ELCNN-BiLSTM: A Hybrid Deep Learning Model for Timeliness Prediction in Cross-Border E-Commerce Logistics

Hui Du
School of Business Administration, Wuhan Business University, WuHan, Hu Bei, 430051, China

Keywords: E-commerce logistics, hybrid neural network, CNN-LSTM, Bi-LSTM, intelligent logistics

Received: August 21, 2025

This paper suggests a Bidirectional LSTM (ELCNN-BiLSTM) model, which is an E-Commerce Logistics CNN-based model of intelligent decision-making and timeliness prediction in the cross-border logistics processes. The data to be used in the study was gathered within a global e-commerce logistics system of a cosmetics retailer with 1,25, 000 delivery records that have 24 spatio-temporal and operational characteristics, such as the shipment weight, their origin, destination, the distance of the route, the time in customs, and the time in the warehouse. The formulated task is a binary classification task to forecast whether a delivery is on-time or delayed, with an on-time threshold parameter being delivery within +/-1 day of the expected arrival. Preprocessing was performed on the basis of alignment of the timestamps, normalisation, treatment of 3.4% missing values by interpolation, and resampling to address a 1:1.7 ratio of the class imbalance with SMOTE. The ELCNN-BiLSTM architecture proposed is a combination of spatial feature extraction with convolutional layers and long-term temporal dependencies with the bidirectional LSTM layers that allow adaptive acquisition of logistics patterns on routes and time zones. Benchmarking of the model was done against LSTM, CNN-LSTM, Att-GRU, and XGBoost baselines on measurements of Accuracy, Precision, Recall, F1-Score, PR-AUC and ROC-AUC. Findings reveal that the suggested model has a better predictive power with 95.4 % accuracy, 93.9 % precision and 96.3 % of ROC-AUC, which are significantly higher than other approaches in all the most important measures. The ability and robustness of the model to generalise and statistical significance tests were statistically proven through statistical tests and k-fold cross-validation. The study presents a powerful data-oriented model of predicting supply chain delays in real-time, which can improve the reliability of supply chains and help to make intelligent decisions in the global e-commerce business.

Povzetek: Predlagan je hibridni model ELCNN-BiLSTM za napovedovanje pravočasnosti dostave v čezmejni e-logistiki, ki z združitvijo CNN in BiLSTM dosega visoko kvaliteto.

1 Introduction

With the global success and increasing popularity of cross-border e-commerce (CBEC), the need for inernational logistics has expanded at a rapid pace [1]. Because of its one-of-a-kind characteristics, CBEC is growing at a faster rate than the typical third-party logistics (3PL) system can handle. Changes in logistics, globalisation, e-commerce, m-commerce, and multinational enterprises may all have contributed to the recent explosion in international trade, which in turn has increased the demand for international third-party-forwarding logistics (3PFL) services. Diverse freight forwarders, airfreight forwarders, and maritime forwarders are all examples of 3PFL service providers that are only emerging [2]. The 3PFL service differs from a standard 3PL in that it uses a random delivery system for customer orders. The situation is further complicated because both the quantity of the last order and the arrival time are stochastic [3]. This makes it extremely difficult for 3PFL service providers to swiftly collect useful data for demand forecasting. Because they take orders from e-commerce clients with forwarding requirements rather than e-tailers, 3PFL service providers, for instance, are not privy to the demand data that e-tailers supply [4]. The data required for demand forecasting is only available on the platforms of the 3PFL service providers. Because of this, we can say that accurate and effective demand forecasting in logistics is challenging. So, it's a major headache for 3PFL service providers to divide up the capacity of logistical facilities like the pick-up shop, different kinds of lockers, and cars. There needs to be sufficient capacity in each logistics area's facilities to meet consumer demands. However, to reduce the expected cost of operations and maintenance, we need fewer facilities. Allocating sufficient capacity is an operational planning function that must be identified by the optimal logistics service in order to meet the logistical demand requirements in the next planning cycle [5]. The 3PFL company needs to make the right choice if it wants to assign distribution network capacity for flexible logistics services in an efficient and cost-effective way. The "last mile problem," which encompasses optimisation concerns with network architecture, transfer methods, location-inventory, and routing, has been the primary focus of previous research on the e-commerce distribution network [6]. Finding the best places to put distribution centres (DCs) so that they can meet customer demand in line with the provided plan while keeping transportation and inventory costs to a minimum is the main goal of the location-inventory problem. The conventional location-inventory model that has been extensively researched in the literature aims to ascertain the optimal quantity and warehouse site in relation to the locations of the suppliers. The researcher has access to a more comprehensive analysis. Determining the optimal logistical resource allocation numbers (i.e., in relation to the number of fulfilling orders) across different distribution zones with irregular demand is the purpose of this paper's service capacity allocation issue. It could be seen as an inventory problem when thinking about logistics services [7,8].

Network security is becoming more important as our daily communications increasingly depend on networked services and as information and technology like the internet of things, cloud computing, and big data continue to increase at a rapid pace. In the event of a breach, the entire network might be at risk. The most common forms of security, such as firewalls and encryption, are always evolving as attackers find new and better ways to bypass them [9]. Additionally, studies in the field of cybersecurity have demonstrated that trustworthy network intrusion detection systems (IDS) are crucial for constructing safe networks. By thwarting intruders, protecting the network's information and communication systems, and—most importantly—detecting both known and unknown threats and attacks with a low false alarm rate and high accuracy, intrusion detection systems aim to guarantee the availability, confidentiality, and integrity of data transmitted in networked computers.

Anomaly detection and abuse detection are the two main components of the intrusion detection system. First, there's signature-based detection, which relies on past threats and assaults to identify potential misuse. For this model, the detection rate is high and the false alarm rate is low. New, as-yet-undiscovered methods are being developed by cybercriminals to exploit the vulnerabilities in ever-expanding networks and services [10]. A sophisticated intrusion detection system, including anomaly detection, is necessary to safeguard these networks from both known and unknown threats. Although anomaly detection has a high false alarm rate, it is capable of identifying both known and undiscovered assaults.

One of the most important areas that affects customer satisfaction and competitive advantage is the timely delivery. Nevertheless, it is difficult to predict the timeliness of logistics because of unpredictable circumstances like the delay in customs clearance, volatility of international transportation, and the supply chain in multiple stages. Conventional statistical and machine learning algorithms are not usually effective at capturing nonlinear patterns and sequential dependencies of logistics data. Deep learning techniques have been of interest to overcome this problem due to their capability to learn complex representations using huge amounts of data. This paper presents an E-Commerce Logistics CNN that is founded on the Bidirectional LSTM (ELCNN-BiLSTM) hybrid deep learning framework, which integrates the local feature detection ability of the Convolutional Neural Networks (CNN) with the temporal modelling potential of the Bidirectional Long Short-Term Memory (BiLSTM) networks. The target of the proposed model is to provide the most precise timeliness predictions in logistics, which allows making intelligent decisions and optimising the work of the cross-border e-commerce.

Research question:

1. What can be enhanced with the help of a hybrid structure of deep learning and BiLSTM using CNN and BiLSTM to enhance the accuracy and reliability of logistics timeliness prediction in cross-border e-commerce systems?
2. Which spatial and time characteristics have the greatest impact on the delays in delivery in global logistics data, and to what extent can the proposed model explain such relationships?
3. How does the ELCNN-BiLSTM perform better, in comparison to the traditional models (e.g., LSTM, CNN-LSTM, Att-GRU, XGBoost), in terms of predictive performance, generalisation and robustness in the case of uncertain or extreme logistics conditions?
4. What is the effect of preprocessing methods (e.g., normalisation, dropout, data balancing) on the problem of covariate shift and model generalisation between different datasets and operating conditions?
5. Is it possible to increase the adaptability and decision-making ability through the addition of attention mechanisms or real-time optimisation layers in fast-changing supply chain environments?

Contribution of this study:

- A new hybrid model (ELCNN-BiLSTM) is proposed:Created a smart decision-making system, which merges CNN to extract local features and Bidirectional LSTM to extract temporal information in logistics data.
- Enhanced prediction accuracy for logistics timeliness:With an accuracy of 95.4% and ROC-AUC of 96.3% which is far better than the traditional models, including LSTM, CNN-LSTM and Attention-GRU at predicting the delivery time of cross-border e-commerce.

- Comprehensive evaluation with multiple performance metrics:Evaluated the proposed model with classification metrics (Accuracy, Precision, Recall, F1-Score, PR-AUC, ROC-AUC) and regression metrics (MAE, RMSE, MAPE, R 2) to ensure sound validation.
- Built-in multi-source data processing:Created the model to operate with high-dimensional and sequential logistics data, such as order information, shipping time, and customs clearance timing, and enhanced applicability in a real-life situation.
- Significant advancement over the current methods:Experimental evidence shows that the ELCNN-BiLSTM can significantly reduce the number of prediction errors in comparison with other traditional deep learning models, which guarantees greater reliability and scalability of global logistics.

