ICCK Transactions on Intelligent Systematics
http://dx.doi.org/10.62762/TIS.2024.300700

RESEARCH ARTICLE

Check for updates

Comparison of Deep Learning Algorithms for Retail Sales Forecasting

Rana Shakeel Ahmed[1], **Muhammad Hasnain**[2,*], **Muhammad Hamza Mahmood**[1] and **Muhammad Abid Mehmood**[1]

[1] Department of Computer Science, Lahore Leads University, Lahore 54130, Pakistan
[2] Department of Information Technology, Lahore Leads University, Lahore 54130, Pakistan

Abstract

We investigate the use of deep learning models for retail sales forecasting in this research. Proper sales forecasting can lead to optimization in inventory management, marketing strategies, and other core business operations. This research evaluates deep learning models such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP), and a hybrid CNN-LSTM model. The models are further improved by adding dense layers to process daily sales data from a major pharmaceutical company. The models are trained on 80% of the dataset and tested on the remaining 20%. Model performance is compared using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results indicate that the CNN-LSTM model outperforms the others, achieving the lowest RMSE and MAE values, making it the most suitable for sales forecasting in this context. This research contributes to the field by demonstrating the superiority of hybrid models in handling complex temporal data for predictive analytics. Future work will explore the integration of additional data sources and advanced deep learning architectures to further improve forecasting accuracy and applicability.

Keywords: sales forecasting, deep learning, CNN, LSTM, retail analytics.

1 Introduction

Nowadays, every company relies on sales forecasts. Increasing product sales is a goal of every business [1]. The ability to make informed decisions about product manufacturing hinges on the business owner's knowledge of the products' future demand. Forecasting methods help us figure out what's coming next. Some statistical approaches, human planning, or a mix of the two is used to accomplish forecasting [2]. Predicting future demands or trends is a challenging process and has become an emerging focus in the market. Because forecasting models can occasionally produce over-or under-forecasting, it is also challenging to construct an accurate prediction model [3]. On the one hand, a business owner runs the risk of missing out on a sales opportunity if they under-estimate or over-estimate their sales, while on the other hand, they run the risk of producing

Academic Editor:
Jianlei Kong

Submitted: 15 September 2024
Accepted: 03 October 2024
Published: 20 October 2024

Vol. 1, **No.** 3, 2024.
10.62762/TIS.2024.300700

***Corresponding author:**
Muhammad Hasnain
drhasnain.it@leads.edu.pk

Citation
Ahmed, R. S., Hasnain, M., Mahmood, M. H., & Mehmood, M. A. (2024). Comparison of Deep Learning Algorithms for Retail Sales Forecasting. *ICCK Transactions on Intelligent Systematics*, 1(3), 112–126.

unnecessary products, which drives up production costs. In order to plan production accordingly and satisfy future demands, manufacturers require reliable predictions. In order to foretell future requests or sales, a variety of forecasting methods are available [4]. Sales forecasting using a back propagation neural network method has its uses, however dealing with huge parameters can be a challenge. An effective approach to address the issue of managing extensive parameters in sales forecasting with a backpropagation neural network (BPNN) is to adopt more advanced deep learning architectures, such as convolutional neural networks (CNNs) (see Figure 1) and long short-term memory networks (LSTMs). These models are capable of processing large datasets while simplifying parameters' complexity by utilizing convolutional layers for feature extraction in CNNs and incorporating memory units in LSTM to preserve temporal information without burdening the system with an excessive number of parameters. Additionally, strategies such as dropout regularization and parameter sharing can further alleviate the parameter burden, leading to more efficient and precise forecasting.

The sales forecast shows the expected quantity of a product sold in a given market at a given price point over a given time period [5]. Many businesses manufacture goods in advance to fulfil future requests, while others manufacture them on an order basis. Estimating the kind, amount, and quality of future activities, such as sales or items, is what forecasting is all about. In order to foretell future requests or sales, a variety of forecasting methods are available [6].

