

Article type:
Original Research

Article history:
Received 10 June 2026
Revised 28 July 2026
Accepted 22 September 2026
Initial Publish 27 September 2026
Published online 01 July 2027

Mehdi. Sadeghi Moein¹, Mahmoud.
Mohammadi^{1*}, Mehرداد.
Allahgholizadeh Azari¹

¹ Department of Industrial Management, CT.C.,
Islamic Azad University, Tehran, Iran

Corresponding author email address:
mah.mohammadi50@iau.ac.ir

How to cite this article:
Sadeghi Moein, M., Mohammadi, M., &
Allahgholizadeh Azari, M. (2027). Enhanced
Bidirectional LSTM (BiLSTM) Model with an
Attention Mechanism for Demand Forecasting in
Manufacturing Systems: A Case Study of the
Polymer Industry. *Future of Work and Digital
Management Journal*, 5(5), 1-19.
<https://doi.org/10.61838/fwdmj.355>



© 2027 the authors. This is an open access article
under the terms of the Creative Commons
Attribution-NonCommercial 4.0 International (CC
BY-NC 4.0) License.

Enhanced Bidirectional LSTM (BiLSTM) Model with an Attention Mechanism for Demand Forecasting in Manufacturing Systems: A Case Study of the Polymer Industry

ABSTRACT

Accurate demand forecasting is a fundamental prerequisite for production planning, inventory control, and operational decision-making in manufacturing industries. In production systems, demand fluctuations can lead to reduced resource productivity, increased inventory levels, higher operating costs, and lower customer service levels. Although deep learning-based models have advanced considerably in recent years, many of them still face limitations in simultaneously capturing long-term temporal dependencies and focusing on the most informative patterns within time-series data. This study proposes a BiLSTM-Attention neural network model for demand forecasting of products in a polyester film (BOPET) manufacturing company. Historical demand data were preprocessed, normalized, and structured into temporal sequences for model training. Model performance was evaluated using a walk-forward validation scheme and the MAE, RMSE, and Safe MAPE metrics, and was compared with the forecasting method currently used in the plant. The results indicate that the proposed model effectively learns nonlinear relationships, temporal dependencies, and fluctuating demand patterns, and substantially reduces forecasting error relative to the existing method. Moreover, the model's forecasts can serve as reliable inputs for make-to-stock (MTS) production planning and other operational decisions. The findings demonstrate that the BiLSTM-Attention architecture not only improves demand forecasting accuracy but also offers practical applicability in real manufacturing environments and potential generalizability to other industries with time-series data.

Keywords: Demand forecasting; BiLSTM-Attention; Deep learning; Time-series forecasting; Make-to-stock (MTS)

Introduction

Demand forecasting is a foundational component of production and operations management because it provides the informational basis for decisions concerning capacity allocation, inventory replenishment, procurement, workforce utilization, production scheduling, and customer-order fulfillment. In contemporary manufacturing environments, these decisions have become increasingly difficult as firms confront greater product variety, shorter product life cycles, volatile markets, changing customer preferences, supply disruptions, and increasingly complex distribution channels. Forecasting errors can propagate throughout production and supply chain systems, resulting in excess inventory, stockouts, inefficient use of machinery and labor, unnecessary setup activities, higher holding and shortage costs, and deterioration in customer service. These consequences are particularly pronounced in multi-stage manufacturing systems, where production activities are technologically interdependent and where inaccurate estimates of future demand may affect not only the quantity produced but also production sequencing, capacity utilization, raw-material requirements, intermediate inventories, and

downstream delivery performance. Accordingly, improvements in demand forecasting have direct implications for operational efficiency and supply chain responsiveness. Recent research has increasingly emphasized data-driven approaches to demand prediction because machine learning can identify relationships that are difficult to represent using conventional forecasting assumptions [1]. Machine-learning-based demand forecasting has also been extended to integrated omnichannel settings, where clustering and neural computational methods have been employed to accommodate heterogeneous demand behavior and complex purchasing patterns [2]. These developments illustrate the transition of demand forecasting from a predominantly statistical task toward an intelligent decision-support capability embedded within production and supply chain management.

Traditional forecasting approaches, including moving averages, exponential smoothing, autoregressive models, and autoregressive integrated moving average models, remain useful when time series exhibit relatively stable and interpretable structures. However, their effectiveness can decline when demand is characterized by nonlinear dynamics, irregular fluctuations, interactions across temporal scales, or changes that cannot readily be described through fixed parametric relationships. The growing availability of high-frequency operational data has therefore encouraged the adoption of artificial intelligence and machine-learning techniques capable of learning nonlinear patterns directly from historical observations. In supply chain settings, optimized machine-learning and deep-learning approaches have demonstrated substantial potential for modeling demand sequences whose temporal structure is too complicated for conventional techniques [3]. Similarly, intelligent approaches to industrial demand prediction have been proposed as alternatives for environments in which demand is affected by heterogeneous and dynamically evolving conditions [4]. Deep learning is particularly attractive because hierarchical neural representations can extract informative structures from large sequential datasets while reducing the need to prespecify the mathematical relationship between predictors and outputs. Consequently, deep architectures have been increasingly applied across domains involving sequential prediction, including energy, transportation, environmental monitoring, finance, manufacturing, and supply chain management.

Among deep-learning architectures, Long Short-Term Memory (LSTM) networks have emerged as one of the principal methods for time-series forecasting. Unlike conventional recurrent neural networks, LSTM architectures use gating mechanisms and internal memory states to regulate information retention and forgetting, allowing them to capture long-range temporal dependencies while mitigating the vanishing-gradient problem. This characteristic is particularly important in demand forecasting because present demand may depend not only on immediately preceding observations but also on seasonal, cyclical, or delayed historical patterns. Empirical applications have demonstrated the usefulness of LSTM models in order-demand forecasting under short lead times and delivery constraints [5], while LSTM autoencoder structures have been applied simultaneously to forecasting and anomaly detection in supply chain systems [6]. Integrated CNN–LSTM architectures have also produced favorable results for short-term load forecasting by combining local pattern extraction with sequential learning [7]. In smart manufacturing, LSTM-based forecasting has been integrated with kernel density estimation to support cost-effective inventory management and multidimensional forecasting [8]. Other applications have shown that LSTM optimization can improve forecasting in electric-load environments [9], while the combination of LSTM with metaheuristic optimization has been applied successfully to oil-price forecasting [10]. These studies collectively demonstrate that LSTM-based models are sufficiently flexible to address different forms of sequential uncertainty and provide a strong methodological foundation for industrial demand prediction.