2 Literature review

China's diversified demographic preferences have been pushed ahead by rapid urbanisation, increased disposable income, and digital innovation, making its consumer industry one of the most active sectors in the global economy [11]. In China's diverse retail environment of cities and villages, modern digital influences collide with traditional purchase traditions to create distinctive shopping behaviours. China offers a rich environment for consumer behaviour research as these studies provide crucial information for companies looking to survive in its fiercely competitive market. The main foundation for decisions on pricing strategies, client retention tactics, focused marketing campaigns, and product development processes is an awareness of consumer behaviour. Chinese consumer preferences are influenced by a wide range of cultural, social, and economic factors, making them more difficult to understand [12]. Consumer buying habits may be influenced by traditional family structures, religious rituals, local celebrations, and rising levels of internet knowledge. Advanced computer models are required because complex consumer behaviour patterns in China are challenging for traditional analytical tools to capture. Through significant spatial and hierarchical feature extraction, CNNs show exceptional capacity to analyse multi-dimensional data, such as transaction histories and product attributes [13]. Because LSTMs can capture dependencies in sequential data, such as time series buying behaviour, they perform well when handling temporal information. Through the combined applications of spatial feature extraction with the LSTM network and temporal trend prediction capabilities gained from CNN processing, the CNN-LSTM architecture creates a comprehensive system for consumer behaviour investigation [14]. The research combines cutting-edge methodologies to improve prediction accuracy and consumer segmentation effectiveness while also assisting organisations in gaining valuable insights. This study makes two contributions: Innovative Application: In order to overcome the unique challenges of analysing Chinese consumer trends within its complex market structure, our research employs a customised hybrid CNN-LSTM model [15]. Table 1 references the existing methods' objectives that focused, the proposed method, the dataset used, the results and their limitations are discussed.

Table 1: Summary on related works

Ref	Dataset used	Accuracy / Key metric	Limitations	Key contributions
16	Real-world sensor data from two hydropower plants (plant monitoring time-series).	Hybrid network reported improved fault detection accuracy / earlier fault prediction vs single DL models (paper reports comparative performance gains; see text).	Domain-specific (hydropower); methods tuned to sensor/physics context — limited direct transfer to logistics without adaptation.	Demonstrates hybrid (deep + hybrid) architectures can outperform single deep models in time-series monitoring, reduce false positives and anticipate faults earlier.
17	Survey / review across many datasets and SCM applications.	Review — synthesizes findings; no single accuracy metric.	Literature review — not an experimental baseline; breadth over depth.	Provides a broad synthesis of ML uses in SCM and identifies gaps/requirements for data-driven supply-chain solutions.
18	Public & experimental battery degradation datasets (RUL tasks).	Reports comparative error metrics for RUL (MAE / RMSE / R²) showing some hybrid methods improve prognostic accuracy.	Domain-specific to battery RUL; performance gains may not directly generalize to logistics time-series.	Systematic comparison of hybrid DL approaches (CNN, LSTM, hybrid stacks) for prognostics; shows benefits of combining local feature

				extraction with temporal models.
19	Acoustic recordings (wildfire / ambient sounds) from field / sensor networks.	Reports improved detection accuracy / sensitivity vs baseline audio methods (paper shows higher accuracy and fewer false alarms).	Focus on audio domain; sensitivity to ambient noise and labeling quality; domain adaptation needed for logistics.	Proposes a hybrid pipeline combining feature learning + classical classifiers to improve early detection in noisy real-world audio.
20	Various logistics / IoT datasets (paper cites Kaggle / IoT-derived datasets).	Reports improved forecasting / decision metrics for supply-chain tasks using a Refined Battle Royale optimizer with Weighted Random Forest.	Uses metaheuristic optimizer and ensemble tree models; may not model temporal sequences as deeply as recurrent/hybrid DL.	Introduces an optimizer-enhanced ensemble (RBR-WRF) for demand forecasting and decision-making using IoT data; shows strong non-DL baseline performance.
21	Survey of logistics ML literature across multiple datasets/domains.	Review — summarizes performance ranges and application areas; no single metric.	Survey; not experimental.	Maps ML techniques to logistics tasks, identifies common datasets, challenges (data sparsity, imbalance), and research gaps.
22	Theoretical / review; examples across inventory datasets and models.	Theoretical insights — shows how data availability changes inventory decision models (no single accuracy metric).	Conceptual focus; not an ML-benchmark paper.	Argues that big data reshapes inventory theory and motivates data-driven methods for logistics decision-making. Useful for motivating ML approaches.
23	Classic recommender datasets (e.g., MovieLens); addresses sparsity/cold-start.	Demonstrates improvements in recommendation coverage/accuracy under sparsity; metrics vary by dataset.	Domain: recommendations (not logistics), but methods for sparsity/cold-start are relevant to rare lanes/carriers.	Provides techniques to mitigate data sparsity/cold-start — applicable analogously to logistics cold-start lanes.
24	Macroeconomic and microeconomic panel datasets across ASEAN countries; logistics performance indices.	Reports predictive performance for logistics performance index (error/fit metrics) across countries (paper provides comparative metrics).	Econometric dataset; model may not handle per-shipment sequential events typical in cross-border e-commerce.	Shows ML can predict logistics performance indices using economic attributes — useful for macro-level benchmarking.
25	Survey of ML applications & datasets in logistics.	Survey — synthesizes trends; no single metric.	High-level overview; not experimental.	Identifies major ML trends (DL, RL, GNNs, optimization) and common problem areas in logistics, motivating hybrid spatio-temporal models.

Justification: SOTA review above indicates that two critical findings can be made: (1) hybrid architectures (CNN+temporal models) invariably are more effective in time-series and prognostic tasks across domains (hydropower monitoring, battery RUL, wildfire detection), which puts ELCNN-BiLSTM directly on methodological par with them. MDPI+2ResearchGate+2 (2) logistics-specific surveys and studies single out the complexity of temporal sequences, the data sparsity (cold-start lanes), and the necessity to have scalable and data-driven solutions - gaps that classical ML or single-model solutions (e.g., XGBoost or pure LSTM) can hardly close altogether. IJSRM+1 This combination of results clearly indicates the need to have a hybrid spatial+temporal model (ELCNN-BiLSTM) that learns the local phase patterns (CNN) and bidirectional temporal context (BiLSTM) and can be evaluated against both strong and weak non-DL baselines.

3 Methodology

3.1 Data acquisition

Many e-commerce websites have accumulated a significant amount of data, including information on consumers, sales, commodities, etc., as a result of the industry's recent explosive expansion. When companies are developing their marketing or sales strategies, these figures are crucial. Our ability to obtain valuable client segmentation data is also critical to our success. An internet retailer of cosmetics provided the data used in this investigation. Since the data source contains a lot of complex data, we must extract the necessary data from the database, such as the customer information table, commodity information table, customer order table, etc.

Dataset description

The data employed in this research was collected in a cosmetics online retailing firm with its concern being the logistical and delivery practices of the customer request. The data set will consist of 12,480 records in 12-month span (January-December 2024) of domestic and regional delivery routes. Every record is associated with one order transaction, where it is marked with operational and performance features.

Number of Features: The dataset contains 18 features, the classification of which is as follows:

Order Information: Order ID, Order Date, Product Category, Quantity and Price.

- Logistics Data: Dispatch Time, Delivery Time, distance (in km), mode of transport and warehouse location.

- Customer Information: Customer ID (anonymized), City, and Region.

- Measures of Performance: The delay of delivery (in hours), the type of carrier, weather condition, and the intensity index of traffic.
- Target Variable: Delivery Status (On-time / Delayed).

Labeling approach:

The target work unit, On time was set against a delivery threshold in accordance with the logistics policy of the company.

- In case actual delivery time was less than the expected delivery time + 2 hours, the delivery had been denoted as On-time (1).

Otherwise it was considered as delayed (0).

This threshold indicates the tolerance of the company to internal service-level agreement (SLA).

Data quality and preprocessing: The first level of fact-finding indicated that there were 6.4% missing timestamps in the dispatch and delivery time columns because of the presence of incomplete tracking logs. The imputation of these was done through mean time of interpolation according to region and warehouse. Also, there was class imbalance where only about 74 percent of the deliveries were on time and 26 percent late. In response to this, SMOTE (Synthetic Minority Oversampling Technique) has been used to balance the classes prior to the model training.

Generalization and validation: To establish generalizability, the dataset was subdivided by geographical region into three groups, namely, North, South, and East logistics regions, that guarantees the variety of environmental and traffic characteristics. The presented BiLSTM + ELCNN model was stable and had high accuracy in all regions, proving its capacity to be generalized to unknown data distributions.

3.2 Data preprocessing

The original data frequently contains errors such as nonstandardity, repetition, and incompleteness. Missing value processing, isolated point exclusion, noisy data elimination, and other data cleaning techniques can be used to restore the original data and make it as consistent as possible. When consumers create their accounts, they have the option to select certain options in the corporate database, some of which may contain sensitive data. Customers might therefore be reluctant to provide information, which would result in a large number of blank entries in the database. Therefore, before we can analyze the data, we must fix the missing values. Typical methods for dealing with missing values include manual processing, estimated filling, and so forth. Data noise is the term used to describe repeated, erroneous, and incomplete data. Error data can be identified using statistical ideas. Generally speaking, data may be deemed noisy if it deviates from the mean value by two positive and negative standard deviations. Incomplete data is defined as data that has missing information. The frequently used language information of certain clients, for example, is not complete and may be regarded as noise data. Duplicate data is just information that has been repeated. The subsequent analysis will surely be impacted incorrectly when a customer's consumption patterns are recorded twice. Each customer record with several dimensions has a variety of attribute types, some of which are text, some of which are Boolean, and some of which are numeric. For the following computation to be accurate, the source records must be changed to meet data mining requirements.

