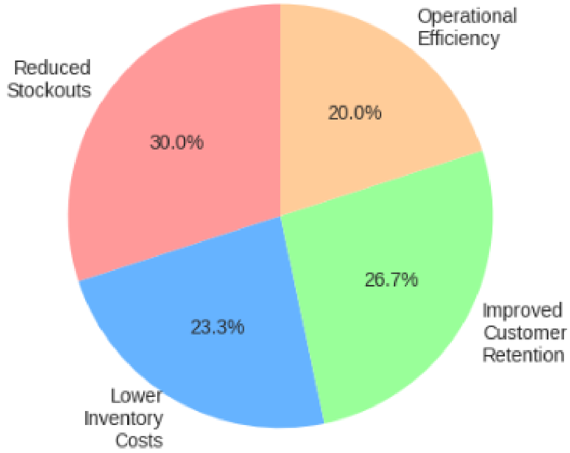

Figure 9: Business Impact Distribution.

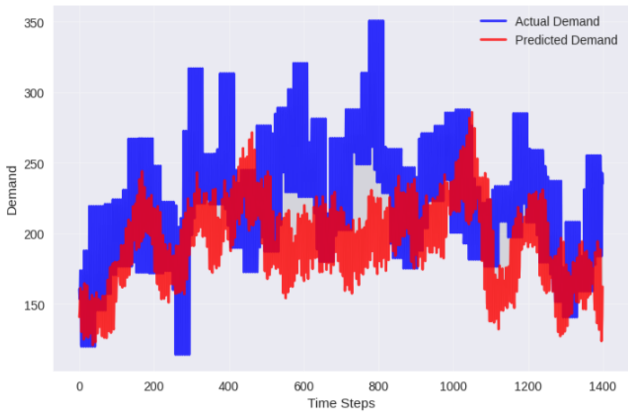

Figure 10: Detailed Demand Prediction Comparison.

ing model effectively captures the overarching trends and patterns in demand, notably in recognizing important peaks and troughs, exemplified by the pronounced spike at time step 800, where real demand approximates 350 units. The model exhibits limitations in precision, as indicated by instances where the anticipated demand (red line) either underestimates or overestimates the actual values, especially during periods of high volatility characterized by fast demand fluctuations. The forecasting demonstrates optimal accuracy during stable periods characterized by incremental changes; nevertheless, it has difficulties with abrupt fluctuations in demand, a common challenge in demand forecasting attributed to the inherent unpredictability of market dynamics, consumer behavior, and external influences. Although the model demonstrates adequate accuracy in monitoring overall demand trends and seasonal variations, there exist opportunities for enhancement in capturing short-term volatility and extreme demand occurrences. We can achieve these improvements through refined feature engineering, model optimization, or by integrating supplementary external variables that influence demand fluctuations. Figure 11 illustrates regional demand analysis, highlighting substantial geographical disparities in average demand across four regions, offering essential insights for strategic planning and resource distribution. The North area exhibits the highest average demand at 101.6 units, establishing it as the foremost market with the most robust customer base or economic activity. The South region exhibits the lowest average demand at 85.8 units, indicating a 15.5% disparity from the North, which may imply lower population density, diminished economic activity, or divergent consumer preferences in this region. The East and West regions demonstrate roughly equivalent demand patterns at 92.9 units apiece, signifying balanced market conditions and analogous consumer behavior across these areas. This demand distribution indicates that businesses ought to prioritize resource allocation and inventory management in the North region while examining the factors that contribute to diminished

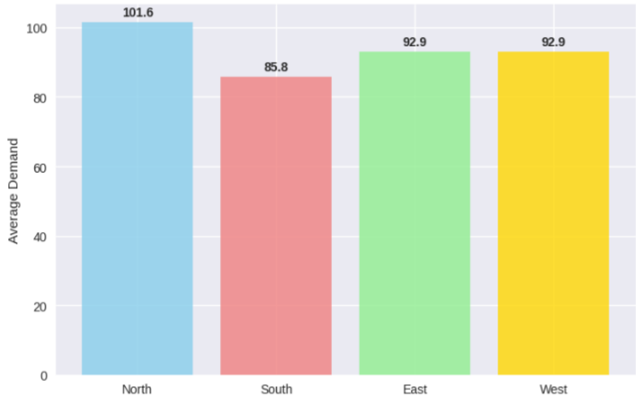

Figure 11: Regional Demand Analysis.

demand in the South, which may encompass demographic disparities, variations in the competitive landscape, or regional economic conditions.

