Simulation Research On Secure Sharing Of Regional E-commerce AI Models Based On Sensor Data De-sensitization

Changqian Wu[1] and Lili Zheng[2*]

[1] *College of Finance and Commerce, Minxi Vocational & Technical College, Longyan 364021, Fujian, China.*
[2] *College of Information Technology and Engineering, Ningde Vocational and Technical College, Ningde 355099, Fujian, China.*
[*] *Corresponding author. E-mail: Liliheng424@outlook.com*

Received Apr.6, 2026; accepted May 15, 2026

Recent developments in e-commerce highlight the importance of AI models, particularly LSTM, for improving sales forecasting and capturing regional patterns. However, challenges remain in handling sensitive customer data, enabling secure cross-regional learning, and addressing regional variability. This study proposes a secure and scalable regional e-commerce forecasting framework integrating Differential Privacy (DP) and Secure Multi-Party Computation (SMPC). The proposed approach achieves strong performance (MAE=0.691, RMSE=0.951, MAPE=0.519, R^2=0.9663), ensuring accurate, scalable, and privacy-preserving forecasting.

Keywords: LSTM, Attention Mechanism, E-commerce Sales Forecasting, Regional Data, Sensor Data De-sensitization, Prediction Accuracy

http://dx.doi.org/10.6180/jase.202610_33.017

1. Introduction

The rapid growth of digital technologies has significantly transformed traditional business [1, 2]. Warehouses and logistics networks and consumer devices [3, 4]. The current technological advancements demonstrate that AI [5].

The novelty of the proposed work is the unified integration of Differential Privacy, LSTM, and SMPC into a single end-to-end framework for regional e-commerce forecasting. DP ensures data anonymization, LSTM captures temporal patterns, and SMPC enables secure model aggregation. This combined approach improves privacy, prediction accuracy, and secure collaborative learning compared to existing methods. A research gap exists in securely exchanging regional e-commerce AI models after sensor data de-sensitization while balancing privacy and performance. Existing methods often reduce accuracy and struggle with heterogeneous data, which the proposed framework addresses using Differential Privacy and SMPC for secure, high-performance collaboration.

The paper is organized as follows: Section 1 presents the background. Section 2 reviews related work. Section 3 describes the methodology with DP, LSTM, and SMPC. Section 4 presents results and analysis. Section 5 concludes the study and suggests future work.

The research conducted by Dong et al. [6], Ni et al. [7], and Pei et al. [8] focused on developing advanced analytical behavior research. Selahattin Bardak et al. [9] and Florin Cornel Dumiter et al. [10] studied how consumers make choices and how the digital economy functions.

2. Materials and methods

The proposed framework follows a modular architecture where data collection, preprocessing, de-sensitization, feature extraction, LSTM forecasting, and SMPC aggregation operate as independent modules. Figure 1 presents the architectural design of the system. The Long Short-Term Memory (LSTM) model was selected because it effectively learns time-dependent and sequential patterns in regional e-commerce data. Unlike traditional models such as SVM, DT, and RF, LSTM captures long-term dependencies through its gated memory architecture, making it suitable for accurate prediction using heterogeneous IoT data.

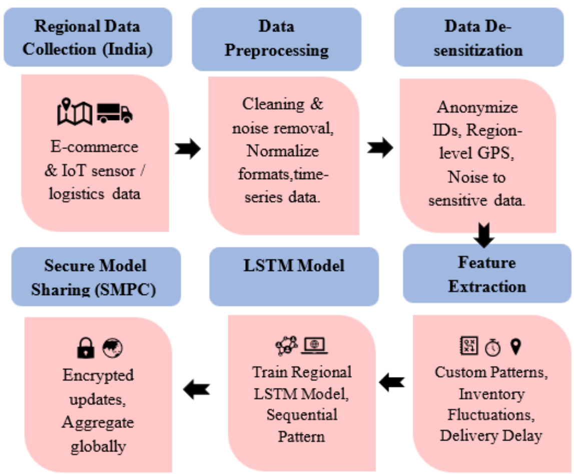

Fig. 1. Architecture of Sensor Data De-Sensitized Regional AI Model Sharing System

The Regional Data Collection dataset [11] was collected from e-commerce transaction logs and IoT logistics systems across Delhi, Mumbai, Chennai, and Kolkata over several months to capture seasonal variations. The dataset was divided into regional subsets for Delhi, Mumbai, Chennai, and Kolkata using transaction identifiers and delivery location data. The target variable was regional sales amount, predicted from historical transaction and logistics sensor data. This regional division helped the framework learn location-specific purchasing and delivery patterns for accurate forecasting.