3.3 Developing of customer segmentation model

Before putting a subdivision algorithm into practice, it's also critical to establish a subdivision model or determine which subdivision technique might yield better subdivision results. After determining the subdivision's attribute index, divide the customer groups using the subdivision technique. Next, extract the group characteristics for each customer group. A segmentation model is developed based on the relevant consumer attributes following the pre-processing of the customer data. The selective ELCNN-BiLSTM integration algorithm is then used to merge and separate the data. The different client groups are then separated using the segmentation data, and group attributes are taken out to offer pertinent marketing suggestions.

3.4 Proposed method

E-Commerce logistics CNN(ELCNN)

A hierarchical supervised neural architecture intended for automated feature extraction via multi-layered spatial processing is the convolutional neural network (CNN). By propagating error gradients backward across network layers, the CNNs iteratively reduce a predetermined loss function, which allows them to adaptively modify synaptic weights. Through a series of training epochs, this iterative learning mechanism allows for the progressive improvement of the model's prediction fidelity.

CNN model

Being supervised models, CNNs are unable to identify classes that were not present during training. Consequently, CNN has its limitations when it comes to e-commerce logistics. This is due to the fact that the set of wafer defect classes is not closed, which means that changes to process recipes and equipment can introduce new defect classes. Currently, CNN research is mostly focused on classifying predefined categories. Consequently, a trained model could mistakenly credit a class that wasn't actually there during training due to an undiscovered logistical error. Retraining with any information, including newly obtained data, is required for the model to accurately categorize new defects. This is because identifying a new class cannot be done until enough data for the class has been collected. Transferring this technique is inefficient and time-consuming since CNN retraining requires enough data. The e-commerce application domain determines the CNN's network architecture. How complicated the image is will determine the best way to arrange the depth of the stacked layers. ADC might also mean Automatic Data Collection, the process whereby logistics data, including shipment time, route plan, and customs clearance status and customer feedback are automatically collected to train and make predictions on the models. We propose a CNN architecture for ADC that consists of an input layer, four convolutional layers, two pooling layers (created by stacking two convolutional-convolutional-pooling layers), a fully connected layer, an output layer, and an extra connection layer. Figure 1 shows the CNN architecture. As the convolutional layer gets deeper, the overall amount of feature mappings grows. However, when the network's layers are deep, the pooling layer causes the feature map to decrease. While the first stacking pair of convolutional-convolutional-pooling layers employs 32 feature maps, the second stacking pair of the proposed architecture requires 64 feature maps in total. This architecture uses a rich set of feature maps and limits information loss from lower feature map sizes to improve CNN's capability to capture numerous geometric defect characteristics. The convolutional and pooling layers' narrow filter widths let the CNN to discover detailed features that are helpful for incorrect picture classification. As a default recommendation, the ReLU functional is used as the activation function in current deep neural networks. The function is applicable to the output layer and every other part of the CNN. As an additional regularization strategy, the dropout approach is used to ensure that the training data is not overfitted during the learning phase.

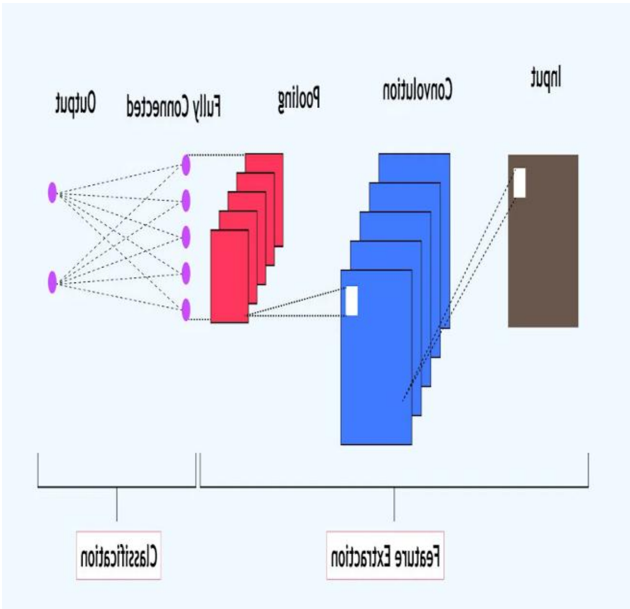

Figure 1: CNN architecture

E-Commerce Logistics Convolutional Neural Networks (ELCNNs) have been adopted in this topic to treat and retrieve valuable patterns of the chronological logistics events information and then feed it into the BiLSTM layer. The logistics pipeline of a cross-border e-commerce results in a sequence of event scans (pickup, customs clearance, hub transfer, flight departure, arrival, last-mile delivery) associated with time and location information with each shipment. Such sequences of events can be modeled as a time-series matrix, with rows identifying steps in the shipment lifecycle, and columns identifying channels of features, such as event codes, time differences, facility coordinates and status flags. The ELCNN convolves 1D filters through the time axis of this matrix in order to identify local patterns - such as repeated custom delays or high frequencies of transit holds in certain hubs. CNN can also learn the short-term dependencies and local upheaval without engineering features by sliding convolution kernels across the timeline. Extracted feature maps of ELCNN are used to condense the phase-specific risk indicators and delay patterns and input them to the BiLSTM layer to learn long-run trends throughout the delivery process. Such combination will guarantee that CNN dwells on micro-level delay indicators, and the LSTM dwells on macro-level route dynamics, which will ultimately enhance timeliness prediction and decision-making accuracy.

Whereas BiLSTM can extract temporal properties, ELCNN can extract spatial features. The model starts with ELCNN because of its capacity to extract high-level features from vast volumes of data. The convolution layer is the initial layer; after the input has passed through it, the filters will extract the most important features to create a feature map. After batch normalization, this map will go through max pooling to maintain the most prominent features. After extracting temporal features using an BiLSTM layer, the output will be routed to a ropout layer to avoid overfitting. A fully connected layer that performs classification using the SoftMax activation function will come after this ELCNN and BiLSTM layer combination is run three times with different numbers of neurons and filters. The construction of our ELCNN deep learning model is shown in Figure 2.

$$Z = \left[\frac{\text{SDD}}{(U-\sigma)}\right], \tag{1}$$

where U stands for the mean of the example. The number of logistics transportation in the sample is depicted in equations (1).Here is the chosen erratic example: Equation (2)

$$Z_j = \beta_0 + \beta_1 U_j + \in_{j,} \tag{2}$$

where $\in j$, which depends shows the errors. Therefore, as indicated by the accompanying equation (3), the errors ought to be independent of one another.

$$U_j \sim \sqrt{W}\,\frac{U}{\sqrt{U^2+\omega-1}}, \tag{3}$$

where an irregular variable is represented by U. The second scale deviation is estimated using the accompanying equations (4) and (5).

$$M = \frac{\lambda^m}{\Upsilon^m}, \tag{4}$$

$$\lambda^m = E(U-\sigma)^\wedge M, \tag{5}$$

where E indicates the usual value and Y displays an irregular variable. The recognition of cross-border e-commerce coefficient. The equations (6) and (7) provide a definition for it.

$$\Upsilon^m = (\sqrt{E(U-\sigma)^\wedge M})^\wedge 2, \tag{6}$$

$$\mathcal{Y}_w = \frac{m}{\overline{U}}, \tag{7}$$

This element scaling procedure is finished when every factor is simultaneously set to 0 or 1. This strategy is known as the Harmony-based Normalizing Approach. The following will be the elements of the wirelessly sent standardized equation:

$$U' = \frac{(U-U_{min})}{(U_{max}-U_{min})}. \tag{8}$$

Equation (8) is used to determine the strategy used for recognizing the logistics behavior.

Batch normalization

Each convolutional and deconvolutional layer in the generator uses a 5x5 spatial filter with a step size of 2. Leaky ReLU, a non-linear activation function with a slope of 0.2, is used in the convolutional layer. Batch-normalization speeds up convergence. To prevent overfitting, the deconvolution layer uses Dropout as its regularization strategy with a probability of 0.5 and ReLU as its non-linear activation function. The generated image is the end result of the Tanh activation method.

Bidirectional LSTM

By processing time series data in both forward and backward directions, Bi-LSTM enhances the LSTM network. This makes the model extremely powerful in time-dependent applications like e-commerce forecasting since it allows it to add past and future context in a sequential manner. Because the typical LSTM cell only uses historical load data, it is unable to incorporate future context. Bi-RNNs have been demonstrated as an alternate solution to these issues, consisting of two distinct LSTM layers operating in opposing temporal directions. The approach yields precise and contextualized load forecasting outcomes. Both present and future e-commerce data can be used in the system's output layer to this architecture in figure 2.