Deep learning, machine learning, a few regression models, the "Autoregressive Integrated Moving Average" ARIMA model, and others are all tools in the toolbox when it comes to building predictive models [7]. A number of factors, including the model's processing speed, the correctness of the findings, the model's robustness, its interpretability, and how easy it is to use, must be validated after the appropriate model has been built [8]. It is standard practice to create multiple models using various methods and then compare these models. For model development, some authors utilize basic or easy techniques like logistic regression, neural networks, or classification trees. To improve accuracy and reduce errors, others use deep learning-based approaches for demand forecasting in specialized retail domains such as fashion [9].

We can choose from a variety of predicting methods. When trying to predict future sales, many businesses employ a combination of strategies [10]. The decision on sales forecasting is made by one or more executives with extensive expertise and solid industry understanding. Quick and easy, this is the way to go. However, LSTM networks are highly effective for future sales predictions due to their ability to capture and learn long-term dependencies in time series data, ensuring accurate and reliable forecasts. It is capable of learning long-term dependence and memorization. In time series prediction, it excels because it can recall past inputs [11].

Recent developments, such as the Informer algorithm, have demonstrated potential in long-term forecasting, especially due to its effective attention mechanism that captures long-term dependencies within time-series data. Nevertheless, our study emphasizes the benefits of hybrid models like the CNN-LSTM architecture, which adeptly handles both spatial and temporal dependencies in retail sales data. The CNN component proficiently extracts intricate features, while the LSTM component preserves information over time, rendering this hybrid model resilient to data variability—an aspect that some advanced algorithms may overlook. This adaptability contributes to enhanced predictive performance.

Businesses can utilize deep learning models [12, 13] to improve sales forecasting accuracy. This enables employees to spend less time and effort on manual analysis and focus more on strategic activities. In the highly competitive retail sector, organizations can gain a significant advantage by leveraging advanced predictive analytics. Compared with competitors that rely on traditional approaches, they can better understand market trends and respond more rapidly to changing customer demands. Furthermore, deep learning models can continuously adapt and learn from new data, allowing businesses to remain competitive and relevant in the dynamic retail market. Motivating this research is the belief that deep learning algorithms, including multimodal approaches, can provide significant benefits to various retail forecasting scenarios [14]. By fixing the problems mentioned earlier, these algorithms might revolutionize how retail companies operate and make decisions. Using modified transformer model, which can learn from large datasets, can substantially improve the accuracy of sales estimates [15]. All aspects of the business plan, including inventory management and marketing, depend on this level of accuracy. Algorithms like these find patterns in the large amount of customer data, which is beyond the ability of a human. As a result of this understanding, we can provide better personalized experiences and increase client happiness.

This paper proposes a hybrid CNN-LSTM model with adjusted dense layers. The main contributions are as follows:

- To propose the application of deep learning models such as CNN, LSTM, MLP and hybrid CNN-LSTM models for retail sales forecasting.
- To evaluate the performance of deep learning models using performance evaluation metrics.

This paper is organized as follows: The overview of literature on related sales forecasting research is presented in Section 2. The proposed approach in this research along with its implementation setup is detailed in Section 3. Applying the suggested methodology to the dataset yielded the results, which are presented and discussed in Section 4. Section 5 concludes the findings and recommendations for further research on performance indicators.

2 Related Work

Recent study investigates a number of machine learning techniques for the purpose of making sales forecasts for retail establishments [16]. The proposed method uses normalization to reduce training time while improving data quality. Another study [17] presents the results of a comparative analysis of machine learning models that are used to sales predictive analytics. Better stock administration and personnel scheduling, which increases transactions and customer loyalty, is a direct result of sales prediction, an integral aspect of modern business intelligence.