The IoT framework used RFID, GPS, environmental, and traffic sensors to monitor inventory, delivery routes, storage conditions, and logistics performance. These sensor data were integrated with customer transaction data to support accurate regional sales forecasting. The transaction amount T_{t_i} for a transaction t_i by a customer c_j can be expressed as

In Equation 1, D_{cust} represent the customer data, where each customer $c_j$is associated with transaction details.

$$D_{\text{cust}} = \Big(\text{customer\_}_j^{\text{age}}, \text{customer\_}_j^{\text{gender}}, \text{membership\_}_j^{\text{level}}, T_{t_i}, \dots\Big) \tag{1}$$

In Equation 2, D_{inv} the database stores warehouse inventory information, which contains individual inventory. This inventory representation is further complemented by delivery and environmental data to capture logistics-related factors influencing regional e-commerce operations.

$$D_{\text{inv}} = \left(L_{I_k}, \text{product\_id}_k, \text{stock\_status}_k, \dots\right) \tag{2}$$

In Equation 3, D_{del} represent the delivery and environmental data. These integrated data components are then subjected to preprocessing to improve data quality and ensure consistency before model development. Let the raw dataset be $D = \{x_1, x_2, ., x_n\}$. Cleaned data is obtained by removing noise ϵ_i:

$$D_{\text{del}} = \left(\text{delivery\_pref}_{t_i}, E_{\text{env}}\right) \tag{3}$$

Equation 4, represents that x_i is the original data value, ϵ_i denotes noise component, and x_i^{clean} The cleaned dataset is subsequently normalized to ensure uniform feature scaling before further processing. The proposed framework applies min-max normalization uniformly across transactional, inventory, logistics, and sensor-based features to transform all attributes into a standardized range between 0 and 1 while preserving relative data distribution characteristics.

$$x_i^{\text{clean}} = x_i - \epsilon_i \tag{4}$$

In Equation 5, x_i denotes original feature value, x_{min}, x_{max} is the minimum and maximum values of the feature, and x_i^{norm} is a normalized value. The normalized data is then aligned temporally to ensure consistency across multiple data sources before further processing.

$$x_i^{\text{norm}} = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{5}$$

The temporal dependency modeling process considered periodic demand fluctuations such as weekly purchasing trends, seasonal shopping behavior, and recurring delivery patterns. A customer transaction recorded at 10:00 AM can be synchronized with warehouse inventory updates, delivery vehicle GPS status, and environmental sensor readings from the same time interval. The proposed preprocessing strategy enhances temporal consistency across the collected data streams. The LSTM model can more effectively capture regional sales and logistics patterns over time. The improved sequence representation contributes to stable and reliable predictive performance.

In this Equation 6, $X(t)$ is aligned dataset at time t, $x_j(t)$ denotes the value of feature $j$at time t, and m represents the number of data sources or features. The preprocessed and time-aligned dataset $X(t)$ is then passed to the Differential Privacy mechanism to protect sensitive information. The regional analysis also considers socio-economic and infrastructural factors such as population density, internet accessibility, transportation efficiency, purchasing capacity, and urban logistics. The AI model training dataset is prepared through secure methods which Figure 2 demonstrates.

$$X(t) = \{x_1(t), x_2(t), \ldots, x_m(t)\} \tag{6}$$

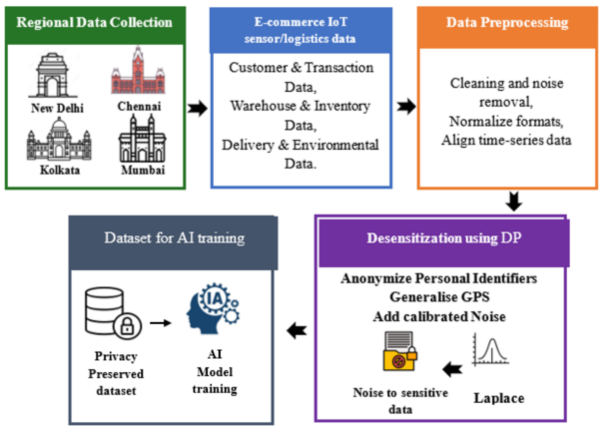

Fig. 2. Privacy-Preserving Data Preparation for Regional E-commerce IoT Data

The privacy guarantee is expressed as:

$$\Pr[M(D_1) \in S] \leq e^{\epsilon} \cdot \Pr[M(D_2) \in S] \tag{7}$$

Equation 7, M represents the privacy mechanism D_1 and D_2 are neighboring datasets differing by one record and ϵ is the privacy budget. A noise injection mechanism is required; therefore, the Laplace mechanism is employed to perturb the data in accordance with the defined privacy budget ϵ.