Nevertheless, standard LSTM networks have important limitations. A conventional LSTM processes a temporal sequence predominantly in the forward direction, which may restrict the richness of the contextual information extracted from an observed sequence. Bidirectional Long Short-Term Memory (BiLSTM) networks address this problem by processing sequential information through forward and backward recurrent layers and combining the resulting representations. The resulting architecture can capture richer dependencies within an observed historical window and has therefore attracted increasing interest in complex forecasting applications. For example, a hybrid CNN–BiLSTM framework has demonstrated strong performance in store-item demand forecasting by integrating convolutional feature extraction with bidirectional temporal learning [11]. BiLSTM has similarly been employed for atmospheric temperature forecasting, where its capacity to represent temporal dependencies was further strengthened through an attention mechanism [12]. The increasing use of bidirectional recurrent architectures reflects a broader recognition that effective time-series prediction requires models capable of representing multiple temporal relationships rather than relying solely on one-directional propagation. Hybridization has further strengthened this line of research: optimization of LSTM–CNN structures has been investigated specifically for improving time-series forecasting performance [13], and CNN–BiLSTM combinations have been successfully transferred to domains beyond conventional demand prediction, demonstrating the general capacity of hybrid neural architectures to learn complex temporal signals.

An additional development with major implications for forecasting is the attention mechanism. Recurrent models ordinarily transform an input sequence into internal representations without explicitly differentiating the importance of every historical observation to the final forecast. Attention mechanisms address this limitation by learning weights that determine which portions or features of an input sequence should contribute more strongly to the prediction. This capability is especially valuable when demand histories contain periods of varying informational significance, because abrupt increases, repeated seasonal structures, recent fluctuations, or particular historical intervals may provide different levels of predictive information. An attention-based CNN–BiLSTM model has achieved promising results in stock-price prediction by integrating local feature extraction, bidirectional sequence learning, and selective weighting of relevant information [14]. Likewise, a dedicated attention-enhanced BiLSTM architecture for demand forecasting has demonstrated strong generalizability across multiple datasets, supporting the direct relevance of attention-based bidirectional models to supply chain decision-making [15]. These findings are consistent with a broader methodological movement toward forecasting systems that do not treat all historical observations as equally informative but instead learn which temporal components should receive greater emphasis during prediction.

The value of hybrid neural forecasting is also evident in studies combining recurrent networks with convolutional models, optimization algorithms, and parallel learning structures. Hybrid CNN–LSTM modeling has been applied to real detergent-product demand data, demonstrating that jointly modeling feature patterns and temporal dependencies can improve product demand prediction [16]. Parallel LSTM–MLP structures have likewise been developed for mid-term minimum-demand forecasting, enabling sequential patterns to be modeled alongside nonlinear relationships involving explanatory information [17]. Comparative investigations of ARIMA, LSTM, and hybrid ARIMA–LSTM architectures have further indicated that combining statistical and neural representations can improve forecasting when both linear and nonlinear structures are present [18]. Model configuration is equally important. Genetic algorithms have been used to optimize LSTM parameters in flash-flood forecasting, illustrating how data-driven hyperparameter optimization can enhance recurrent-network

performance [19]. Modified genetic algorithms have similarly been integrated with LSTM architectures for air-quality forecasting, providing evidence that evolutionary optimization can increase the effectiveness of deep sequential models [20]. Earlier work using an improved genetic algorithm for LSTM optimization in traffic-flow forecasting also showed that the predictive performance of recurrent architectures can be strengthened through systematic optimization rather than reliance on manually selected configurations [21]. In financial forecasting, hybrid LSTM–CNN architectures assisted by genetic algorithms have likewise been investigated, further illustrating the cross-domain utility of jointly optimizing feature extraction and sequential learning [22].

More recent research has expanded these concepts by integrating BiLSTM and attention with sophisticated feature-extraction mechanisms. CNN–BiLSTM–Attention architectures have been applied to asphalt-pavement performance prediction, demonstrating the capacity of such models to characterize nonlinear degradation patterns and temporal dependencies [23]. A similar structure incorporating feature engineering has been developed for carbon-emission forecasting, where convolutional extraction, bidirectional sequence learning, and attention-based weighting were jointly exploited to improve predictive accuracy [24]. In geophysical applications, CNN–BiLSTM–Attention architectures have been used to predict total organic carbon from logging information, again highlighting their effectiveness in extracting useful patterns from complex sequential data [25]. The architecture has also been enhanced through heuristic optimization for evapotranspiration forecasting under limited meteorological data availability, indicating that optimized hybrid networks can retain predictive capacity even under constrained information conditions [26]. These developments are relevant to manufacturing demand forecasting because industrial demand data similarly contain nonlinearities, temporal dependencies, abrupt fluctuations, and potentially limited observations for individual products.

Attention-enhanced bidirectional learning has additionally been extended to dynamic trajectory and motion forecasting. An adaptive-attention MSC–BiLSTM model has been developed for ship-track prediction, combining multiscale feature extraction with bidirectional temporal representations and dynamically weighted information [27]. Reinforced fusion of CNN, BiLSTM, and attention has also been used for forecasting the motion of floating platforms, demonstrating strong applicability to unstable and highly nonlinear sequential behavior [28]. The relevance of these approaches extends beyond their immediate application domains: manufacturing demand also represents a dynamically evolving sequence whose future states may depend on patterns occurring at multiple temporal scales. In addition, hybrid deep-learning frameworks are increasingly being embedded within larger intelligent industrial systems. A CNN–LSTM architecture integrated with multiobjective evolutionary optimization has been proposed as part of a closed-loop framework for adaptive quality optimization in smart manufacturing [29]. Such research indicates that deep forecasting should not be viewed merely as an isolated prediction exercise; rather, its operational value emerges when predictive outputs are connected to planning, control, optimization, and adaptive decision-making systems.

This transition from isolated forecasting toward operational decision support is particularly significant in supply chains and manufacturing systems. Recent work on multi-node supply chains has integrated Transformer and LSTM architectures to capture both local temporal patterns and long-range dependencies associated with demand fluctuations across interconnected supply chain nodes [30]. In e-commerce, attention-enhanced BiLSTM models combined with Bayesian optimization have been developed for sales forecasting, demonstrating the applicability of optimized bidirectional architectures to inventory and demand-planning problems [31]. Multi-scale attention-based CNN–LSTM models have also

been applied to electricity-load forecasting, with their architecture designed to extract heterogeneous patterns while selectively emphasizing influential temporal information [32]. Together, these studies indicate that contemporary forecasting research is converging on architectures that integrate several complementary capabilities: extraction of local features, representation of long-term dependencies, bidirectional interpretation of temporal structure, adaptive weighting of important information, and systematic optimization of model parameters. Such capabilities are especially relevant to manufacturing demand series, which may exhibit irregular cycles, periods of low and high activity, abrupt transitions, and product-specific temporal behavior.