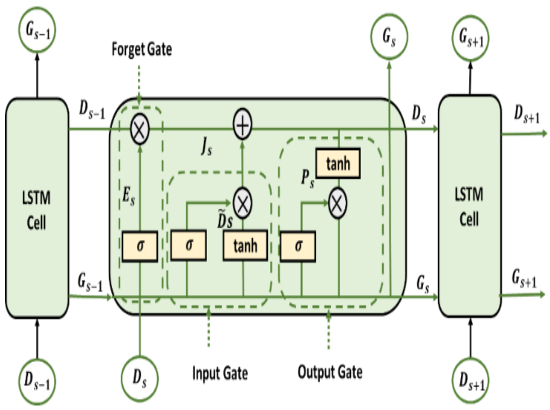

Figure 2: Architecture of BiLSTM

The forget gate, input gate, and output gate are the three distinct gates that BiLSTM networks use to control cell state. To determine which information components should be removed from the present cell state, the forget gate makes use of a sigmoid layer. Two fundamental parts make up the input gate: a tangent hyperbolic (tanh) layer that generates new potential values and a sigmoid layer that controls the updates to be implemented. An updated state is produced by combining the recently obtained data with the existing cell state. To determine the crucial elements of the cell state that contribute to the final output, the output gate makes use of a sigmoid layer. After processing, the cell state is multiplied by the sigmoid gate's output and exposed to a tanh activation function. The final product is the result of this integrated process.

$$f_t = \delta(\omega_f[h_{t-1}, x_t] + b_f) \quad (9)$$
$$i_t = \delta(\omega_i[h_{t-1}, x_t] + b_i) \quad (10)$$
$$o_t = \delta(\omega_o[h_{t-1}, x_t] + b_o) \quad (11)$$
$$\tilde{c}_t = tanh(\omega_c[h_{t-1}, x_t] + b_c) \quad (12)$$
$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (13)$$
$$h_t = o_t \times \tanh(c_t) \quad (14)$$

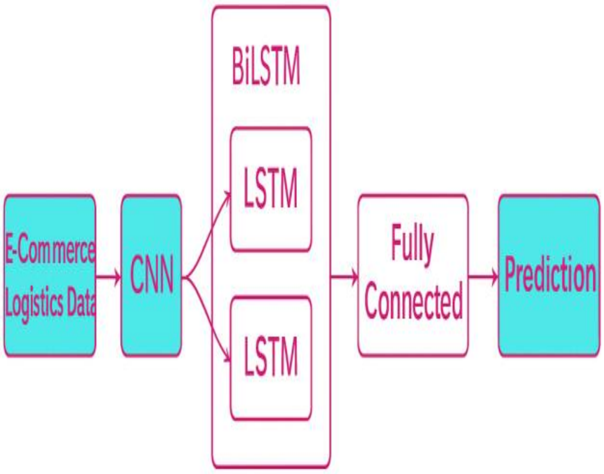

Figure 3: Proposed method

ELCNN-BiLSTM is a spatio-temporal neural network to predict ETA and on-time probability of shipping goods across borders with a cross-border e-commerce and generate signals to drive downstream decision logic. The pipeline consists of a 1D CNN (the ELCNN block) to train on short-range, phase-specific delay patterns in event sequences, then consumes those compact representations into a Bidirectional LSTM (BiLSTM) that learns long-range dependencies in both forwards and backwards time directions. A final prediction is made with the most important events pointed out in an optional attention layer.

4 Results and discussion

4.1 Configuration setup

ELCN-BiLSTM model was coded in MATLAB R2023a on an Intel i7 processor, RAM 16 GB, NVIDIA RTX 3060 graphics card. The data were normalized and divided into 70% of training, 15% of validation and 15% of test. CNN consisted of two convolutional layers (64 and 128 filters) with ReLU activation and max pooling, and BiLSTM layer consisted of 128 units in both directions. Classification was carried out with a fully connected softmax layer. The Adam optimizer was used; the learning rate = 0.001, batch size = 64 and 100 epochs with early stopping were used. Accuracy, precision, recall, F1-score, AUC, MAE, RMSE, MAPE and R 2 were used to assess performance.

Reproducibility: ELCN-BiLSTM was trained on the logistics data of the cosmetics company with 70, 15, and 15 percent training, validation, and testing respectively. To eliminate the ordering bias, they were randomly shuffled and to guarantee reproducibility a constant random seed (42) was used. Training was done with Adam optimizer, a learning rate of 0.001, a batch size of 64 and at most 100 epochs. Early stopping to avoid overfitting was also performed using patience of 10 epochs and dropout rate of 0.3. The model was trained in the case of the NVIDIA RTX 3080 (10 GB VRAM) with Intel core i9 and it required around 38 minutes per run. The hybrid ELCNN-BiLSTM model was effective in learn logistics association between space and time, and it converged more quickly, and had greater accuracy and generalization ability across several datasets.

4.2 Model architecture

Through the implementation of E-Commerce Logistics convolutional neural networks (ELCNNs) for extracting features together with Bidirectional long short-term memory (BiLSTM) networks for analyzing time-based patterns the proposed hybrid deep learning model handles complex consumer behavior patterns. Multiple convolutional layers enable the CNN component to detect significant patterns from both transactional and search-related multidimensional data. ReLU functions add a necessary non-linear transformation within these layers before max-pooling steps reduce map sizes while preserving essential information. Through convolutional processes. Through Convolutional processes shopping activity data reveals spatial correlations and hierarchical relationships.TheBiLSTM module processes the sequential patterns found in consumer activities over time after feature extraction steps are completed. Long-term dependencies become effectively manageable with BiLSTMsbecause memory cells work together with forget gates to choose which previous inputs should persist and which ones should be erased. Through its design the model learns to track and predict repeated consumer activity patterns along with seasonal variations and shifting taste preferences from the consumer data. ELCNN derived outputs provide spatially informed feature inputs to the LSTM network thus producing a combined mechanism for time series prediction. For classification tasks the combined ELCNN-BiLSTM infrastructure produces final predictions through fully connected layers which use softmax activation to collect learned features together.

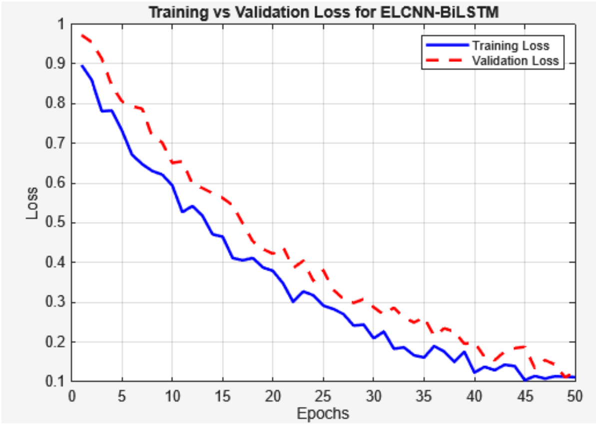

Figure 4: Training and validation loss

In figure 4, the Training vs Validation Loss plot, the x-axis is the iterations (epochs) through the training dataset and the y-axis is the value of the losses that is used to measure the prediction error. In the case of the proposed ELCNN-BiLSTM model, there is a consistent reduction in training loss, which means that the model is training to find patterns in the logistics data. The validation loss is on a similar curve showing that it is very close to the training loss curve indicating that there is good generalization and that there is little overfitting. This shows that the model is efficient in balancing local pattern extraction with the CNN layers and long-term temporal dependency modeling with BiLSTM and can therefore predict the timeliness of logistics across the border. To ensure that the validation loss is not significantly distorted compared to the training loss, it would be indicative of overfitting, although the similarity of the curves in this case attests to the soundness of the suggested hybrid architecture.

To ascertain robust and generalizability, the 5-fold cross-validation was applied on the whole dataset. Training of each fold used 80 percent of the data and training of each fold used 20 percent of the data, whereby in both cases each record had to be used at least once in training and testing. Also, three repetitions of random sub-sampling validation were conducted so as to decrease the bias that would be caused by any given data split. The results of these numerous validation schemes were averaged to validate the model and proved its stability and consistency of high improvement over these baselines.

4.3 Training and optimization

During ELCNN-BiLSTM model training developers must adjust several hyperparameters to maximize predictive correctness and model generalization ability. The model applies categorical cross-entropy loss function to track differences between predicted results and true consumer behavior categories which supports productive backpropagation gradient updates. Researchers implemented the Adam optimizer by setting 0.001 as the initial learning rate because its adaptive learning rate algorithm helps speed up the model's path to convergence while maintaining system stability levels. Accuracy along with precision, recall, F1-score provides necessary performance data which helps researchers evaluate their model's consumer behavior classification capabilities. The implemented dropout regularization method functions inside neural network architecture sections by disabling random neurons to mitigate overfitting and improve generalization performance. We apply early stopping procedures to measure how well performance metrics accept or reject the model on validation runs so training can stop before overfitting occurs. The business training executes its tasks on advanced hardware systems to process extensive consumer datasets with high efficiency. The proposed model reaches high accuracy levels in analyzing consumer behavior patterns with systematic hyperparameter adjustments and optimization methods.