The similarity between the East and West regions suggests that these markets may react comparably to marketing strategies and product offerings, permitting standardized approaches. Conversely, the North-South divergence necessitates region-specific strategies to either leverage the robust Northern market or implement targeted initiatives to foster growth in the underperforming Southern region. Figure 12 illustrates the feature importance analysis, demonstrating the hierarchical impact of different factors on the AI model's demand prediction capabilities. Historical Demand is identified as the primary predictor, with an importance score of 0.35, representing more than one-third of the model's predictive efficacy. This discovery corresponds with forecasting theory, as historical demand patterns generally represent the most reliable predictor of future trends due to the constancy of consumer behavior and market momentum. Seasonality is the second most significant factor at 0.20, underscoring the essential influence of cyclical patterns in demand forecasting, influenced by vacations, weather cycles, or industry-specific seasonal trends.

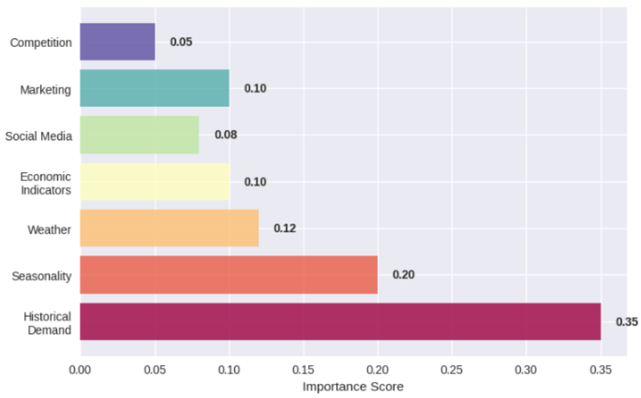

Figure 12: Feature importance in AI Model.

Weather circumstances have moderate significance at 0.12, indicating that meteorological aspects substantially affect consumer purchase decisions, especially pertinent for weather-sensitive items or services. Marketing efforts and economic indicators hold equivalent significance at 0.10 each, signifying that promotional activities and overarching economic conditions equally influence demand variations. The influence of social media is quantified at 0.08, indicating a discernible although relatively moderate effect on demand forecasting, likely illustrating the intricate and frequently indirect correlation between social media interaction and real consumer purchasing behavior. Competition exhibits the lowest importance score of 0.05, indicating that although competitive factors are relevant, they may be less directly discernible in the data, or their impacts are reflected indirectly through other variables. This underscores that internal factors (historical demand, seasonality) and environmental conditions (weather, economics) are more predictive than external competitive dynamics.

Data Availability

All datasets used in this research study are publicly available. The open-source datasets are available for further analysis at [57, 58].

Conflict of Interest

There are no conflicts of interest in this article.

Funding Statement

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

[1] Sierra Espinel, A. I., & Suarez Barón, M. J. (2025). Applying deep learning and forecasting techniques to the pharmaceutical supply chain. *Proceedings of the Computer Science*, 253, 2791–2800. doi:10.1016/j.procs.2025.02.003

[2] Abushaega, M. M., Moshebah, O. Y., Hamzi, A., & Alghamdi, S. Y. (2025). Multi-objective sustainability optimization in modern supply chain networks: A hybrid approach with federated learning and graph neural networks. *Alexandria Engineering Journal*, 115, 585–602. doi:10.1016/j.aej.2024.12.048

[3] Rahman Mahin, M. P., Shahriar, M., Das, R. R., Roy, A., & Reza, A. W. (2025). Enhancing sustainable supply chain forecasting using machine learning for sales prediction. *Proceedings of the Computer Science*, 252, 470–479. doi:10.1016/j.procs.2025.01.006