Despite substantial methodological progress, the transfer of these advances to real or realistically structured manufacturing demand environments remains incomplete. Much of the existing literature evaluates advanced forecasting architectures in energy systems, environmental processes, financial markets, transportation, or benchmark datasets, whereas fewer studies directly address operational demand forecasting in process-manufacturing environments. Even within supply chain research, forecasting performance is frequently evaluated mainly through statistical error metrics, without explicitly connecting model outputs to concrete manufacturing policies. Yet the economic value of demand forecasting does not arise solely from minimizing numerical error; forecasts become operationally useful when they improve decisions involving inventory positioning, production quantities, capacity allocation, and scheduling. This issue is particularly important under make-to-stock (MTS) strategies, in which production decisions are made in anticipation of future demand and inaccurate forecasts can therefore generate substantial overproduction or shortages. Existing studies on product and supply chain forecasting clearly demonstrate the potential of intelligent methods [2, 3], but further research is required to determine how advanced attention-based bidirectional recurrent models perform when applied to demand patterns representative of actual manufacturing systems and when their outputs are intended to support production-planning decisions.

The polyester film industry provides a particularly relevant setting for investigating this problem. Biaxially oriented polyethylene terephthalate (BOPET) manufacturing is characterized by continuous and technologically interdependent processing activities, substantial capital utilization, multiple product specifications, and the need to coordinate output with anticipated market requirements. Under such conditions, erroneous demand estimates may affect machine loading, material consumption, inventory accumulation, production sequencing, and service performance. More accurate forecasts can therefore contribute simultaneously to production stability and more effective MTS planning. The broader literature demonstrates the utility of deep and hybrid forecasting systems across manufacturing, inventory, supply chain, and demand-management settings [8, 16], while recent developments in attention-enhanced architectures provide a methodological basis for improving how demand histories are represented and interpreted. Nevertheless, the effectiveness of a comparatively streamlined BiLSTM–Attention architecture deserves direct evaluation in a realistic industrial setting because highly elaborate hybrid architectures are not always necessary or practical for operational implementation. A model that combines bidirectional temporal learning with explicit attention weighting may provide a favorable balance between predictive capability, computational feasibility, and operational applicability. The manuscript's underlying industrial framework specifically concerns daily product demand in BOPET manufacturing and evaluates forecasts as inputs to MTS production planning, making the connection between forecasting accuracy and manufacturing decision support central rather than incidental.

Accordingly, the research gap addressed in this study lies at the intersection of advanced time-series forecasting and practical manufacturing decision support: although LSTM, CNN–LSTM, BiLSTM, optimization-based recurrent models, and attention-enhanced hybrid networks have demonstrated substantial forecasting potential across diverse application domains, comparatively limited evidence exists regarding the performance and operational usefulness of BiLSTM–Attention forecasting for demand patterns representative of a real multi-stage BOPET manufacturing environment. Addressing this gap is important because manufacturing forecasting systems must not only identify nonlinear temporal relationships but also maintain predictive reliability under sequential validation conditions that resemble actual forecasting practice. Moreover, their outputs should be interpretable as actionable inputs to production planning rather than merely as improvements in statistical model fit. The use of walk-forward validation is particularly pertinent in this respect because it preserves temporal ordering and evaluates forecasts under progressively advancing information sets, thereby approximating the circumstances under which an industrial forecasting model would operate in practice. The study therefore extends the existing literature by situating BiLSTM–Attention forecasting within a concrete manufacturing-planning context and by linking predictive performance to the informational requirements of an MTS production system. Accordingly, the aim of this study is to develop and evaluate a BiLSTM–Attention-based model for forecasting product demand in a polyester film (BOPET) manufacturing environment and to assess its applicability as a reliable forecasting input for make-to-stock (MTS) production planning.

Methodology

This study is applied in nature and was conducted using a data-driven model development approach. The main objective is to develop and evaluate an intelligent model based on a BiLSTM–Attention neural network for forecasting product demand in a real manufacturing environment. The proposed methodological framework includes the stages of data collection and preparation, time-series data preprocessing, construction of the model input structure, neural network training, performance validation, and generation of the final forecasts.

The data used in this study were extracted from the historical demand records of products from a polyester film (BOPET) manufacturing company. The dataset includes daily demand quantities for different products and their prices, and reflects demand behavior comparable to that observed in an industrial setting.

Table 1

Dataset Sample

product_id	product_description	date	year	month	is_shutdown	day	season	demand	sales_price
101	Metallized BOPET film 23 micron – width 1020	1/1/2023	2023	1	1	1	1	0	100000
101	Metallized BOPET film 23 micron – width 1020	1/2/2023	2023	1	1	2	1	0	100000
.	.	.	.	.	.	.	.	.	.

Prior to using the data, a data quality assessment process was carried out, including the identification of missing values, examination of outliers, correction of possible inconsistencies, and standardization of the temporal structure. After these preparation steps, the processed data were used to train the BiLSTM–Attention model and to evaluate its forecasting capability.

Table 2*Characteristics of the dataset used*

Feature	Value
Industry under study	Polyester film (BOPET) manufacturing
Data type	Daily product demand
Data source	Comparable to the company's real data
Data nature	Demand time series
Historical data period	At least one year of daily data for each product
Training and validation period	Historical data through 2024
Forecasting period	2025

Data preprocessing is one of the key stages in preparing inputs for deep neural networks. In this study, demand data were first examined in terms of quality, and invalid values were corrected. Then, to reduce the effect of differences in demand scale and to improve the stability of the training process, the data were normalized to a specified range using Min–Max scaling.

After normalization, the time-series data were transformed into a structure suitable for neural network input. In this process, a sliding time-window method was used, such that the demand values of the previous 60 days were defined as the model input and the demand value of the following day was defined as the network's target output. The selection of a 60-day time window enables the model to learn short-term and medium-term patterns present in demand behavior.

Table 3*Main data preparation parameters*

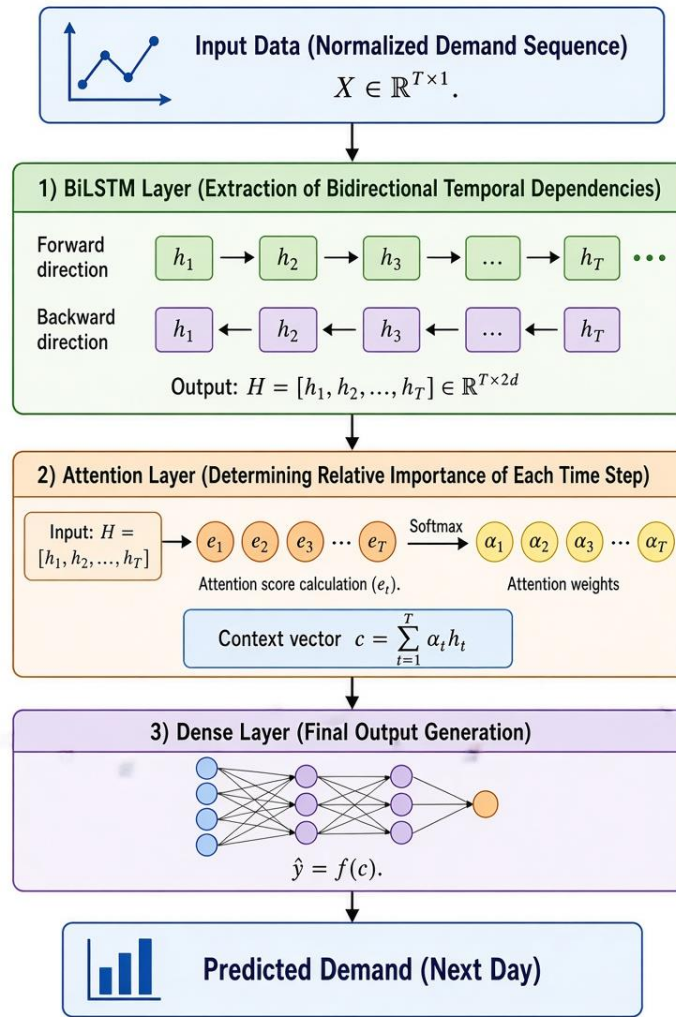
Parameter	Value
Look-back window length	60 days
Forecasting type	One-step-ahead (next day)
Normalization method	Min–Max Scaling
Data frequency	Daily
Target variable	Demand quantity