In Equation 8, the Laplace mechanism in Differential Privacy by adding noise to the original data value x. Here, $\tilde{x}$ is the perturbed output, and Laplace $(0, \Delta f/\epsilon)$ represents noise controlled by the sensitivity Δf and privacy budget ϵ. The interaction between Differential Privacy and secure parameter exchange is formalized as a sequential process. The LSTM model is then trained on D' to generate local model parameters θ_i. These parameters are encrypted and securely shared using SMPC without exposing raw data. The perturbed output$\tilde{x}$ satisfies the Differential Privacy constraint while maintaining controlled distortion, thereby enabling reliable feature extraction without compromising sensitive information. The extracted feature set F is used as input to the LSTM model for learning temporal patterns and forecasting regional trends. A variable screening process was applied during preprocessing. Principal Component Analysis (PCA) reduced feature redundancy and identified important attributes from transactional. as Figure 3 demonstrates.

$$\tilde{x} = x + Laplace\left(0, \frac{\Delta f}{\epsilon}\right) \tag{8}$$

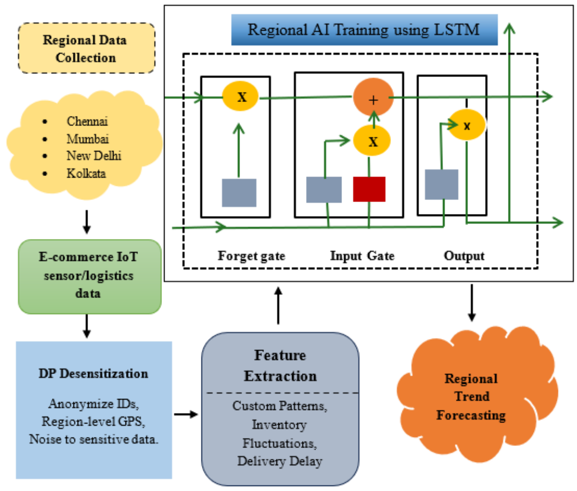

Fig. 3. Architecture of Regional AI Training LSTM with DP -Based Data De-sensitization

The LSTM computation begins with the forget gate, which determines the amount of past information to be retained from the previous cell state.

Equation 9 shows output f_t, input vector x_t at time t, previous hidden state h_{t-1}, weight matrices $W_f$and U_f, bias vector b_f, and sigmoid activation function σ.

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{9}$$

Equation 10 uses the input gate value i_t, input gate weight matrices $W_i$and U_i. The input gate determines the extent to which new information from the current input is incorporated into the cell state.

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{10}$$

Equation 11, is the updated cell state, c_{t-1} is the previous cell state, f_t is the forget gate output, i_t is the input gate output, $W_c$and U_c The updated cell state is computed by combining retained past information with newly learned input features.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{11}$$

The hidden state output at time t in Equation 12 is represented by h_t. The predicted output h_t from the LSTM model is used to compute the loss function, which measures the deviation between predicted and actual sales values for model optimization. The LSTM model is trained by minimizing a loss function that measures the difference between predicted and actual regional sales values.

$$h_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \odot \tanh(c_t) \tag{12}$$

In this Equation 13, y_t is actual sales value at time t, $\widehat{y}_t$– predicted sales value, N– number of observations, $L_i(\theta_i)$– loss function for region i.

$$L_i(\theta_i) = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2 \tag{13}$$

In this Equation 14, θ_i denotes model parameters of region i, η represents learning rate, $\nabla L_i(\theta_i)$ is the gradient of the loss function for region i. The proposed framework uses additive homomorphic encryption. Each regional LSTM secure aggregation on encrypted data.