Authors suggested the model that uses recurrent neural networks and transformers to predict the daily cash flow of power sales [18]. This model retrieves the daily cash flow results after mining and analyzing payment data based on previous data, learning the internal characteristics automatically, and extracting the information. Prediction and analysis of previous data are handled by a 3-layer GRU unit. Following the last experimental results, the author claims that a recurrent neural network model outperforms an ARIMA model in terms of prediction accuracy. The development, investment, and consumption of electricity capital can benefit from knowing the correct estimate of daily sales cash flow. Researchers in a study [19] propose a hybrid deep learning framework combining CNN and Bi-directional LSTM for store item demand forecasting, demonstrating superior performance over standalone models in terms of RMSE, MAE, and MAPE. Next, they trained their weights with a differential evaluation model after selecting a structure with an adaptive learning model.

A previous research focused on the applications of DL algorithms to sales forecasting [20]. The fashion sector is the focus of their empirical research, which employs a variety of attributes in their dataset. They do their analysis using the fashion dataset given by companies. They make use of a broad variety of product factors and physical attributes. The performance of the proposed model is compared with various existing approaches, including decision trees, random forests, SVMs, SVRs, artificial neural networks, and linear regression. They also propose a model using deep learning techniques and discover a decent model for fashion market forecasting. The proposed model isn't just for the fashion business; it can be applied to any prediction analysis. However, they suggested that in order to do more targeted analyses, we should modify the dataset and features.

A forecasting model has been developed to reduce drug supply shortages in the pharmaceutical retail context [21]. The study constructs a two-level hierarchical forecasting framework that operates at both the individual pharmacy level and the retail chain level, incorporating outlier detection and missing data treatment as preprocessing steps. Among the candidate methods evaluated — including the naive method, moving average, exponential smoothing variants, Holt's linear method, and ARIMA-based approaches — Theil's U2 statistic is employed to assess forecasting accuracy across planning horizons. The findings indicate that disaggregated forecasting at the individual pharmacy level, subsequently integrated into chain-level plans, yields superior results, and that weekly availability monitoring is strongly recommended. The study is empirically validated using Lithuanian pharmaceutical retail data, demonstrating that incorporating outlier and shortage-period correction into the forecasting pipeline is essential to avoid systematic oversupply errors.

There are multiple branded traditional forecasting algorithms, for example ARIMA and Exponential smoothing that fail to capture the complexity of multiple influencing factors in time-series data which exhibit non-linear patterns. These approaches need manual feature engineering and typically posit a stationary time series [11]. On the other hand, deep learning methods, such as the models used in this research, are good at learning complex patterns in raw data automatically and capable of handling non-linear relations between variables that make them more appropriate for intricate forecasting like retail sales.

Several models have been used to predict retail sales rates [22]. This study aimed at evaluating these models and find the one with the highest accuracy rate. Machine learning methods such as boosting algorithms and linear regression are utilized. Multiple regression, polynomial regression, ridge regression, and lasso regression are among the forecasting approaches they employ. Among the numerous boosting algorithms, they employ are AdaBoost, Gradient Tree Boosting, and XGBoost. For this study, they scoured 10 stores in different cities, using 1559 goods total. According to the results, the boosting algorithms outperformed the standard regression methods in terms of prediction accuracy.

Forecasting retail sales is an essential aspect of business organizations, enabling them to enhance inventory management, refining marketing strategies, and improving overall operational efficiency. The advent of deep learning models has markedly transformed this domain by offering predictive capabilities that surpass traditional prediction techniques. Among the most frequently utilized models are Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Multilayer Perceptrons (MLPs). Notably, the hybrid CNN-LSTM model excels in capturing both spatial and temporal relationships within the data, rendering it particularly effective for sales forecasting. Research indicates that the incorporation of dense layers and the enhancement of data representation can significantly boost prediction accuracy, particularly in scenarios where daily sales data are influenced by external variables such as promotions or seasonal variations.

While deep learning methodologies have been utilized for forecasting retail sales, there remain significant gaps in the research. The majority of studies have concentrated on either spatial or temporal patterns in isolation, overlooking the interaction between these dimensions within retail data. Furthermore, there is a scarcity of investigations into the efficacy of hybrid models that integrate CNN and LSTM architectures, particularly in the context of large-scale retail sales datasets [16]. Additionally, the existing literature frequently fails to provide a thorough comparison of these models under consistent conditions, which is essential for determining the most effective strategy for managing large datasets.