In this study, a hybrid BiLSTM-Attention architecture was developed for demand forecasting. This architecture was designed to extract complex temporal patterns and improve forecasting accuracy and consists of three main components:

- **BiLSTM layer:** By processing data sequences in both temporal directions, this layer enables the extraction of past and future temporal dependencies. The use of a bidirectional structure allows more information from historical demand behavior to be incorporated into the model's learning process.
- **Attention layer:** By assigning different weights to the input temporal observations, this layer determines the relative importance of each part of the sequence. As a result, the model is able to focus more strongly on the time periods that are more influential in forecasting.
- **Dense layer:** As the output component of the network, this layer receives the extracted features and generates the final predicted demand value.

The combination of these three components enables the proposed model not only to learn long-term dependencies in demand data but also to focus on important temporal patterns, thereby providing more accurate performance than conventional time-series models.

Figure 1*Architecture of the BiLSTM-Attention model*



To provide a realistic evaluation of model performance under conditions close to an operational environment, the walk-forward validation method was used. In this approach, the training and testing process is carried out sequentially over time, and future information is not used in model training.

For each month of 2024, all available data prior to that month were used as the training set, and the data of the same month were used as the test set. After forecasting the target month, data comparable to the actual observations of that month were added to the training set, and the process was repeated for the next month. This method better evaluates the model's ability to handle real future forecasting conditions.

To prevent overfitting and improve the model's generalizability, the early stopping method was applied. Under this approach, training was terminated if no improvement in the loss function was observed over 8 consecutive training epochs.

Table 4

Training parameters of the BiLSTM-Attention model

Parameter	Value
Look-back window length	60 days
Number of training epochs	60
Batch size	32
Optimizer	Adam
Loss function	Mean Squared Error (MSE)
Early stopping patience	8 epochs

To evaluate the performance of the proposed model, three commonly used metrics for time-series forecasting were applied:

This metric represents the average absolute difference between actual and predicted values and measures the average error of the model.

The MAE is calculated as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where:

- y_i : actual demand value
- $\hat{y}_i$: predicted demand value
- n : number of observations

By placing greater emphasis on larger errors, this metric evaluates the extent to which forecasts deviate from actual values.

The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where:

- y_i : actual demand value
- $\hat{y}_i$: predicted demand value
- n : number of samples

Safe Mean Absolute Percentage Error (Safe MAPE)

This metric is used to avoid the division-by-zero problem in cases of low or zero demand and measures the relative forecasting error.

$$Safe\ MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i + \varepsilon} \right|$$

where:

- y_i : actual demand value
- $\hat{y}_i$: predicted demand value
- ε : a very small number to avoid division by zero
- n : number of observations

The simultaneous use of these three metrics enables a comprehensive assessment of the accuracy, stability, and generalizability of the proposed model.

All stages of data preprocessing, model training, validation, and forecast generation were carried out in the Python environment. The neural network architecture was developed using the TensorFlow/Keras framework, and the Pandas, NumPy, and Scikit-learn libraries were used for data processing and numerical operations.

To ensure the reproducibility of the results, the random settings used in the training process were fixed, and all experiments were executed under identical computational conditions.

Findings and Results

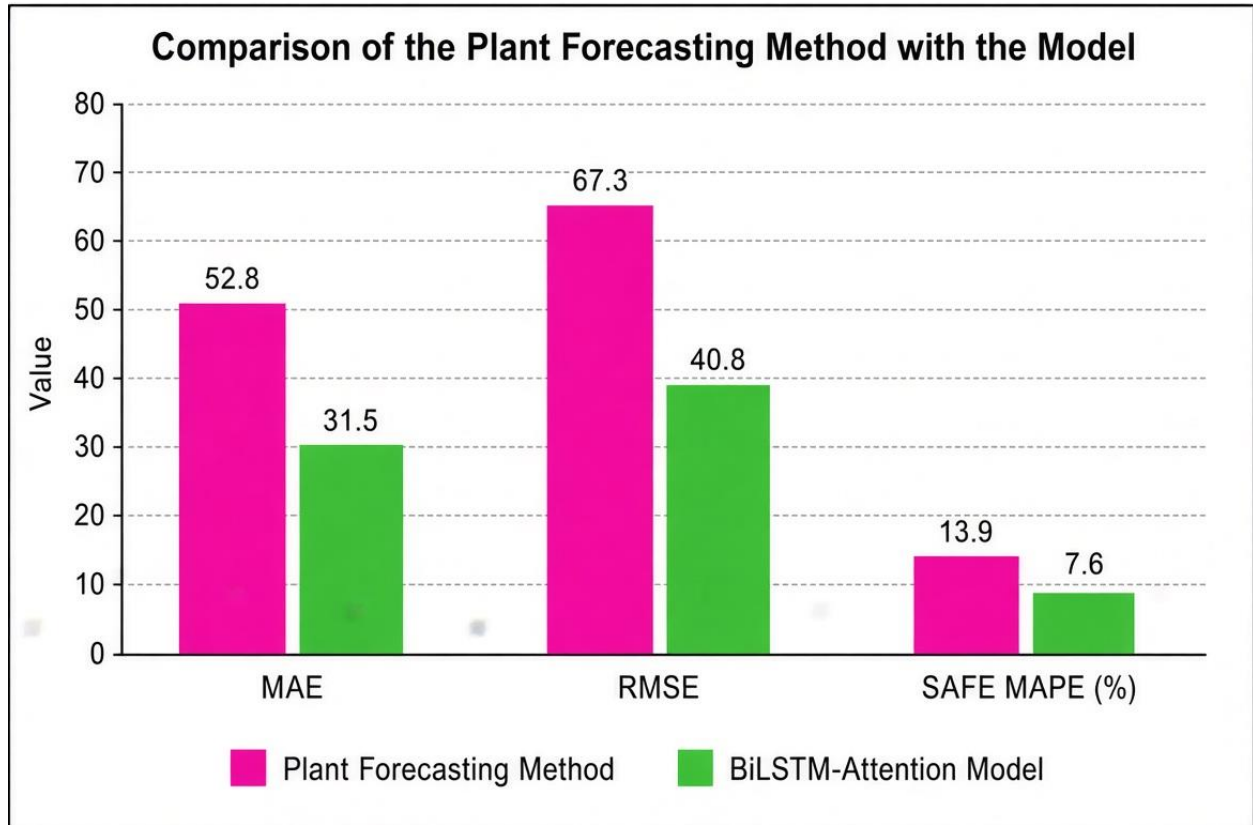
The results of this study are consistent with recent findings on the application of deep neural networks to time-series forecasting and show that combining the BiLSTM architecture with an attention mechanism can improve forecasting accuracy relative to conventional methods. This performance improvement stems from BiLSTM's ability to extract bidirectional temporal dependencies and from the attention mechanism's capacity to identify and weight the temporal observations that are most influential in the forecasting process.