$$\theta_i = \theta_i - \eta \nabla L_i\left(\theta_i\right) \tag{14}$$

In Equation 15, E_i is encrypted model update of region i, Enc denotes encryption function, θ_i refers local model parameters, k_i is the secret encryption key. In the protocol level, encrypted aggregation uses additive homomorphic encryption where each region encrypts its model parameters before sending them. The final aggregated result is decrypted only after computation. The trained local model parameters θi are then prepared for secure sharing through encryption and SMPC-based aggregation. The secure collaboration strategy combines secret-sharing mechanisms with privacy-preserving communication protocols to protect sensitive information during distributed model training. This approach enables participating nodes to exchange model updates securely without exposing raw data.

$$E_i = Enc\left(\theta_i, k_i\right) \tag{15}$$

In Equation 16, E_{agg} is the aggregated encrypted update, E_i is encrypted update from region i, n refers number of participating regions. The encrypted local model updates are then aggregated using SMPC to compute a combined global representation.

$$E_{\text{agg}} = \sum_{i=1}^{n} E_i \tag{16}$$

In Equation 17, where n is the number of regions and s is parameter size. This shows linear growth of communication overhead with increasing nodes.

$$C(n) = n \cdot s \tag{17}$$

In this Equation 18, θ_{global} denotes final global model parameters. The aggregated global model θ_{global} is then used for regional sales.

$$\theta_{\text{global}} = Dec\left(E_{\text{agg}}, k\right) \tag{18}$$

3. Results and discussion

Regional e-commerce transaction and IoT sensor datasets were used to evaluate customer behavior, inventory changes, delivery performance, and sales forecasting. The experimental setup was implemented in Python using TensorFlow, Keras, and scikit-learn on a system with 16 GB RAM and an Intel Core i5 CPU. Datasets from Delhi, Mumbai, Chennai, and Kolkata were split into training, validation, and testing sets with ratios of 80%, 10%, and 10%, while 5-fold cross-validation ensured reproducibility and stability.

The LSTM model was trained using a batch size of 32 and the Adam optimizer with a learning rate of 0.001 for stable convergence. The training process used 500 epochs to capture long-term temporal patterns in regional e-commerce data. These hyperparameters were selected based on preliminary experiments and standard deep learning practices. The experiments were conducted using regional e-commerce transaction and IoT sensor datasets collected from Delhi, Mumbai, Chennai, and Kolkata, ensuring consistency with the proposed framework. The implemented LSTM architecture contained two LSTM layers with 128 and 64 hidden units, followed by a dense output layer for sales prediction. A sequence length of 30-time steps was used to capture temporal dependencies, while the Adam optimizer with a learning rate of 0.001 and batch size of 32 ensured stable training. The model was trained for 500 epochs using mean squared error as the loss function. The LSTM model uses a two-layer stacked architecture with 128 hidden units in the first layer and 64 in the second to capture and refine temporal patterns in regional e-commerce data, improving both efficiency and reproducibility for stable forecasting performance. Adaptive learning-rate scheduling reduces the learning rate when validation loss plateaus, and early stopping is applied to halt training when improvements become negligible.

Figure 4 shows the distribution of customer transactions, Delhi, Mumbai, Chennai, and Kolkata.

Figure 5 presents normalized customer activity scores across Delhi.

Figure 6 demonstrates how DP ϵ is the privacy budget parameter that controls the trade-off between privacy protection and data utility, where smaller values provide stronger privacy guarantees through higher noise injection.