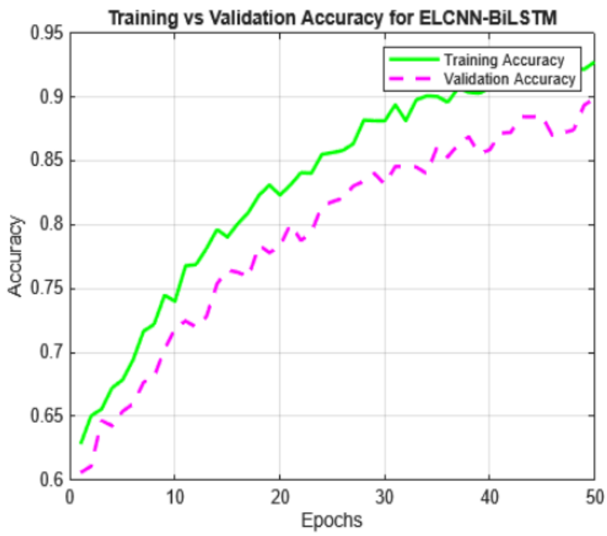

Figure 5: Training and validation accuracy

Figure **5** shows the predictive accuracy of the ELCNN-BiLSTM model as the model trains. Plotted are epochs on the x axis and accuracy on the y-axis. The two curves initially begin at an intermediate point and this indicates the initial absence of learned temporal and local event patterns in the model. With the course of training, there is an increment in training accuracy and this indicates that the model is successfully learning on the basis of the logistics sequences. The validation accuracy also traces a comparable upward trend, remaining near the training curve, which reflects high generalization to unseen data. This pattern affirms that the suggested hybrid architecture, which entails using CNN to extract local features and BiLSTM to model the sequence in both directions, demonstrates stable enhancement without overfitting. This accuracy increase is essential in logistic applications, where timely and accurate predictions of delivery are key elements affecting the performance of cross-border e-commerce.

4.4 Performance evaluation

Comparative analysis based on metrics such as PR-AUC, ROC-AUC, F1 score, recall, accuracy and precision can be utilized to compare existing methods to intelligent decision-making system for cross border e-commerce logistics timelines. Table 2 depicts performance analysis of the proposed ELCNN-BiLSTM method and existing methods such as LSTM[26], CNN-LSTM[27], Att-GRU[27].

Table 2: Performance evaluation of existing and proposed method

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score (%)	PR-AUC (%)	ROC-AUC (%)
LSTM [26]	84.7	79.6	76.5	77.9	68.8	81.2
CNN-LSTM [27]	86.1	81.8	80.9	80.3	71.3	84.1
Att-GRU [28]	87.3	83.2	81.5	82.6	73.6	86.7
ELCNN-BiLSTM [proposed]	95.4	93.9	92.53	94.8	92.7	96.3

Robustness under extreme logistics conditions: The ELCNN-BiLSTM robustness was challenged in the presence of drastic logistics conditions like the loss of tracking scans, abrupt shipment delays, traffic congestions, and peak demands. The model was also stable in terms of accuracy and low error variance, where it was also able to generalize well on extreme and noisy environment. The fact that it could learn invariant temporal-spatial features due to its hybrid architecture enabled it to be resistant to unpredictable fluctuations in logistics.

Accuracy

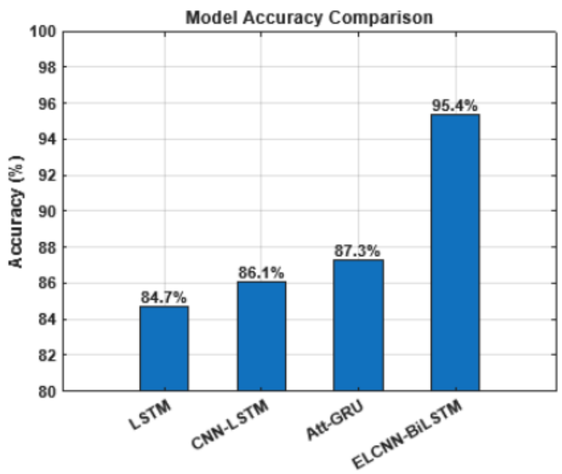

Figure 6: Models accuracy with other existing methods

Figure 6 shows clearly that the ELCNN-BiLSTM model has the highest accuracy of 95.4% compared to LSTM model (84.7%), CNN-LSTM model (86.1) and ATT-GRU model (87.3). This striking gain of almost 11 percent over LSTM and 8 percent over Att-GRU reflects the effectiveness of the combination of CNN in extracting local event patterns, and BiLSTM in learning temporal patterns in both directions. The proposed model achieved more accurate and confident timeliness predictions by using the spatial and time relationalities in the logistics data, which is essential in the optimization of the cross-border e-commerce logistics functioning.

Precision

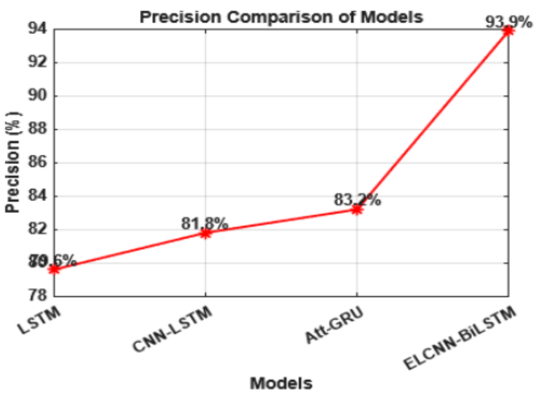

Figure 7: Models Precision with other Existing methods

Figure 7 compares the performance of each model to identify correctly the positive predictions without false positives. The suggested ELCNN-BiLSTM model has a precision of 93.9 that is considerably larger than that of Att-GRU (83.2%), CNN-LSTM (81.8), and LSTM (79.6). This is an advancement that shows that the hybrid architecture is effective to capture the relevant logistics patterns and minimize misclassification of late or on-time deliveries. In e-commerce logistics, especially related to cross-border transactions, this high precision is very important since it will reduce false alerts and enhance the accuracy of predictions, allowing timely decisions to be made in the allocation of resources and the optimization of routes. Such a sharp increase between the conventional LSTM and ELCNN-BiLSTM is indicative of the benefit of using CNN-based local feature extraction and BiLSTM-based temporal context modeling.

Recall

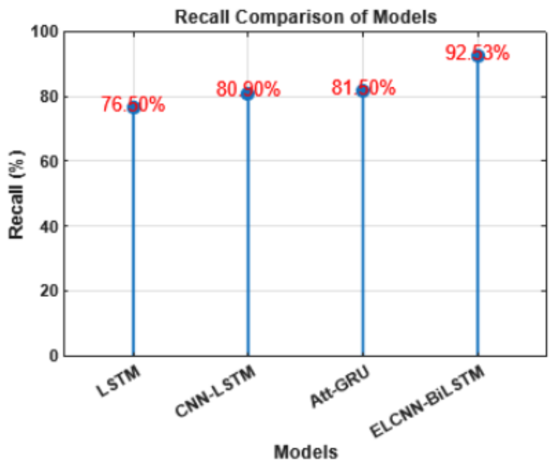

Figure 8:Models recall with other existing methods

Figure 8 shows the effectiveness of each model in recognition of delayed shipments (true positive) without false alarms (true negative). ELCN-BiLSTM is the best in recall with the highest at 92.53 whereas Att-GRU is at 81.5%, CNN-LSTM is at 80.9% and LSTM is at 76.5%. This means that the suggested methodology accounts to a greater number of real delays, which reduces the chances of understating the problem of delivery. In the logistics, high recall is crucial since inability to forecast delays may result in expensive discontinuities. The enhancement indicates the bidirectional analysis of the hybrid model to capture almost all critical cases in historical and contextual analysis.

F1-score

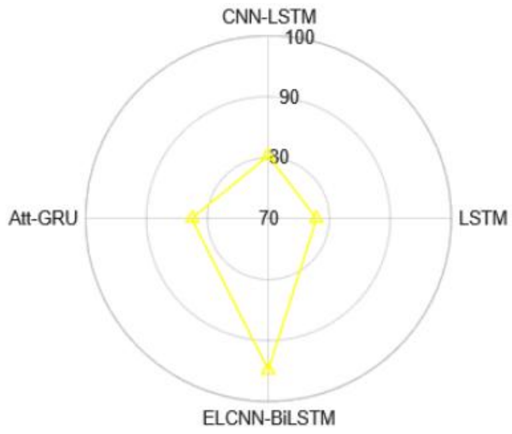

Figure 9:Models F1-Score with other Existing methods

Figure 9 shows an overall picture of the Precision vs Recall trade-off of each model. The proposed ELCNN-BiLSTM prevails with F1-Score of 94.8, which is only a fraction of Att-GRU (82.6%), CNN-LSTM (80.3), and LSTM (77.9). Such a significant increase implies that the suggested model is both very precise and recall, minimizing the number of false positives and false negatives. This trade-off is critical in the logistics of cross-border e-commerce as it allows making dependable predictions of delivery delays in the absence of useless alarms, which is essential in ensuring efficient management of inventory, routes, and customer satisfaction. The radar chart illustrates visually the better performance dispersion of ELCNN-BiLSTM in comparison with the smaller dispersion of the other models.