This issue is particularly important in multi-stage manufacturing environments, where product demand is typically characterized by nonlinear behavior, periodic fluctuations, and abrupt changes. Under such conditions, more accurate forecasting models can play a significant role in reducing uncertainty in production planning, improving inventory-related decisions, and increasing the utilization of production capacity.

Overall, the findings indicate that the proposed model not only improves statistical forecasting metrics but also has practical applicability in real manufacturing environments. Forecasts generated by the BiLSTM-Attention model can serve as reliable inputs for make-to-stock (MTS) production planning systems and can support more intelligent decision-making in manufacturing industries.

Table 5*Comparison of the proposed model's performance with the forecasting method used in the plant*

Evaluation metric	Plant forecasting method	BiLSTM-Attention model	Improvement
MAE	52.8	31.5	40.3%
RMSE	67.3	40.8	39.4%
Safe MAPE (%)	13.9	7.6	45.3%

Figure 2*Comparison of Plant Forecasting Method with the Study Model*

According to the results in Table 5, the proposed BiLSTM-Attention model outperformed the plant's forecasting method on all evaluation metrics. The MAE decreased from 52.8 to 31.5, corresponding to a 40.3% reduction in average forecasting error. Likewise, the reduction of RMSE from 67.3 to 40.8 indicates a 39.4% decrease in larger errors and greater model stability in the face of demand variations.

The greatest improvement was observed for Safe MAPE, which declined from 13.9% to 7.6%, representing a 45.3% reduction in relative error. This result indicates that the proposed model performs more effectively in forecasting products with different demand scales and provides higher reliability for use in production decision-making processes.

Overall, the results confirm that the BiLSTM-Attention architecture can increase demand forecasting accuracy in manufacturing environments and provide more precise forecast information to support make-to-stock (MTS) production policies.

Discussion and Conclusion

The findings of the present study demonstrate that the proposed BiLSTM–Attention architecture substantially improved demand forecasting performance in the polyester film manufacturing environment compared with the forecasting method currently used in the plant. The model produced lower forecasting errors across all three evaluation criteria. Specifically, MAE decreased from 52.8 under the plant forecasting method to 31.5 using the proposed model, representing a 40.3% reduction in average absolute forecasting error. RMSE similarly decreased from 67.3 to 40.8, corresponding to a 39.4% improvement, while Safe MAPE declined from 13.9% to 7.6%, indicating a 45.3% reduction in relative forecasting error. These results indicate that the BiLSTM–Attention architecture was better able to reproduce the temporal behavior of product demand and generate forecasts that were closer to actual observations. The simultaneous reduction in MAE and RMSE is particularly important because it indicates improvement not only in the average magnitude of forecasting errors but also in the occurrence of relatively large errors. The substantial reduction in Safe MAPE further suggests that the proposed approach maintained improved performance across products or periods characterized by different demand magnitudes. Collectively, the results support the usefulness of combining bidirectional recurrent learning with an attention mechanism for forecasting nonlinear and fluctuating demand patterns in manufacturing environments.

The improved performance of the proposed model can first be explained by the temporal learning capability of the BiLSTM architecture. Manufacturing demand is rarely determined exclusively by the immediately preceding observation. Rather, it may reflect recurring demand patterns, delayed responses, short- and medium-term fluctuations, market cycles, and interactions among observations distributed across the historical sequence. Conventional forecasting procedures may not represent these relationships effectively, particularly where nonlinear patterns dominate. LSTM-based approaches were developed specifically to retain relevant information across sequential observations and have consequently shown considerable usefulness in demand and supply chain forecasting. Wang et al. demonstrated the suitability of LSTM for order-demand forecasting in environments characterized by short lead times and strict delivery requirements [5], while Nguyen et al. showed that LSTM-based approaches can simultaneously identify temporal patterns and anomalies in supply chain data [6]. The present results extend this evidence by suggesting that bidirectional processing can further strengthen the representation of dependencies contained within the historical input window. By integrating information extracted in two directions within the observed sequence, BiLSTM can construct a richer representation of the temporal structure than a conventional unidirectional recurrent network.

The findings are also consistent with previous studies that have directly used bidirectional recurrent networks in complex forecasting tasks. Joseph et al. developed a hybrid CNN–BiLSTM framework for store-item demand forecasting and found that bidirectional temporal learning improved the modeling of product-demand behavior [11]. Hao et al. similarly showed that a BiLSTM architecture combined with attention could effectively capture complex temporal relationships in atmospheric temperature forecasting [12]. More directly related to supply chain demand, Cui et al. proposed an attention-based BiLSTM framework and reported strong forecasting accuracy and generalizability across multiple demand datasets [15]. The agreement between these findings and the present results suggests that the performance gain observed in polyester film demand forecasting is not limited to a single dataset or industrial application. Instead, it appears to reflect a broader strength of bidirectional recurrent models in representing temporal sequences whose future behavior depends on complex combinations of earlier observations.

A second explanation for the observed improvement concerns the role of the attention mechanism. Although recurrent networks can retain sequential information, not all historical observations are equally informative for forecasting future demand. Certain periods may contain abrupt demand changes, recurring peaks, unusual declines, or patterns that are especially relevant to the next forecasting point, whereas other observations contribute little useful information. The attention mechanism allows the model to assign different weights to the representations generated from the temporal sequence and thereby emphasize observations that carry greater predictive information. This mechanism can reduce the influence of less informative observations while strengthening the contribution of critical temporal patterns. Zhang et al. found that combining BiLSTM with attention improved stock-price prediction by enabling the model to focus on more informative temporal features [14]. Cui et al. likewise demonstrated that the integration of self-attention with BiLSTM strengthened temporal feature extraction and forecasting generalizability in demand forecasting applications [15]. The substantial reduction in Safe MAPE observed in the current study is consistent with this interpretation, because adaptive weighting may have helped the model respond more effectively to changes in demand scale and to periods in which particular historical observations were disproportionately important.