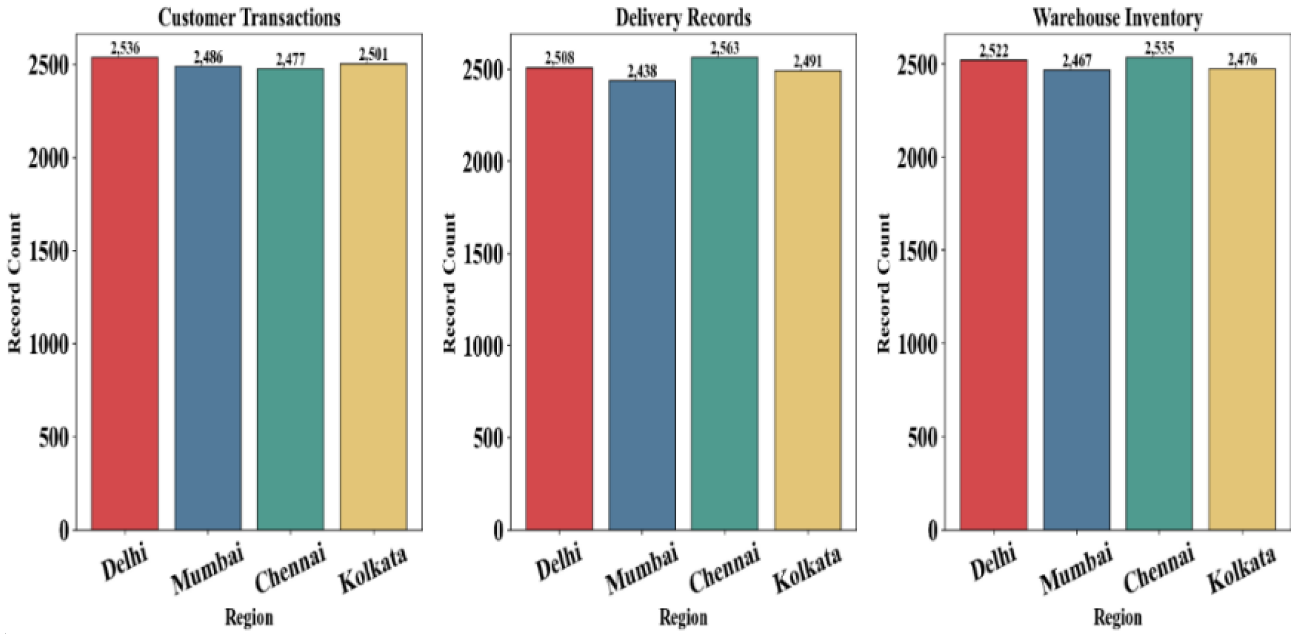

Fig. 4. Regional Data Distribution of E-commerce IoT Datasets

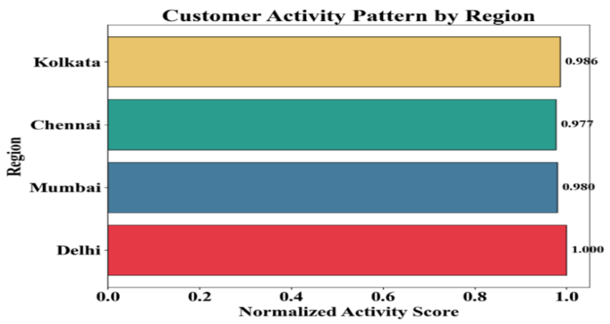

Fig. 5. Regional Customer Activity Pattern in E-commerce Transactions

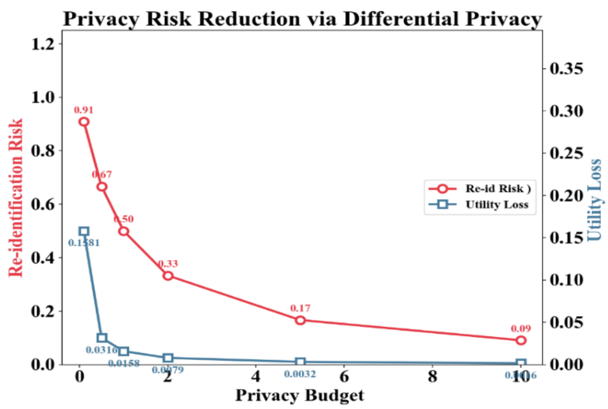

Fig. 6. Privacy Risk Reduction and Utility Loss under DP

Figure 7 displays IoT dataset according to different privacy budget values (ϵ). A parameter tuning strategy was applied to optimize the privacy budget (ϵ) by evaluating multiple values (0.1 to 10) using cross-validation. The optimal ϵ was selected based on a trade-off between privacy risk reduction and prediction accuracy, ensuring a balanced performance between anonymization strength and forecasting capability.

In Figure 8, High-activity regions show stable trends and lower prediction errors, while variable regions show higher error due to inconsistent demand. Despite Differential Privacy noise, LSTM maintains strong accuracy. SMPC overhead increases with more regions but stays low due to efficient aggregation. The proposed framework demonstrates a balanced trade-off between predictive accuracy, privacy preservation, and communication efficiency. The results indicate that reliable model performance can be achieved while limiting communication overhead in distributed environments. Furthermore, the incorporation of privacy-aware mechanisms supports secure data exchange during training.