This research enhances the current body of literature by performing a comprehensive analysis of CNN, LSTM, MLP, and hybrid CNN-LSTM models utilizing a real-world retail sales dataset. It fills the void of limited investigation into hybrid models and showcases their effectiveness in managing intricate temporal data [19]. Additionally, the incorporation of dense layers to represent daily sales data presents a novel strategy for enhancing forecasting accuracy, an area that has not been extensively explored in previous studies. By concentrating on an extensive dataset from the pharmaceutical sector, this study offers valuable insights that may be relevant to various other industries.

The research presented above illustrate how sophisticated algorithms can enhance customer delight, reduce costs, and streamline business processes. This categorization applies to a plethora of other meaningful applications as well, such as precise real estate valuation or accurate short-term sales forecasting at supermarkets.

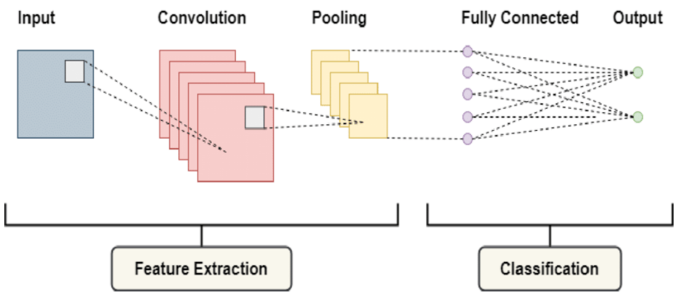

Figure 1. CNN architecture [2].

3 Methodology

The following research approaches were utilized in this paper to address the research questions. To begin, in order to discover improved findings, we learn from the existing relevant research. We employed deep learning algorithms to analyze retail sales based on the results of a number of literature evaluations that surveyed various relevant existing researches. Evaluating the performance of several deep learning algorithms, such as CNN, MLP, LSTM, and CNN-LSTM, is the primary objective of this research. In order to achieve very high accuracy with very little error rate in this research, we additionally added some dense layer for all of the CNN. To do this, we applied as many as four dense layers to each algorithm before assessing their performance. In our model, we train and test four different deep learning algorithms on our dataset, each with its own set of additional dense layers. We then analyze their performance to determine which model is best for sales forecasting.

3.1 Algorithms

Algorithms used in this study such as CNN, MLP, LSTM, and CNN-LSTM, are popular known models. In the following, we provide the architectures of these models. To explain the functionalities of CNN, MLP, and LSTM models of deep learning, the architectures of each model are presented below. The following are the diagrams depicting the characteristics and working of these models as an indication of what is different about them and how the models work.

CNN converts the input sequences into features through the convolution layers and then switch to pooling layers to minimize dimensionality. This extracted features are taken to fully connected layers and then to softmax layer which is used for classification. This architecture is specifically very good at preserving sparsity at different levels of the structure.

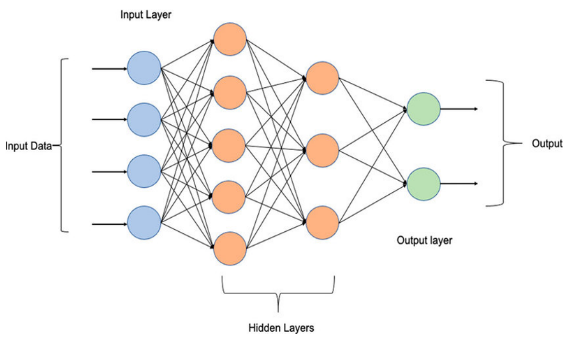

Figure 2. MLP Architecture [3].

ANN architecture of MLP is that of a feed forward neural network and its architecture has one input layer, one or more hidden layers and an output layer and all the layers are fully connected, as shown in Figure 2. Input data is passed through the network of neurons based on different layers where each neuron applies the activation function for its outputs. This architecture can learn non-linear relationships in the data by using a derivative free approach.