The present findings also align with the expanding literature on hybrid deep-learning architectures for forecasting. Ghazouani et al. reported that a CNN–LSTM model improved detergent-product demand forecasting by combining the strengths of convolutional feature extraction and recurrent temporal learning [16]. Alfialali et al. similarly demonstrated the effectiveness of optimized LSTM architectures for demand forecasting in supply chain management [3]. Soltani et al. showed that machine-learning methods incorporating temporal clustering and neural networks can provide useful solutions for integrated omnichannel demand forecasting [2]. These studies collectively indicate that complex demand patterns frequently benefit from models capable of learning nonlinear structures directly from data rather than depending exclusively on fixed statistical assumptions. The results of the current research support this argument because the substantial reductions in forecasting error indicate that the proposed neural architecture was able to extract predictive structure from daily industrial demand data more effectively than the existing plant forecasting procedure.

The decrease in RMSE deserves particular attention. Because RMSE gives greater weight to relatively large deviations between forecasted and actual values, its reduction from 67.3 to 40.8 indicates that the proposed model was not merely improving forecasts during ordinary periods but was also reducing the severity of larger forecasting errors. This characteristic is operationally important in manufacturing systems because major forecast deviations can generate disproportionately large consequences. A severe underforecast may result in inventory shortages, delayed customer deliveries, urgent production changes, and inefficient rescheduling, whereas a severe overforecast may produce unnecessary inventories, excess use of working capital, and storage costs. Hybrid models have repeatedly demonstrated an ability to handle nonlinear or abrupt changes in time series. Rafie et al. reported favorable performance from integrated CNN–LSTM forecasting in short-term load prediction [7], and Debnath et al. showed that attention-based CNN–LSTM architectures improved forecasting under complex and rapidly changing load patterns [32]. The current reduction in RMSE is compatible with these findings and suggests that the BiLSTM–Attention architecture may be particularly useful where operational decision-making is sensitive to occasional large demand-prediction errors.

The findings are also supported by research showing that combining multiple learning mechanisms can improve forecasting robustness. Kim et al. developed a parallel LSTM–MLP model for mid-term demand forecasting and demonstrated

that combining different learning components improved the representation of temporal and nonlinear relationships [17]. Hamian et al. similarly found that a hybrid ARIMA–LSTM architecture could outperform individual forecasting approaches by combining different forms of predictive information [18]. In another manufacturing-related context, Kim et al. integrated kernel density estimation and LSTM in a two-dimensional forecasting framework designed to support cost-effective inventory management [8]. Such studies reinforce the interpretation that forecasting accuracy can be improved when architectures are designed to capture complementary dimensions of data behavior. In the current study, the complementary functions were provided by bidirectional temporal representation and attention-based weighting, allowing the model both to learn dependencies within the historical sequence and to identify which components of that sequence were most relevant to the forecast.

The magnitude of improvement observed in the present study is also consistent with evidence that optimized and attention-enhanced recurrent models outperform simpler neural architectures. Zhang et al. showed that optimizing LSTM through an improved genetic algorithm enhanced short-term traffic forecasting [21], while Jung et al. reported that genetic-algorithm-based parameter optimization improved LSTM performance in flash-flood forecasting [19]. Bouakline et al. similarly demonstrated improved air-quality forecasting through an LSTM architecture optimized using a modified genetic algorithm [20]. Al-Zaidawi applied optimized LSTM modeling to oil-price forecasting [10], and Al-Bayati employed LSTM with grey wolf optimization for electric-load forecasting [9]. Nikhil et al. further combined LSTM, CNN, and genetic optimization for financial time-series prediction [22]. Although the present model does not rely on an external metaheuristic optimization algorithm, these studies are relevant because they collectively demonstrate that standard recurrent structures often benefit from architectural or algorithmic enhancement. In the present research, the attention mechanism provides such enhancement by improving the selective processing of temporal information.

More recent evidence from highly complex forecasting domains further supports the results obtained here. Huang et al. showed that a CNN–BiLSTM–Attention architecture was effective in predicting pavement performance by integrating local feature extraction, bidirectional temporal learning, and attention weighting [23]. Yu et al. reported improved carbon-emission forecasting using a hybrid CNN–BiLSTM–Attention model supported by feature engineering [24]. Wang et al. likewise demonstrated that a CNN–BiLSTM–Attention framework was effective for total organic carbon logging prediction [25]. Gao et al. combined CNN–BiLSTM–Attention with heuristic optimization and obtained favorable forecasting performance under limited meteorological information [26]. Although these applications differ substantially from industrial product demand, they share important statistical characteristics with the present problem, including nonlinearity, temporal dependency, variability, and the need to identify informative signals within noisy sequential data. The convergence of findings across these domains strengthens the interpretation that attention-enhanced bidirectional recurrent architectures possess broad capability for complicated time-series forecasting.

The results also agree with studies emphasizing adaptive attention and multi-scale temporal representation. Wu et al. developed an MSC–BiLSTM model with an adaptive attention mechanism for ship trajectory forecasting and showed that dynamically weighting informative features improved forecasting performance [27]. Zheng et al. similarly demonstrated the effectiveness of reinforced CNN–BiLSTM–Attention fusion for forecasting nonlinear and unstable motion sequences [28]. Hu et al. combined attention-enhanced BiLSTM with Bayesian optimization for e-commerce sales prediction and reported improved predictive accuracy with direct relevance to inventory and demand-planning applications [31]. These findings

provide particularly relevant support for the current study because e-commerce sales and industrial product demand both require forecasts that can subsequently influence inventory and operational decisions. Moreover, the present findings complement the work of Feng, who showed that Transformer–LSTM architectures incorporating attention mechanisms can improve multi-node supply chain demand forecasting and model demand fluctuations across interconnected systems [30]. The consistency among these studies suggests that selective attention to temporally important information is increasingly becoming a central element of high-performance forecasting systems.

Beyond statistical performance, one of the most important findings of the present study concerns the operational applicability of the generated forecasts. The manuscript indicates that the forecasts produced by the BiLSTM–Attention model can serve as reliable inputs for make-to-stock production planning. In an MTS environment, production quantities are determined before actual customer demand is fully realized; consequently, forecasting accuracy directly affects the balance between product availability and inventory exposure. A reduction in average and relative forecasting error can therefore potentially reduce both shortages and unnecessary production. This interpretation is consistent with research linking advanced forecasting methods to inventory and supply chain management. Kim et al. explicitly connected LSTM-based forecasting with cost-effective inventory management in smart manufacturing [8], while Hu et al. emphasized the applicability of attention-enhanced sales forecasting to demand planning and inventory management [31]. Sorour and Amin further demonstrated the broader role of hybrid deep learning in adaptive smart-manufacturing decision systems, highlighting the movement from predictive modeling toward integrated operational optimization [29]. Therefore, the value of the proposed model should be considered not only in terms of lower MAE, RMSE, and Safe MAPE but also in terms of its potential to improve the quality of information supplied to production-planning decisions.