Table 1 the model demonstrates efficiency through low communication overhead and computation time, while the 100% aggregation accuracy confirms consistent integration of encrypted regional model updates. The attention mechanism assigns weights to LSTM hidden states by computing relevance scores for each time step and normalizing them using softmax.

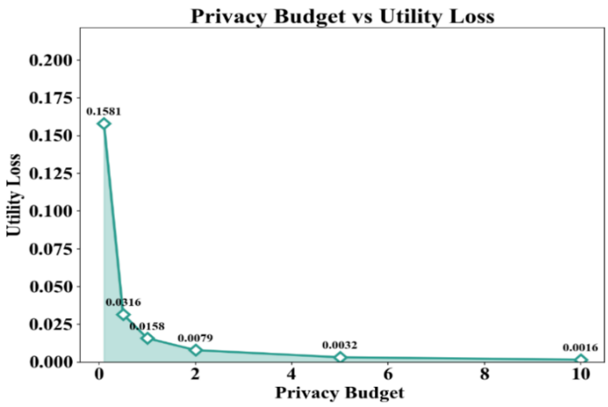

Fig. 7. Effect of Privacy Budget (ϵ) on Utility Loss in DP

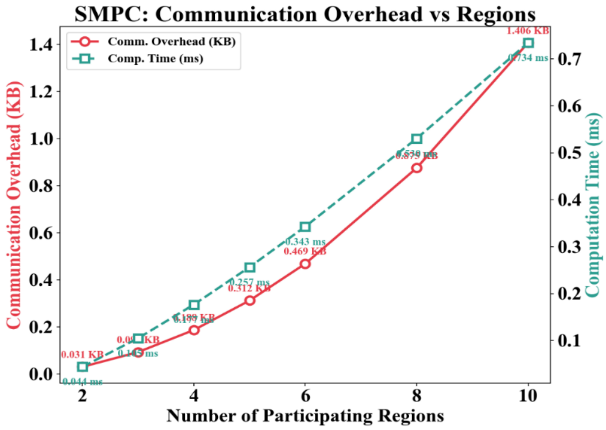

Fig. 8. SMPC Communication Overhead vs Regions

Table 2 compares the FL-LSTM enables decentralized learning across regions while preserving data privacy, aligning with the SMPC-based framework. long-term dependencies using self-attention but typically require higher computational resources and communication cost. The baseline models in Table 2 were selected because they are widely used deep learning and hybrid their architectures such as BiLSTM, attention mechanisms, stacking models, and TCN-LSTM are suitable for regional e-commerce forecasting and provide effective benchmarks for evaluating the proposed framework.

Dataset al.ignment was maintained by applying uniform preprocessing steps, including normalization, missing value handling, and feature standardization across all experimental scenarios. All experiments were conducted under controlled parameter settings to maintain reproducibility. Ethical AI considerations were recognized in the proposed regional forecasting framework is responsible and transparent model deployment. The proposed approach therefore contributes toward fair, privacy-aware, and ethically responsible AI-driven e-commerce forecasting.

Table 1. SMPC Evaluation Metrics

Metric	Value
Communication Overhead (KB)	0.19
Computation Time (ms)	0.1768
Aggregation Accuracy (%)	100.0000
Scalability (nodes)	4

4. Conclusion

Scalability constraints remain a key factor when extending the proposed framework to large-scale or high-dimensional datasets. Future research can explore cross-domain applicability to evaluate the generalizability of the approach across diverse application contexts. Addressing these aspects can enhance the overall effectiveness and broaden the impact of the proposed system.

Declaration

Author contributions

Changqian Wu: Conceptualization, Methodology, Formal analysis, Data Curation, Writing – Original Draft. **Lili Zheng**: Supervision, Validation, Writing – Review & Editing, Funding acquisition. Both authors have read and approved the final manuscript.

Funding declaration

This work was supported by the Vocational Education Research Project of Fujian Provincial Association of Vocational and Technical Education in the 2025-2026 academic year, titled 'Research and Practice on the Co-construction and Sharing Mechanism of AI Vertical Models under the Framework of City-level Industry-Education Consortium' (FJZJXHA25010).

Conflict of interest declaration

The authors declare no conflict of interest regarding the publication of this paper.

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors consent to the publication of this manuscript.