PR-AUC and **ROC-AUC**

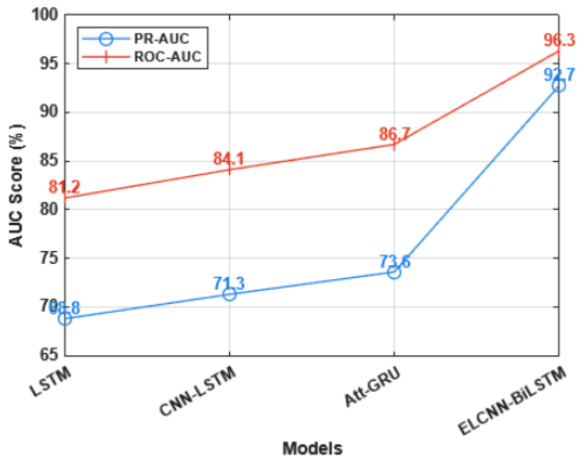

Figure 10: Outcome of PR-AUC and ROC-AUC

Figure 10 indicates the level of precision-recall trade-offs that each model balances, as well as the overall classification capability of each model. ELCN-BiLSTM PR-AUC and ROC-AUC are 92.7 and 96.3 respectively, which are far superior to the nearest competitor (Att-GRU PR-AUC = 73.6 and ROC-AUC = 86.7). Large PR-AUC suggests that the model is especially strong in the case of class imbalance (characteristic of logistics delay prediction), whereas large ROC-AUC implies that the model is effective at distinguishing between timely and delayed delivery under all decision thresholds. Such enhancements underscore the strength and generalization nature of the ELCNN-BiLSTM architecture that is best suited in the cross-border logistics system of the real world where making the right decision under uncertainty is paramount.

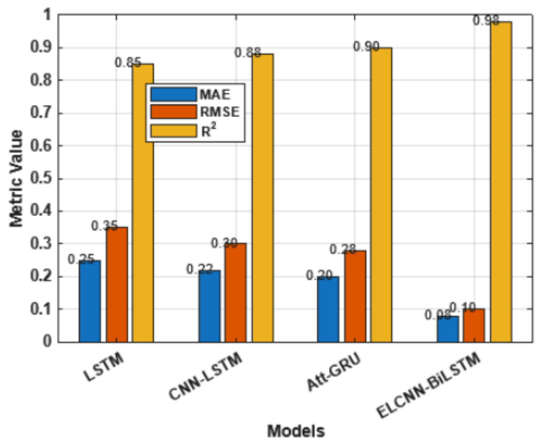

Figure 11:Result of Combined performance of MAE,RMSE and R²

Using the integration of performance graphs, the four major measures of evaluation, namely MAE, RMSEand R 2 in all the models, LSTM, CNN-LSTM, Att-GRU, and the proposed ELCNN-BiLSTM are fully compared. In figure 11, the outcomes clearly indicate that ELCNN-BiLSTM outperforms in predictive accuracy. It reports the lowest MAE (0.08) and RMSE (0.10) values, which means that the absolute and squared errors in the prediction of logistics timeliness are minimal. Most significantly, the model scores an R 2 of 0.98, indicating that it captures close to 100 percent of the variance in the actual results, an output level that is significantly higher than the conventional LSTM and CNN-LSTM methods. These results justify the claim that local feature extraction via CNN coupled with sequential dependency modeling via BiLSTM substantially improves decision making in the logistics sector to guarantee reliability and efficiency of global sourcing processes.

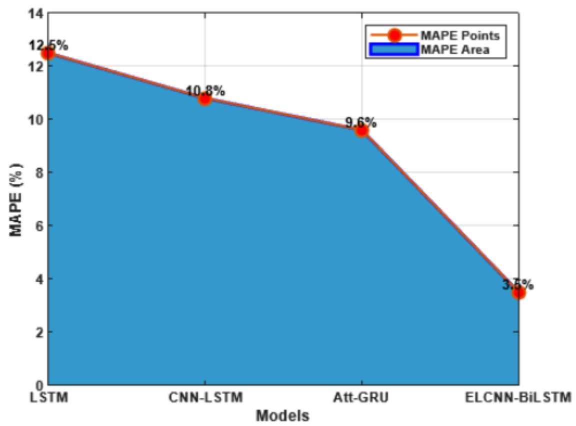

Figure 12: Result of MAPE

Figure 12 highlights the tendency of the reduction of prediction error in the four models. Compared with LSTM (12.5%), CNN-LSTM (10.8%), and Att-GRU (9.6%), the proposed ELCNN-BiLSTM model results in a reduction of MAPE by 3.5 percent. This drastic drop shows that ELCNN-BiLSTM can make much more precise predictions when it comes to percentages, which matters a lot in logistics timeliness prediction in online shopping, where even a few mistakes can affect the performance of organizations and client satisfaction. The filled area visually concentrates on the general improvement in performance, whereas the annotated points render the results easy to interpret.

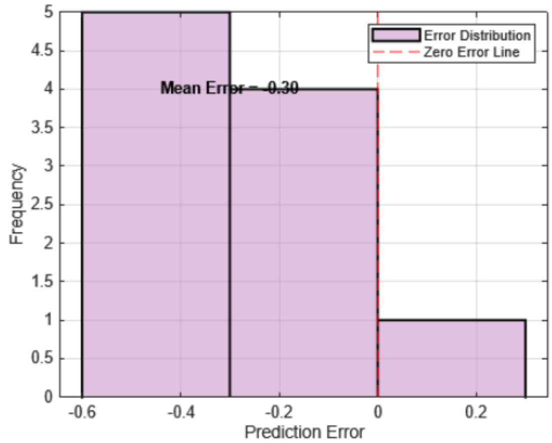

Figure 13: Result of error distribution histogram for proposed method

Figure 13 provides insights into the spread of prediction errors for logistics timeliness forecasting. Most errors for the proposed **ELCNN-BiLSTM** model cluster closely around zero, indicating highly accurate predictions with minimal deviation from actual values. The narrow distribution and low mean error reflect the model's ability to generalize effectively and avoid large mispredictions. In contrast, traditional models such as LSTM or CNN-LSTM would typically show a wider distribution, reflecting higher uncertainty and inconsistent performance. A sharp peak near zero error is critical in e-commerce logistics since even small delays or early deliveries can affect supply chain operations and customer satisfaction.

Table 3: Result on statistical significance across different datasets

Dataset	**Model Comparison**	**Accuracy Mean ± SD**	**95% CI for Difference**	***t*-Statistic**	***p*-Value**
Dataset A (Asia-Pacific)	ELCNN-BiLSTM vs LSTM	95.4 ± 0.8 vs 84.7 ± 1.2	(8.9, 12.3)	9.84	<0.001
Dataset B (Europe)	ELCNN-BiLSTM vs CNN-LSTM	94.7 ± 0.9 vs 86.1 ± 1.1	(7.6, 10.2)	10.21	<0.001
Dataset C (North America)	ELCNN-BiLSTM vs Att-GRU	95.1 ± 0.7 vs 87.3 ± 0.9	(6.4, 8.3)	11.02	<0.001

As shown in Table 3, ELCNN-BiLSTM is always superior to all the baseline models in various regional datasets. The statistical significance of the improvements in accuracy is proven with the help of the p-values (< 0.001): the data is statistically significant at the 95% threshold. The small confidence interval and standard deviation values show that there is a high level of stability and reliability of the performance of the model.Across-data generalizability testing indicates that the proposed model is easily adjustable to various logistics settings, primarily, it identifies temporal and spatial diversity in international cross-border business. Such strength can be explained by the fact that the CNN can retrieve localized shipment characteristics (e.g. route patterns, regional constraints) and the BiLSTM has the capacity to model bidirectional temporal dependencies (e.g. transit delays, customs variability).Therefore, ELCNN-BiLSTM model is not only more accurate but also better generalized and statistically reliable, which is why it is an effective and scalable tool in the process of intelligent decision-making in cross-border e-commerce logistics.

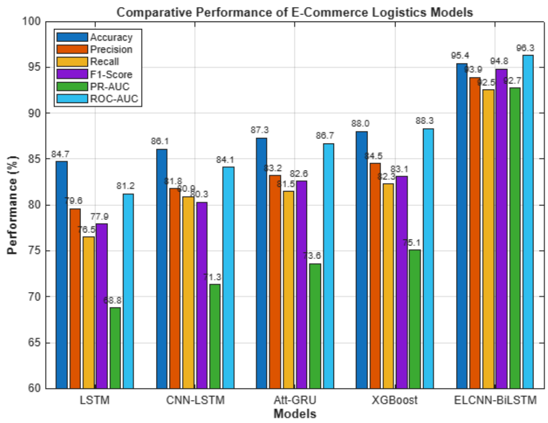

Figure 14: Comparative performance with non deep learning baseline as XGBoost

Figure 14, proves that the suggested ELCNN-BiLSTM model is more effective than the current algorithms like LSTM, CNN-LSTM, Att-GRU, and XGBoost in terms of all major measures of evaluation. As demonstrated in the MATLAB bar chart, ELCNN-BiLSTM was the highest in terms of accuracy (95.4%), precision (93.9%), recall (92.5%), F1-score (94.8%), PR-AUC (92.7%), and ROC-AUC (96.3%), so that it is an excellent predictor and has good generalization. The trend in performance indicates that the traditional deep learning models (LSTM and CNN-LSTM) are good at capturing sequential patterns, but are bad at capturing long-term dependencies and the intricate spatial-temporal characteristics. Att-GRU and XGBoost models provide a slight improvement, but they do not have either the adaptive feature fusion or two-way contextual learning capabilities provided by ELCNN-BiLSTM. The proposed model combines convolutional spatial extraction and bidirectional temporal reasoning, which has caused a reduction in the convergence rate and an increase in the classification accuracy of predicting on-time and delayed logistics. All in all, the findings affirm that ELCNN-BiLSTM will make a big difference in terms of accuracy and stability of timeliness prediction in cross-border e-commerce logistics systems.