LSTM networks on the other hand consist of input gate; forget gate, and output gate which control the flow of data in the cell. The forget gate removes information unnecessary in the current scenario, the input gate adjusts the cell's state based on incoming information, and the output gate transmits information to subsequent time-step. This architecture performs very well in capturing long term dependencies in sequential data, as shown in Figure 3.

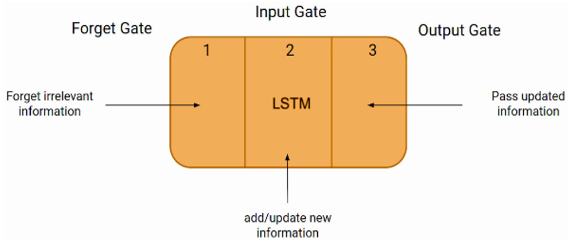

Figure 3. LSTM Architecture [4].

Convolutional Neural Networks (CNNs) are highly effective in automatic feature extraction through convolutional layers, making them particularly suitable for identifying spatial patterns and relationships in multivariate retail sales data. Multilayer Perceptrons (MLPs) excel at capturing non-linear relationships between variables; however, they lack the capability to model temporal dependencies inherent in time-series sales data. In contrast, Long Short-Term Memory (LSTM) networks are specifically designed for sequential data, utilizing memory cells to effectively capture long-term temporal dependencies, which makes them well-suited for time series forecasting tasks such as retail sales prediction. The hybrid CNN-LSTM model leverages the strengths of both architectures efficiently: the CNN component extracts relevant spatial and local features from the sales data, while the LSTM component models the temporal dynamics and long-term trends. The present research demonstrates that these advanced techniques, particularly the CNN-LSTM hybrid enhanced with additional dense layers, achieve superior prediction accuracy and lower error rates compared to individual models in prior studies for retail sales forecasting. For example, [23] implemented a single LSTM model for sales demand forecasting but did not incorporate convolutional layers for enhanced feature extraction. [24] applied MLPs for retail sales prediction, which, while effective for non-linear modeling, struggled to capture sequential temporal patterns in daily sales data. Similarly, authors in [25] utilized a standalone CNN model for sales forecasting, which performed well in feature extraction but was limited in handling long-term temporal dependencies. The proposed hybrid CNN-LSTM approach with dense layers outperformed these models in terms of RMSE and MAE on pharmaceutical retail sales data. While some

Table 1. Software requirements.

Item Name	Description
Python	An open-source, high-level programming language used for performing different deep learning models.
Pandas	Data manipulation tool that is open source, fast, powerful and flexible. It provides numerical and time series data structures and operations for solving these problems.
Numpy	Base N-dimensional array package for SciPy library It offers high level n-dimensional arrays and numerical computing tools, available across various scientific domains.
Matplotlib	The holoviews package is a high-level tool for convenient visualization of different types of data in Python Matplotlib: It helps in plotting two-dimensional graphs for creating publication-quality figures.
Seaborn	An appealing package for Python visualizations built on Matplotlib that can generate high-level, meaningful statistical images.
Sklearn	A quick and easy way to analyze data built on Matplotlib, Scipy, and Numpy. It is accessible to everyone and reusable in various contexts.
Keras	A library open to everyone, makes it easier for a large group of people to use artificial neural networks with a Python interface that opens the door to the TensorFlow library.

Table 2. Hardware requirements.

Item Name	Description
High-Performance GPU	A high-performance Graphics Processing Unit (GPU) for efficient deep learning model training and inference.
High RAM Capacity	Sufficient Random Access Memory (RAM) to handle large datasets and complex computations efficiently.
SSD Storage	Solid State Drive (SSD) for fast data access and storage.
Multi-core CPU	A multi-core Central Processing Unit (CPU) to support parallel processing and enhance computation speed.

researchers have relied on single models, the current study evaluates multiple architectures and integrates their complementary strengths to achieve more robust and accurate retail sales forecasts.