Another important methodological aspect is the use of walk-forward validation, which strengthens the practical interpretation of the findings. In conventional random training-test splits, observations from different points in time may be mixed, which can create unrealistic evaluation conditions for sequential forecasting. By contrast, walk-forward validation maintains temporal ordering and repeatedly evaluates the model using information available only before the prediction period. The present study used this sequential procedure to approximate real forecasting operations, thereby reducing the possibility that future information could influence model training. As a result, the observed improvements more closely represent the performance that might be expected when the model is implemented operationally. The favorable results therefore provide evidence not simply that the BiLSTM–Attention architecture can fit historical demand, but that it can generalize to subsequently observed periods. This distinction is important because generalization is essential if deep-learning models are to move from experimental environments to practical industrial decision support.

Overall, the findings indicate that the combination of BiLSTM and attention offers an effective framework for industrial demand forecasting. The BiLSTM component enables rich temporal representation of historical demand sequences, while the attention layer identifies and emphasizes the observations that contribute most strongly to prediction. The resulting architecture substantially reduced average, large, and relative forecasting errors compared with the existing plant method and generated forecasts with direct relevance to MTS production planning. The findings are consistent with previous studies involving demand forecasting, supply chains, energy systems, financial markets, environmental forecasting, transportation, and smart manufacturing, in which recurrent, bidirectional, attention-based, and hybrid architectures have repeatedly demonstrated advantages in modeling complex temporal data [1, 4, 13]. Accordingly, the present study adds industry-specific

evidence that a comparatively focused BiLSTM–Attention architecture can provide substantial forecasting improvements without requiring an excessively complex multi-model framework, while simultaneously generating information capable of supporting operational production decisions.

The present study has several limitations that should be considered when interpreting the findings. First, the model was developed and evaluated within the context of a single polyester film manufacturing environment, and demand behavior in this industry may differ from that of discrete manufacturing, food production, consumer electronics, pharmaceuticals, or other sectors. Second, the forecasting framework primarily relied on historical demand sequences and therefore did not explicitly incorporate potentially influential exogenous variables such as economic conditions, customer characteristics, marketing activities, competitor behavior, raw-material prices, exchange-rate fluctuations, promotional campaigns, or macroeconomic shocks. Third, the model was designed primarily for one-step-ahead daily forecasting using a fixed historical window, and its performance may differ for longer forecasting horizons or alternative temporal resolutions. Fourth, although walk-forward validation provides a realistic evaluation framework, the available period of historical data limits the range of unusual market conditions that can be represented during training and validation. Finally, the study evaluated forecasting performance primarily through MAE, RMSE, and Safe MAPE and did not directly quantify the financial effects of improved forecasting on inventory cost, stockout cost, service level, production utilization, or profitability.

Future research should validate the proposed architecture using datasets from multiple manufacturing companies and industries to determine the extent to which its performance generalizes across different demand structures and production systems. Future studies should also incorporate exogenous predictors such as product prices, economic indicators, customer orders, promotional events, seasonal factors, raw-material conditions, and market variables into multivariate forecasting frameworks. Comparative evaluations involving Transformer architectures, Temporal Fusion Transformers, convolutional-recurrent models, probabilistic forecasting approaches, and other attention-based networks would help determine the conditions under which BiLSTM–Attention provides the greatest advantage. Additional research could examine multi-step and multi-horizon forecasting, product-specific models, hierarchical forecasting across product families, and transfer-learning approaches for products with limited historical observations. An especially valuable extension would be the development of an integrated forecasting-and-optimization framework in which predicted demand is directly connected to production scheduling, inventory optimization, capacity allocation, and raw-material procurement decisions. Future studies could also examine forecast uncertainty through prediction intervals or probabilistic outputs rather than relying exclusively on point forecasts.

From a practical perspective, manufacturing firms can use the proposed forecasting framework as a decision-support component within production planning and inventory-management systems. Daily forecasts generated by the model can be incorporated into MTS planning to determine more appropriate production quantities, reduce unnecessary inventory accumulation, lower the risk of stockouts, and improve the allocation of production capacity across different products. Managers should implement the system as a continuously updated forecasting process in which new demand observations are regularly incorporated and model performance is monitored through operationally meaningful indicators. The model should also be integrated with existing enterprise resource planning, manufacturing execution, and inventory-control systems so that forecasting improvements translate into actual planning decisions. Particular attention should be given to products with highly volatile demand or high inventory and shortage costs, as these products are likely to benefit most from improved

forecasting accuracy. Firms should additionally establish periodic model-retraining procedures and monitor changes in demand behavior to detect model deterioration. Ultimately, the greatest practical value of the proposed approach will be achieved when its forecasts are not treated as isolated statistical outputs but are embedded directly within production, inventory, procurement, and capacity-planning processes.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

- [1] R. Shirazizadeh, "An Intelligent Demand Forecasting Model in the Supply Chain Using Machine Learning Approaches," Yazd University, 2022.
- [2] M. Soltani, S. M. Khatami, M. Amiri, and M. Hajian Heidari, "A Hybrid Approach to Integrated Omnichannel Demand Forecasting Using Machine Learning: Time-Series Clustering with the Dynamic Time Warping Algorithm and Artificial Neural Networks," *Journal of Production Management, University of Isfahan*, vol. 14, no. 32, pp. 121-140, 2023.
- [3] A. Alfialali, H. Ben Lamar, S. Alfialali, and M. Kasbaya, "Applications of Machine Learning in Supply Chain Management: An Optimized LSTM Model for Demand Forecasting," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, 2022.
- [4] S. S. Moradi Sanjed Bouri, "A Hybrid Meta-Learning and Generative Adversarial Network Approach for Dairy Product Demand Forecasting," Mazandaran University of Science and Technology, 2022.
- [5] C. C. Wang, C. H. Qin, and W. J. C. Therapy, "Application of ARIMA and LSTM for Order Demand Forecasting Based on Short Lead Time and On-Time Delivery Requirements," *Processes*, vol. 9, no. 7, p. 1157, 2021, doi: 10.3390/pr9071157.