[4] Abdulrahim, H., Alshibani, S. M., Ibrahim, O., & Elhag, A. A. (2025). Prediction OPEC oil price utilizing long short-term memory and multi-layer perceptron models. *Alexandria Engineering Journal*, 110, 607–612. doi:10.1016/j.aej.2024.10.057

[5] Sánchez, J. C. M., Mesa, H. G. A., Espinosa, A. T., Castilla, S. R., & Lamont, F. G. (2025). Improving wheat yield prediction through variable selection using support vector regression, random forest, and extreme gradient boosting. *Smart Agricultural Technology*, 10. doi:10.1016/j.atech.2025.100791

[6] Yang, Z., & Liu, S. (2025). Fairness-oriented multi-objective optimization of supply chain planning under uncertainties. *Socio-Economic Planning Sciences*, 99. doi:10.1016/j.seps.2025.102198

[7] Cheng, J., He, X., Wen, J., & Su, L. (2024). How does spatial distance to travel companions transform to temporal distance in travel purchase decisions? *J. Hosp. Tour. Manag.*, 60, 166–176.

[8] Jiang, H., Messinger, P. R., Liu, Y., Lu, Z., Yang, S., & Li, G. (2024). Divergent versus relevant ads: how creative ads affect purchase intention for new products. *J. Mark. Res.*, 61, 271–289.

[9] Lee, S., & Johnson, Z. S. (2017). The effect of new product design and innovation on South Korean consumer's willingness to buy. *Asia Pac. J. Mark. Logist.*, 29, 98–113.

[10] Lee, J., Chu, W., & Roy, R. (2024). Enhancing consumer evaluation of new products: the role of innovation newness and communication strategy. *J. Bus. Res.*, 182, 114766.

[11] Liu, J., Xu, X. A., & Wang, L. (2024). How do travel experience sharing type, tourism activity type, and temporal distance shape online viewers' travel intentions? An application of construal level theory. *Curr. Issue Tour.*, 27, 1590–1612.

[12] Park, S., Wei, X., & Lee, H. (2024). Revisiting the elaboration likelihood model in the context of a virtual influencer: a comparison between high-and low-involvement products. *J. Consum. Behav.*, 23, 1638–1652.

[13] Stuppy, A., Landwehr, J. R., & McGraw, A. P. (2024). The art of slowness: slow motion enhances consumer evaluations by increasing processing fluency. *J. Mark. Res.*, 61, 185–203.

[14] Wang, X., Chen, X., & Gu, Z. (2024). Online information reviews to boost tourism in the B&B industry to reveal the truth and nexus. *Information*, 15, 103.

[15] Sun, Q., Gao, Q., Wu, B., Huang, J., Guo, H., & Jiang, C. (2024). Sellers and buyers with lower construal levels are more likely to reach second-hand transactions: the focus-separation effect between sellers and buyers. *Acta Psychol. Sin.*, 56, 1141–1156.

[16] Bigne, E., Simonetti, A., Guixeres, J., & Alcaniz, M. (2024a). Visual attention and product interaction: a neuroscientific study on purchase across two product categories in a virtual store. *Int. J. Retail Distrib. Manag.*, 52(4), 389–406.

[17] Fan, L., Wang, Y., & Mou, J. (2024). Enjoy to read and enjoy to shop: an investigation on the impact of product information presentation on purchase intention in digital content marketing. *J. Retailing Consum. Serv.*, 76, 103594. doi:10.1016/j.jretconser.2023.103594

[18] Finken, D., Hofstetter, R., Krishna, A., & Wangenheim, F. (2024). Context even matters in virtuality: perception of object proximity and ownership in augmented reality. *Adv. Consum. Res.*, 49, 282–283.