3.2 Software and Hardware Requirements

This research employs Python, an open-source high-level programming language, to implement the proposed hybrid approach. We present software and hardware requirements in the following Tables 1 and 2.

3.3 Dataset Description

Daily historical sales data from a globally recognized pharmaceutical company were collected and used in this research study. The dataset also includes the total number of products sold per day, representing daily sales information. Data were retrieved from multiple sources using PLSQL queries, as the Oracle database system supports PLSQL operations. Oracle was

Table 3. Represent data fields description.

Field Name	Description
ID	Presents the (item) in the tuple
item_id	Unique identifier for products
item_category_id	Unique identifier for categories of item
date_num	Represents consecutive data numbers
Date	Date in dd/mm/yyyy format
item_sale_day	Number of items sold in a day
item_name	Name of the item
item_category_name	Name of the item category

therefore utilized for data collection and management. The data were gathered over the period from January 1, 2021, to December 31, 2023. Among the collected data, 80% were used for training and the remaining 20% for testing. The training dataset contains 1,048,576 rows, including information such as dates, item identifiers, prices, and sales days. Using Oracle, the required data were extracted by combining multiple tables through inner and outer joins. The dataset fields are presented in Table 3.

3.4 Data Preprocessing

Numerous established approaches to sales forecasting are at the disposal. In terms of efficiency, precision, and error rate, each of these approaches is unique. After we transform the data into days, we look for outliers, missing numbers, and null values. To prepare our dataset for testing and training, we eliminated any outliers and filled in any missing numbers.

While performing data preprocessing we made a proper check list for handling outliers and missing values. They exclude outliers in order to reduce their impact on the assessment of the model performance, as well as to overcome the possible influence of outliers on the results of the sales analysis. Finally, for the missing values we have used imputation methods in order not to distort the data set. In addition, we had to address the missing numerical data using the mean or median of the respective fields of more frequent value imputation with categorical data. This made preprocessing an important step, as it produces a clean and accurate dataset for both training and testing the models, thereby improving the overall performance of the predictive models [20].

3.5 Feature Selection

The succession of deep learning models is influenced by numerous critical parameters. When implementing deep learning algorithms, feature selection is crucial. It improves the model's accuracy while decreasing training time and protecting data from overfitting by eliminating data dependencies. In order to make our model more efficient, we narrow down the list of input variables and pick the ones that will be most helpful. To choose features, we employ a selection procedure, correlation statistics, and transform variables. Regression is also employed to eliminate data dependencies.

In feature selection, we used the Pearson correlation statistics that proved useful in filtering out the features with a high linear dependency with the target variable and by doing so we eliminated any input variables that were inconsequential or repetitive. Also, various techniques, including stepwise regression, were used in order to eliminate those features that have little contribution to the efficiency of the model. This made it possible to capture only the relevant variables which enhance the efficiency of the model hence reducing the chances of getting a wrong model during retail sales forecasting.

3.6 Dense Layer

The neural network is strongly linked to this layer. This layer of neural networks is common and utilized extensively. The adjective "dense" describes these layers perfectly. In neural networks, it is one of many specialized layers. The fundamental unit of artificial neural networks, dense layers aggregate all the inputs and outputs from each layer. There is a weight W, an activation function a, and a bias vector b in this layer. Applying the dense layer at the end of the network and doing it again up to four times yields the best performance. Every neuron in the layer below the dense layer is connected to every neuron in the layer above it, as shown in Figure 4.

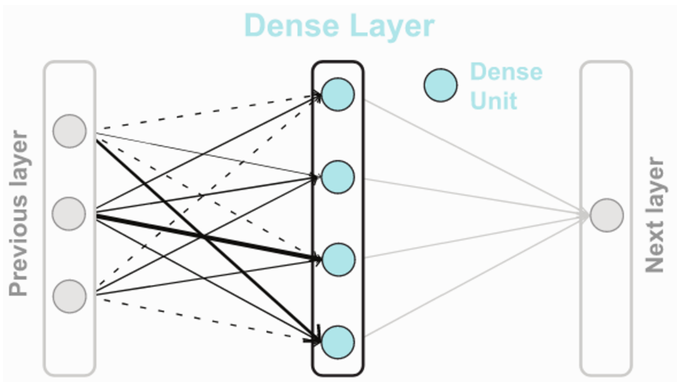

Figure 4. Visualization of the Additional Dense Layer in the Neural Network.