Table 2. Performance Comparison

Framework / Method	MAE	RMSE	MAPE	R^2 Score
BiLSTM [12]	0.798	1.124	0.610	0.9445
Transformer (TFT) [6]	0.712	0.984	0.528	0.9625
FL-LSTM (Federated Learning [8]	0.735	1.021	0.541	0.9587
Stacking-Based Fusion Model[13]	0.762	1.102	0.555	0.9562
Multi-Feature Attention Mechanism with Neural Networks[14]	0.824	1.190	0.675	0.9381
TCN-LSTM-Attention[15]	0.833	1.212	0.689	0.9335
LSTM Framework	**0.691**	**0.951**	**0.519**	**0.9663**

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] L. Jian, S. Guo, and S. Yu, (2023) *"Effect of Artificial Intelligence on the Development of China's Wholesale and Retail Trade"* **Sustainability** ***15***(13): 10524. DOI: 10.3390/su151310524.

[2] H. Alserhan, R. Altarawneh, N. Alyami, Y. Alsheyyab, R. Alrababah, and H. Alshamayleh, (2025) *"The challenges and opportunities of implementing predictive analytics in marketing strategies and e-commerce personalisation techniques"* **Asia Pacific Management Review** ***30***(4): 100409. DOI: 10.1016/j.apmrv.2025.100409.

[3] M. Shili, S. Hammedi, and M. Elkhodr, (2025) *"Spatial Intelligence in E-Commerce: Integrating Mobile Agents with GISs for a Dynamic Recommendation System"* **Algorithms** ***18***(1): 28. DOI: 10.3390/a18010028.

[4] X. Zhang and C. Guo, (2024) *"Research on Multimodal Prediction of E-Commerce Customer Satisfaction Driven by Big Data"* **Applied Sciences** ***14***(18): 8181. DOI: 10.3390/app14188181.

[5] Y. M. Tang, K. Y. Chau, Y. Lau, and Z. Zheng, (2023) *"Data-Intensive Inventory Forecasting with Artificial Intelligence Models for Cross-Border E-Commerce Service Automation"* **Applied Sciences** ***13***(5): 3051. DOI: 10.3390/app13053051.

[6] H. Dong, D. Wang, and S. Bashar, (2026) *"E-Commerce Supply Chain Resilience and Sustainability Through AI-Driven Demand Forecasting and Waste Reduction"* **Sustainability** ***18***(1): 360. DOI: 10.3390/su18010360.

[7] L. Ni, Z. Huang, and N. Fu, (2025) *"A Stacking-Based Fusion Framework for Dynamic Demand Forecasting in E-Commerce"* **Mathematics** ***13***(21): 3436. DOI: 10.3390/math13213436.

[8] Y. Pei, J. Zhu, and J. Cao, (2025) *"Intergenerational Differences in Impulse Purchasing in Live E-Commerce: A Multi-Dimensional Mechanism of the ASEAN Cross-Border Market"* **Journal of Theoretical and Applied Electronic Commerce Research** ***20***(4): 268. DOI: 10.3390/jtaer20040268.

[9] S. Bardak, (2026) *"Predicting Smart Tablet Preferences in Turkish E-Commerce Platforms Using Artificial Neural Networks and Machine Learning Techniques"* **Applied Sciences** ***16***(2): 832. DOI: 10.3390/app16020832.

[10] F. C. Dumiter and K. B. Schebesch, (2025) *"Using Artificial Intelligence to Determine the Impact of E-Commerce on the Digital Economy"* **Journal of Theoretical and Applied Electronic Commerce Research** ***20***(3): 219. DOI: 10.3390/jtaer20030219.

[11] Kaggle. *Regional Data Collection*. Accessed: 2026-03-16. 2026.

[12] H. Hu, J. Cai, and C. Xu, (2026) *"A Mathematical Framework for E-Commerce Sales Prediction Using Attention-Enhanced BiLSTM and Bayesian Optimization"* **Mathematical and Computational Applications *31***(1): 17. DOI: 10.3390/mca31010017.

[13] S. Yu, M. Guo, X. Chen, J. Qiu, and J. Sun, (2023) *"Personalized Movie Recommendations Based on a Multi-Feature Attention Mechanism with Neural Networks"* **Mathematics *11***(6): 1355. DOI: 10.3390/math11061355.

[14] Z. Huang and J. Liu, (2024) *"TransTLA: A Transfer Learning Approach with TCN-LSTM-Attention for Household Appliance Sales Forecasting in Small Towns"* **Applied Sciences *14***(15): 6611. DOI: 10.3390/app14156611.