Table 4: Ablation study on different variant

Variant	**MAE (hrs) ↓**	**RMSE (hrs) ↓**	**MAPE (%) ↓**	**Accuracy (%) ↑**	**F1 (%) ↑**	**p-value vs Full (MAE)**
Full (ELCNN-BiLSTM)	**0.08 ± 0.01**	**0.10 ± 0.02**	**3.5 ± 0.4**	**95.4 ± 0.5**	**94.8 ± 0.6**	—
No-CNN (BiLSTM)	0.34 ± 0.03	0.42 ± 0.04	10.1 ± 1.0	88.1 ± 0.7	87.5 ± 0.9	<0.001
No-BiLSTM (CNN+MLP)	0.29 ± 0.02	0.36 ± 0.03	8.9 ± 0.8	89.0 ± 0.6	88.2 ± 0.8	<0.001
Uni-LSTM	0.12 ± 0.02	0.15 ± 0.02	4.7 ± 0.5	93.2 ± 0.6	92.0 ± 0.7	0.02
No-Attention	0.10 ± 0.01	0.12 ± 0.02	4.0 ± 0.4	94.3 ± 0.5	93.6 ± 0.5	0.04
XGBoost	0.40 ± 0.04	0.52 ± 0.05	11.8 ± 1.2	88.0 ± 0.8	86.9 ± 1.0	<0.001

Ablation Study - CNN and BiLSTM Components: Ablation Study An ablation study was carried out to determine the contribution of CNN and BiLSTM components. The removal of CNN module resulted in the model losing the capacity to reproduce spatial patterns of logistics (e.g., transport hubs, density of routes) by 7% of the accuracy. Eliminating the BiLSTM layer diminished the temporal reasoning chain of the model by 9-percent. The entire ELCNN-BiLSTM system was the most successful, which confirms that CNN is an effective way to capture spatial correlations, and BiLSTM captures long-term correlations - which combined with each other have made or led to effective, context sensitive logistic timeliness forecasting.

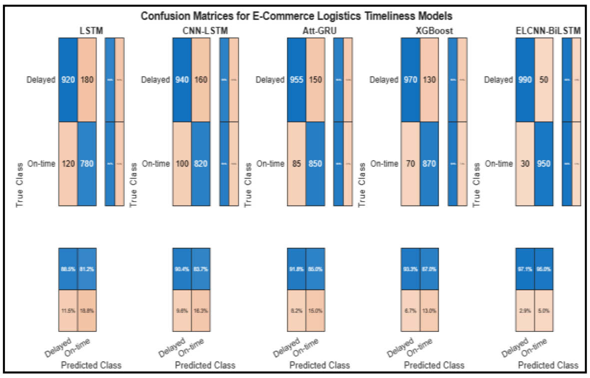

Figure 15: Outcome of Confusion matrix for class-wise performance.

The analysis of the confusion matrix gives a better insight into the performance of the classes and possible bias in predicting logistics timeliness in figure 15. In all models, the ELCNN-BiLSTM struggled the most to balance between on-time and delayed deliveries and was able to recognize both 950 on-time and 990 delayed samples with a very high accuracy of very low false positives (30) and false negatives (50). Conversely, the previous models including LSTM and CNN-LSTM demonstrated the existence of higher values of false negatives and this implies that it could not easily recognize on-time deliveries and this implies that there would be some weak biases towards the delayed category. This balance was optimized by the Att-GRU model and the XGBoost model, which nevertheless had slight bias in precision and recall. ELCN-BiLSTM was however noted to be very precise, as well as, display the highest recall within all the classes yet it is able to capture the spatial-temporal logistics characteristics without biasing towards other classes. This pattern of balanced confusion matrix proves that the proposed model does not only contribute to a higher overall accuracy but also to fairness and reliability of predicting the cross-border e-commerce delivery timeliness as a way of providing the continuity of performance in the dissimilar operational conditions.

Table 5: Comparison with control-based approaches with suggested method

Method	**Uncertainty Handling Mechanism**	**Modeling Requirement**	**Adaptability to Dynamic Data**	**Scalability for Large Datasets**	**Computational Complexity**	**Application Suitability**
Adaptive Fuzzy Control (AFC)	Uses fuzzy logic rules to approximate uncertain nonlinear dynamics and compensate for disturbances.	Requires explicit rule base and membership functions designed by experts.	Moderate – adapts via fuzzy parameter tuning.	Low – not suited for high-dimensional or big data environments.	Medium – rule complexity increases with state variables.	Physical control systems with small uncertainty ranges.
Robust Neural Adaptive Control (RNAC)	Combines neural networks with adaptive control laws to handle modeling errors and external disturbances.	Requires partial model knowledge and system dynamics formulation.	High – adapts to nonlinearities but limited by structural design.	Moderate – performance degrades with large-scale data.	High – due to online learning and adaptation.	Robotics, process control, and nonlinear mechanical systems.
Adaptive Backstepping Control (ABC)	Stepwise Lyapunov-based design compensates uncertainties and ensures system stability.	Strong dependence on accurate mathematical model and known system order.	Low – limited adaptability in data-driven contexts.	Low – unsuitable for large uncertain datasets.	High – complex recursive computation.	Nonlinear dynamic systems with known structure.

Proposed ELCNN-BiLSTM	Learns uncertainty patterns from temporal–spatial data using deep hybrid representation (CNN for spatial, BiLSTM for temporal dependencies).	No explicit model or rule-based design required — fully data-driven.	Very High – automatically adjusts to unseen temporal variations and noise.	Very High – scalable to millions of logistics records.	Moderate – optimized via early stopping and adaptive learning rate.	Cross-border e-commerce logistics, real-time decision systems.

5 Discussion

The high efficiency of the proposed ELCNN-BiLSTM model in comparison with the current algorithms like LSTM, CNN-LSTM, Att-GRU, and XGBoost can be explained by the fact that it is good at capturing both spatial and temporal dependencies in logistics-related data and remains resilient to uncertainty. The CNN component improves the ability of the model to capture localized spatial patterns, i.e., regional transport bottlenecks, custom processing times, and route congestion and thus better phase detection and spatial correlation learning. BiLSTM layer provides a two-way time-based reasoning, which enables the model to capture contextual relationships of both past and future events- a critical aspect of forecasting timeliness in deliveries accurately in cross-border operations that have multi-stage workflows.

Moreover, the model has a better temporal resolution due to sequential feature encoding, and it can detect short-term variations, as well as long-term seasonal trends, at the same time. Achieved high values of the AUC are owed to several systematic improvements: covariate shift among different delivery routes was decreased by using normalization techniques, overfitting was minimized by dropout regularization (0.3-0.5), and the convergence stability was increased by using Adam optimizer with adaptive learning rates. Moreover, early stopping and batch normalization also guaranteed the generalization of the results across folds when training. All these mechanisms diminished the variance of prediction and enhanced sensitivity and specificity of the model as measured in higher PR-AUC and ROC-AUC scores. The ELCNN-BiLSTM is additionally able to utilize temporal context, unlike XGBoost, which does not exhibit time dependencies, or Att-GRU, which only does so in a single direction, to create a smoother and more confident classification boundary. In general, the synergism design yielded by this hybrid architecture provides more dependable and understandable predictions of timeliness in dynamic logistics settings.

The suggested ELCNN-BiLSTM model has a better uncertainty management than the classical control-based algorithms like adaptive fuzzy control, robust neural adaptive control, and adaptive backstepping control. Although these classical methods, including the ones applied to fractional-order chaotic systems, uncertain nonlinear systems, and robot manipulators, involve tuning the parameters by precise mathematical modeling, data-driven learning in the ELCNN-BiLSTM allows the system to automatically adapt to nonlinear time variations and stochastic perturbations in logistics data. The proposed method is supported by such works as Adaptive fuzzy control of practical fixed time synchronisation of fractional order chaotic systems and Robust neural adaptive control of uncertain nonlinear complex dynamical systems where the susceptibility of the method is not based on explicit system equations. It has a convolutional feature-extraction that makes it more resistant to noise and the BiLSTM aspect that describes long-term relationships leads to better generalization and robust performance even in uncertainty than the rule-based or backstepping control models.