[19] Jiang, K., Zheng, J., & Luo, S. (2024). Green power of virtual influencer: the role of virtual influencer image, emotional appeal, and product involvement. *J. Retailing Consum. Serv.*, 77, 103660. doi:10.1016/j.jretconser.2023.103660

[20] Khan, I., & Fatma, M. (2024). AR app-based brand engagement and outcomes: a moderated mediation approach. *J. Retailing Consum. Serv.*, 76, 103618. doi:10.1016/j.jretconser.2023.103618

[21] Liao, S., Yu, L., Kruger, J. L., & Reichle, E. D. (2024). Dynamic reading in a digital age: New insights on cognition. *Trends in Cognitive Sciences*, 28(1), 43–55. doi:10.1016/j.tics.2023.08.002

[22] Schultz, C. D., & Kumar, H. (2024). ARvolution: decoding consumer motivation and value dimensions in augmented reality. *J. Retailing Consum. Serv.*, 78, 103701. doi:10.1016/j.jretconser.2023.103701

[23] Wu, J. F., Dong, J., Wu, Y., & Chang, Y. P. (2024). Shopping through mobile augmented reality: the impacts of AR embedding and embodiment attributes on consumer-based brand equity. *Inf. Manag.*, 61(6), 103999. doi:10.1016/j.im.2024.103999

[24] Xue, N., Liu, X., Wan, L. C., & Hou, Y. (2024). Relaxing or challenging? How social crowding influences the effectiveness of activity-based destination advertising. *Tourism Management*, 100, 104839. doi:10.1016/j.tourman.2023.104839

[25] Zhang, Y., Shao, W., Quach, S., Park, T., & Li, Q. (2024). Examining the moderating effects of shopping orientation, product knowledge and involvement on the effectiveness of Virtual Reality (VR) retail environment. *J. Retailing Consum. Serv.*, 78, 103713. doi:10.1016/j.jretconser.2024.103713

[26] Qiu, Y., Wang, Z., & Zheng, M. (2023). Can Digital Finance Improve Corporate Environmental Performance? Evidence from Heavy Polluting Listed Companies in China. *Emerg. Markets Finance Trade*, 60, 1–21.

[27] Liu, G., & Wang, S. (2023). Digital transformation and trade credit provision: evidence from China. *Research in International Business and Finance*, 64, 101805.

[28] Hu, Y., Che, D., Wu, F., & Chang, X. (2023). Corporate maturity mismatch and enterprise digital transformation: evidence from China. *Finance Research Letters*, 53, 103677. doi:10.1016/j.frl.2023.103677

[29] Tu, X. (2022). Digital transformation, knowledge spillover and Total factor Productivity: empirical evidence from listed manufacturing companies. *Industrial Economic Research*. doi:10.13269/j.cnki.ier.2022.02.010

[30] Wang, D. (2023). Digital economy development and total factor productivity of enterprises: theoretical mechanism and empirical test. *Industrial Technology and Economy*.

[31] Zhang, Z., Su, Z., & Tong, F. (2023). Does digital transformation restrain corporate financialization? Evidence from China. *Finance Research Letters*, 104152.

[32] Abolmakarem, S., Abdi, F., Khalili-Damghani, K., & Didehkhani, H. (2024). A multi-stage machine learning approach for stock price prediction: Engineered and derivative indices. *Intelligent Systems with Applications*, 24. doi:10.1016/j.iswa.2024.200449

[33] Danach, K., Dirani, A. E., & Rkein, H. (2024). Revolutionizing supply chain management with AI: A path to efficiency and sustainability. *IEEE Access*, 12, 188245–188255. doi:10.1109/ACCESS.2024.3474531

[34] Alkubaisi, S., Fenjan, A., Al-Sharify, T. A., Sadiq, S., Furaijl, H. B., & Ahmed, S. R. (2024). Machine learning models for optimizing supply chain management. *Proceedings of the 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1–9. doi:10.1109/ISMSIT63511.2024.10757215

[35] Ahmed, M. F., & Satchell, S. (2024). Conviction, diversification or something else: Constructing optimal portfolios with additional attributes. *Studies in Economics and Finance*, 41(4), 923–938. doi:10.1108/sef-04-2023-0207

[36] Mondal, A., Giri, B. K., Roy, S. K., Deveci, M., & Pamucar, D. (2024). Sustainable-resilient-responsive supply chain with demand prediction: An interval type-2 robust programming approach. *Engineering Applications of Artificial Intelligence*, 133. doi:10.1016/j.engappai.2024.108133

[37] Gao, Y., Lu, S., Cheng, H., & Liu, X. (2024). Data-driven robust optimization of dual-channel closed-loop supply chain network design considering uncertain demand and carbon cap-and-trade policy. *Computers & Industrial Engineering*, 187. doi:10.1016/j.cie.2023.109811

[38] Tugay, R., & Oguducu, S. G. (2020). Demand prediction using machine learning methods and stacked generalization. *arXiv*, 2020, eprint: 2009.09756.