- [6] H. D. Nguyen, S. T. Tran, A. Tomasi, and M. Hamad, "Forecasting and Anomaly Detection Approaches Using LSTM Autoencoder Techniques with Applications in Supply Chain Management," *International Journal of Information Management*, vol. 57, p. 102282, 2021, doi: 10.1016/j.ijinfomgt.2020.102282.
- [7] S. H. Rafie, N. Almasud, S. Rahman, and K. Hossain, "A Short-Term Load Forecasting Approach Using an Integrated CNN and LSTM Network," *IEEE Access*, vol. 10, 2021, doi: 10.1109/ACCESS.2021.3060654.
- [8] M. S. Kim, J. Y. Lee, C. Lee, and J. Jeong, "A KDE and LSTM-Based Two-Dimensional Forecasting Framework for Cost-Effective Inventory Management in Smart Manufacturing," *Applied Sciences*, vol. 12, no. 5, p. 2380, 2022, doi: 10.3390/app12052380.
- [9] S. K. A. Al-Bayati, "Electric Load Forecasting Using an LSTM Neural Network and the Grey Wolf Algorithm," Imam Reza International University, 2022.
- [10] A. M. M. Al-Zaidawi, "Oil Price Forecasting Using an LSTM Network Optimized by the Grey Wolf Algorithm," Imam Reza International University, 2023.
- [11] R. V. Joseph, A. Mohanty, S. Tyagi, S. Mishra, S. K. Satpati, and S. N. Mohanty, "A Hybrid Deep Learning Framework with CNN and Bidirectional LSTM for Store Item Demand Forecasting," *Computers & Electrical Engineering*, vol. 103, p. 108358, 2022, doi: 10.1016/j.compeleceng.2022.108358.
- [12] X. Hao, Y. Liu, L. Pei, W. Li, and Y. Du, "Atmospheric Temperature Prediction Based on a BiLSTM-Attention Model," *Symmetry*, vol. 14, no. 11, p. 2470, 2022, doi: 10.3390/sym14112470.
- [13] Z. Shahim Garmi, "Optimization of LSTM-CNN Hybridization for Time-Series Forecasting," Islamic Azad University, Safadasht Branch, 2022.
- [14] J. Zhang, L. Ye, and Y. Lai, "Stock Price Prediction Using CNN-BiLSTM-Attention Model," *Mathematics*, vol. 11, no. 9, p. 1985, 2023, doi: 10.3390/math11091985.
- [15] L. Cui, Y. Chen, J. Deng, and Z. Han, "A Novel AttLSTM Framework Combining Bidirectional LSTM and Self-Attention Mechanism for Demand Forecasting," *Expert Systems with Applications*, vol. 254, p. 124409, 2024, doi: 10.1016/j.eswa.2024.124409.
- [16] A. Ghazouani, A. Masmoudi, A. Mejri, and S. B. Laib, "A Hybrid Deep Learning Model CNN-LSTM for Detergent Product Demand Forecasting," *International Journal of Supply and Operations Management*, vol. 11, no. 4, pp. 417-429, 2024.
- [17] D. Kim, Y. Lee, S. K. Jo, and Y. M. Wai, "Mid-Term Minimum Demand Forecasting Based on Parallel LSTM-MLP Model," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3520986.
- [18] S. Hamian, Y. Kano, H. Khalifi, and M. Telmam, "Comparative Analysis of LSTM, ARIMA, and Hybrid Models for GDP Forecasting," *International Journal of Safety and Security Engineering*, vol. 29, no. 3, pp. 853-861, 2024, doi: 10.18280/isi.290306.
- [19] Y. Jung, C. X. Chen, B. C. Jung, C. H. Tsai, and S. Y. Yang, "Optimizing LSTM Parameters for Flash Flood Forecasting Using Genetic Algorithms," *Water Resources Management*, vol. 38, no. 1, pp. 1141-1164, 2024, doi: 10.1007/s11269-023-03713-8.
- [20] O. Bouakline, Y. El Marabet, A. El Idrissi, K. Khamsi, and R. Laghreib, "A Hybrid Deep Learning Model Integrating LSTM and Modified Genetic Algorithm for Air Quality Applications," *Environmental Monitoring and Assessment*, vol. 196, p. 1264, 2024, doi: 10.1007/s10661-024-13447-8.
- [21] J. Zhang, S. Ko, Z. Zhang, and S. Cheng, "Improved Genetic Algorithm and Optimized LSTM Model for Short-Term Traffic Flow Forecasting," *PeerJ Computer Science*, vol. 8, p. e1048, 2022, doi: 10.7717/peerj-cs.1048.
- [22] S. Nikhil, R. K. Sah, S. K. Parki, and S. Reddy, "Stock Market Prediction Using a Hybrid LSTM-CNN Model with Genetic Algorithm," 2023: IEEE.
- [23] Y. Huang, C. Chen, and X. Dai, "A CNN-BiLSTM-Attention-Based Deep Learning Approach for Predicting Asphalt Pavement Performance," *Buildings*, vol. 16, no. 6, p. 1150, 2026, doi: 10.3390/buildings16061150.
- [24] S. Yu, Y. Tang, X. Xu, and T. Liu, "A Hybrid CNN-BiLSTM-Attention Model with Feature Engineering for Accurate Carbon Emission Forecasting," *Air Quality, Atmosphere & Health*, vol. 19, p. 56, 2026, doi: 10.1007/s11869-026-01954-1.
- [25] Y. Wang, R. Tian, X. Liu, and R. Tan, "TOC Logging Prediction Method Based on CNN-BiLSTM-Attention," *Journal of Geophysical Prospecting*, vol. 2, no. 3, pp. 207-217, 2026, doi: 10.1016/j.jgp.2026.05.003.

- [26] Y. Gao, T. Fu, M. He, F. Yang, and X. Li, "Heuristic Algorithm Optimization of CNN-BiLSTM-Attention for Reference Crop Evapotranspiration Forecasting under Limited Meteorological Data Availability," *Atmosphere*, vol. 17, no. 4, p. 382, 2026, doi: 10.3390/atmos17040382.
- [27] N. Wu, D. Chen, R. Gu, C. Wen, W. Tian, and J. Lu, "A Study on the MSC-BiLSTM Ship Track Prediction Model Incorporating an Adaptive Attention Mechanism," *Journal of Marine Science and Engineering*, vol. 14, no. 10, p. 924, 2026, doi: 10.3390/jmse14100924.
- [28] H. Zheng *et al.*, "Accurate Time-Series Forecasting of Floating Platform Motion via a Reinforced Fusion CNN-BiLSTM-Attention Model," *PLOS ONE*, 2026, doi: 10.1371/journal.pone.0342081.
- [29] S. E. Sorour and A. E. Amin, "EvoDeep-Quality: A Closed-Loop Hybrid Framework Integrating CNN-LSTM and NSGA-III for Adaptive Quality Optimization in Smart Manufacturing," *Sustainability*, vol. 18, no. 8, p. 3679, 2026, doi: 10.3390/su18083679.
- [30] J. Feng, "Multi-Node Supply Chain Demand Forecasting and Fluctuation Modeling Based on Transformer-LSTM Hybrid Model," *Journal of Engineering, Project & Production Management*, vol. 16, no. 5, p. 1, 2026, doi: 10.32738/JEPPM-2025-366.
- [31] H. Hu, J. Cai, and C. Xu, "A Mathematical Framework for E-Commerce Sales Prediction Using Attention-Enhanced BiLSTM and Bayesian Optimization," *Mathematics*, vol. 31, no. 1, p. 17, 2026, doi: 10.3390/mca31010017.
- [32] S. Debnath, M. U. Mia, M. Abubakkar, M. R. Islam, M. S. I. Mridul, and A. K. Biswas, "Hybrid Multi-Scale Deep Learning Enhanced Electricity Load Forecasting Using Attention-Based Convolutional Neural Network and LSTM Model," *IEEE Access*, vol. 14, 2026, doi: 10.1109/ACCESS.2026.3656545.