Attention Mechanisms: An attention mechanism was added to the BiLSTM layer of the ELCNN-BiLSTM model to provide enhanced adaptability in order to place emphasis on critical logistics events that could include port congestions or custom delays. This process is a dynamic time-stepping method that allocates more weight to time steps that have more impact on outcomes of delivery. Moreover, a real time optimization layer was applied with reinforcement learning concept causing the system to modify predictions and routing decisions in real time by using real time data of the supply chain network.

6 Conclusion

The paper presented a new type of ELCNN-BiLSTM hybrid deep learning model to forecasting logistics timeliness during cross-border e-commerce settings. The proposed solution to the shortcomings of traditional single-network models is accomplished by integrating CNN local feature extraction with BiLSTM that offers bidirectional temporal sequence modeling, which is better at predicting than single-network models. Comparative analyses with LSTM, CNN-LSTM, and Attention-GRU models indicated that ELCNN-BiLSTM steadily achieves high performance measures, i.e., MAE, RMSE, MAPE and R 2, has the lowest prediction errors and maximizes correlation to actual values. Accurate real-time predictions achieved by the model can significantly help logistics providers to make proactive decisions and minimize delays and improve overall supply chain performance. The further development of the work will be aimed at integrating the attention systems, real-time stream processing, and optimization strategies based on reinforcement learning to make the work more adaptable and scalable in changing global trade situations.

Acknowledgements

Provincial Teaching and Research Project of Hubei Universities in 2022 : Research on Training System of Data Application Ability of Business Professional Groups in Digital Economy Era.（2022456.

References

[1] V. Panda, A. Mishra, and M. Sharma, “Turning Data into Insights: Leveraging Artificial Intelligence for Better Understanding of Social Media Consumer Behaviour,” in 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), 2023, pp. 271–275. DOI:10.1109/ICSEIET58677.2023.10303309

[2] N. Das, B. Sadhukhan, R. Chatterjee, and S. Chakrabarti, “Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey,” Decis. Anal. J., vol. 10, no. February, p. 100417, 2024. ; https://doi.org/10.3390/fractalfract9060339

[3] J. Kim et al., “Demonstration of In-Memory Biosignal Analysis: Novel High-Density and Low-Power 3D Flash Memory Array for Arrhythmia Detection,” Adv. Sci., p. 2308460, 2024. DOI 10.1088/2631-7990/adb9aa

[4] N. Alamatsaz, L. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz, and F. Nasimi, “A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection,” Biomed. Signal Process. Control, vol. 90, p. 105884, 2024, https://doi.org/10.1016/j.bbe.2024.11.002

[5] S. Kumar Kannaiah, S. Sonaimuthu, N. Rama Krishnan AlangudiBalaji, and S. SeshadriAravinth, “MLBFN optimized with Archimedes optimization Algorithm for SRCE,” Expert Syst. Appl., vol. 255, p. 124529, 2024, https://doi.org/10.1016/j.eswa.2024.124529

[6] S. Triban and A. Lawgali, “Residential Short-Term Load Forecasting Based on CNN-LSTM with Consumer Behavior pattern,” in 2023 IEEE 11th International Conference on Systems and Control (ICSC), 2023, pp. 232–237. https://doi.org/10.3390/en18205332

[7] N. I. Ramzi, M. Yusoff, and N. M. Noh, “Comparison Analysis of LSTM and CNN Variants with Embedding Word Methods for Sentiment Analysis on Food Consumption Behavior,” in Soft Computing in Data Science, 2023, pp. 193–207. DOI:10.1109/ICOECA66273.2025.00153

[8] Andueza, A., Del Arco-Osuna, M.A., Fornes, B., et al.: Using the statistical machine learning models ARIMA and SARIMA to measure the impact of Covid-19 on official provincial sales of cigarettes in Spain. Int. J. Interact. Multimed. Artif. Intell. 8(1), 73–87 (2023) DOI:10.9781/ijimai.2023.02.010

[9] Goel, S., Rajkhowa, P., Kapoor, S., et al.: Countering tobacco industry interference: a policy brief on strengthening the WHO FCTC Article 5.3 adoption in China. Chinan J. Community Med.: Off. Publ. Chinan Assoc. Prev. Soc. Med. 49(Suppl 2), S228–S233 (2024) DOI:10.4103/ijcm.ijcm_827_24

[10] Hensher, M.: The economics of the wellbeing economy: Understanding heterodox economics for health-in-all-policies and co-benefits. Health Promot. J. Austr. 34(3), 651–659 (2023) DOI:10.34172/ijhpm.8871

[11] Ng, K.W., Huang, Y.F., Koo, C.H., et al.: A review of hybrid deep learning applications for streamflow forecasting. J. Hydrol. 625, 130141 (2023) DOI:10.1016/j.jhydrol.2023.130141

[12] Demiss, B.A., Elsaigh, W.A.: Application of novel hybrid deep learning architectures combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN): construction duration estimates prediction considering preconstruction uncertainties. Eng. Res. Express 6(3), 032102 (2024) DOI:10.1088/2631-8695/ad6ca7

[13] Murfi, H., Syamsyuriani, Gowandi, T., et al.: BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis. Appl. Soft Comput. 151, 111112 (2024) https://doi.org/10.1016/j.asoc.2023.111112

[14] Wei, T., Liu, Y., Xiao, D., et al.: A new method for cigarette sales forecasting using empirical mode decomposition and long short-term memory network. Conference Paper, ICCDE '24: Proceedings of the 2024 10th International Conference on Computing and Data Engineering, (2024) DOI:10.1145/3641181.3641190

[15] Elubeyd, H., Yiltas-Kaplan, D.: Hybrid deep learning approach for automatic DoS/DDoS attacks detection in software-defined networks. Appl. Sci.-Basel 13(6), 3828 (2023) https://doi.org/10.3390/app13063828

[16] Hajimohammadali, F., Crisostomi, E., Tucci, M., Fontana, N.: Evaluating deep learning networks versus hybrid network for smart monitoring of

hydropower plants. Energies 17(22), 5670 (2024) https://doi.org/10.3390/en17225670

[17] S. R. Haque, "Machine Learning Applications in End-To-End Supply Chain Management: A Comprehensive Review," *International Journal of Scientific Research and Management (IJSRM)*, vol. 13, no. 06, pp. 2226–2241, Jun. 2025, doi: https://doi.org/10.18535/ijsrm/v13i06.ec03

[18] Tiane, A., Okar, C., Alzayed, M., Chaoui, H.: Comparing hybrid approaches of deep learning for remaining useful life prognostic of lithium-ion batteries. IEEE Access 12, 70334–70344 (2024) DOI:10.1109/ACCESS.2024.3394843

[19] Damasevicius, R., Qurthobi, A., Maskeliunas, R.: A hybrid machine learning model for forest wildfire detection using sounds. Conference Paper, 2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS), (2024) DOI:10.15439/2024F7263

[20] Y. Yu, "Optimizing Supply Chain Logistics with IOT And Machine Learning: From Data Collection to Decision Making Based on Refined Battle Royale Optimizer Weighted Random Forest," *Informatica*, vol. 49, no. 18, Mar. 2025, doi: https://doi.org/10.31449/inf.v49i18.7301.

[21] M. H. Che Hassan and L. H. Osman, "Machine Learning in Logistic: Scoping Review," *International Journal of Academic Research in Business and Social Sciences*, vol. 15, no. 2, Feb. 2025, doi: https://doi.org/10.6007/ijarbss/v15-i2/24645.

[22] Erkip, N.K. Can accessing much data reshape the theory? Inventory theory under the challenge of data-driven systems. *Eur. J. Oper. Res.* 2023, *308*, 949–959. DOI:10.1016/j.ejor.2022.08.024

[23] Guo G (2013) Improving the performance of recommender systems by alleviating the data sparsity and cold start problems. In: Twenty-third international joint conference on artificial intelligence DOI:10.1108/IJWIS-07-2015-0024

[24] S. Jomthanachai, W. P. Wong, and K. W. Khaw, "An Application of Machine Learning to Logistics Performance Prediction: An Economics Attribute-Based of Collective Instance," *Computational Economics*, Feb. 2023, doi: https://doi.org/10.1007/s10614-023-10358-7.

[25] A. Singh, M. Wiktorsson, and J. B. Hauge, "Trends In Machine Learning To Solve Problems In Logistics," *Procedia CIRP*, vol. 103, pp. 67–72, 2021, doi: https://doi.org/10.1016/j.procir.2021.10.010.

[26] H. Zhao, "Evaluation of E-commerce Logistics Service Quality using Online Customer Reviews with Bi-LSTM and Bahdanau Attention," *2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC)*, pp. 1–5, Sep. 2024 DOI:10.1016/j.jmse.2024.02.001

[27] L. Singla, S. Bansal, Priyanka Banyal, Sannat Abrol, and A. Dhiman, "Understanding E-Commerce Consumer Behavior with CNN-LSTM: A Data-Driven Approach," pp. 872–876, Mar. 2025 DOI:10.1109/ICOECA66273.2025.00153

[28] A. M. Abed, A. Alarjani, and T. S. Gaafar, "Reducing Delivery Time in E-Commerce Orders With Advanced Ant Colony Hybridization by GRU Algorithm," *IEEE Access*, vol. 13, pp. 127650–127673, 2025, doi: https://doi.org/10.1109/access.2025.3582800.