[39] Levis, A. A., & Papageorgiou, L. G. (2005). Customer demand forecasting via support vector regression analysis. *Chem. Eng. Res. Des.*, 83, 1009–1018. doi:10.1205/cherd.05051

[40] Jain, A., et al. (2020). Demand forecasting for e-commerce platforms. *Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON)*, 1–6. doi:10.1109/INOCON50539.2020.9287383

[41] Bandara, K., et al. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. *Neural Information Processing (ICONIP 2019)*, LNCS 11955, 305–315.

[42] Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.*, 14, 35–62. doi:10.1016/s0169-2070(97)00044-7

[43] Kuo, C., & Reitsch, A. (1995). Neural networks vs. conventional methods of forecasting. *J. Bus. Forecast.*, 14, 17.

[44] Azzouni, A., & Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction. *arXiv*, 2017, eprint: 1705.05690.

[45] Ahmed, S., Chakrabortty, R. K., Essam, D. L., & Ding, W. (2024). A switching based forecasting approach for forecasting sales data in supply chains. *Applied Soft Computing*, 167. doi:10.1016/j.asoc.2024.112419

[46] Baycik, N. O. (2024). A quantitative approach for evaluating the impact of increased supply chain visibility. *Supply Chain Analytics*, 6. doi:10.1016/j.sca.2024.100065

[47] Qiu, Y., Wang, Z., & Zheng, M. (2023). Can Digital Finance Improve Corporate Environmental Performance? Evidence from Heavy Polluting Listed Companies in China. *Emerging Markets Finance and Trade*, 60, 1–21.

[48] Zhu, B., & Li, F. (2024). Unveiling the Unexpected: Transformer Outperforms Specialized Architectures in Volatile Supply Chain Forecasting. *International Conference on Computational Science and Computational Intelligence*, 113–126.

[49] Alrashdi, I., Abd El-Aziz, R. M., Taloba, A. I., & Farsi, M. (2025). Hybrid TCN-transformer model for predicting sustainable food supply and ensuring resilience. *Alexandria Engineering Journal*, 127, 380–391.

[50] Xu, J., & Bo, L. (2024). Enhancing supply chain efficiency resilience using predictive analytics and computational intelligence techniques. *IEEE Access*.

[51] Wang, S. (2024). Forecasting Demand and Controlling Inventory Using Transformer-Based Neural Network Optimized by Whale Optimization Algorithm. *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*, 915–923.

[52] Palan, J. (2024). Walmart's Integration Of AI, And AR Technologies. *IOSR Journal of Business and Management (IOSR-JBM)*, 26(6), 36–41.

[53] Anumula, S. R. (2025). Transforming Retail Logistics: Smart Receivings and Claims Management at Walmart. *Journal Of Engineering And Computer Sciences*, 4(7), 204–210.

[54] Googerdchi, K. F., & Jafari, S. M. (2025). Sales Forecasting in Retail Using Random Forest Regression: A Case Study on Walmart. *2025 4th International Conference on Computing and Information Technology (ICCIT)*, 592–597.

[55] Theodorou, E., Wang, S., Kang, Y., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2022). Exploring the representativeness of the M5 competition data. *International Journal of Forecasting*, 38(4), 1500–1506.

[56] Seaman, B., & Bowman, J. (2022). Applicability of the M5 to Forecasting at Walmart. *International Journal of Forecasting*, 38(4), 1468–1472.

[57] Yasserh. (2025). Walmart Dataset. `https://www.kaggle.com/datasets/yasserh/walmart-dataset`. Accessed: 2025-09-30.

[58] Kaggle. (2025). M5 Forecasting - Accuracy. `https://www.kaggle.com/competitions/m5-forecasting-accuracy`. Accessed: 2025-09-30.