The dense layer in our study contributes significantly in the improvement of the performance of the proposed model. A dense layer is used so that every neuron in one layer is connected with each other neuron in the next layer and the architecture of network fully connected. In our setup, we added four dense layers at the end of the network using the inputs from subsequent layers – weights (W), activation (a) and bias (b). Precise configuration of the dense layer proves vital in enhancing the feature extraction process hence increases the probability of high accurate prediction since the model can easily relate between the data. Dense layer placement along with the sufficient repetition has been specifically located to be most effective for our forecasting models; making it a critical innovation to this study.

3.7 Workflow of the Proposed Method

First of all, as discussed earlier, we collect dataset at the point of sale in the pharmaceutical business. At the next stage, blanks or missing values are removed from the dataset. Next we extract the features from the data set. Before analyzing our training dataset using the remaining 20% of the data, we trained it using a new deep learning model with extra dense layers on 80% of the rows in the dataset. Following the assessment, we choose the most suitable model for making predictions according to the outcomes.

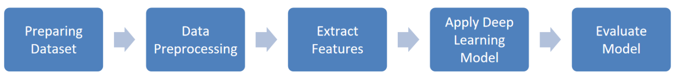

Figure 5. Proposed workflow of the models.

The Figure 5 depicts the proposed approach of utilizing deep learning models for retail sales forecasting. We begin by gathering all of the necessary data, getting it ready for analysis, and then cleaning it up. Preprocessing the data is the first step after data collection since it is crucial for learning models to ensure improved accuracy. We choose the data feature. After that, we ran the dataset through the deep learning models and drew conclusions about the findings depending on the accuracy and number of errors.

3.8 Evaluation Metrics

This study used Mean Absolute Error and Root Mean Square Error as evaluation measures for retail sale forecasting.

3.8.1 Mean Absolute Error (MAE)

This metric measures the average magnitude of prediction errors. It is computed as the mean of the absolute differences between the predicted and actual values. In order to summarize how well the deep learning models performed, this statistic is used for performance evaluation. We can use it to check how well the predictions match up with the real results. Over time, the utter inaccuracy levelled off. A lower error rate indicates a more accurate model. The following is the formula for the Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\gamma_i - \hat{\gamma}_i| \tag{1}$$

where γ_i represents actual values and $\hat{\gamma}_i$ represents the forecasted values.

3.8.2 Root Mean Square Error (RMSE)

It is the square root of the mean squared error, i.e., the square root of the average of the squared differences between the predicted and actual values. A lower error rate indicates a more accurate model.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\gamma_i - \hat{\gamma}_i)^2} \tag{2}$$

where γ_i represents actual values and $\hat{\gamma}_i$ represents the forecasted values.

4 Results and Discussion

This section provides results of the implemented approach and their discussion in the following. Before this, we provide the experimental parameters and their settings of chose models.

4.1 Experimental Parameters and their Settings

4.1.1 CNN Model

A CNN model combines 256 filters and kernel size equals 3, which takes 36 input observations from a three-year dataset. In trained 100 epochs using the batch size of 100, the model learns the spatial patterns of the unit sales data effectively. The RMSE of 0.08442 and MAE of 0.36069 tell us that we have well accuracy in terms of spatial features which is good.

4.1.2 LSTM Model

The LSTM aimed at capturing temporal feature in the sales data consists of 50 units with difference order of 12. Consequently, the model that has been trained up to 100 epochs was able to predict with an RMSE of 0.07331 and the MAE of 0.32517. These results show that it has advantage over the CNN model in terms of time-sequence data handling, which makes it fit in handling the temporal aspect of the sales trends.

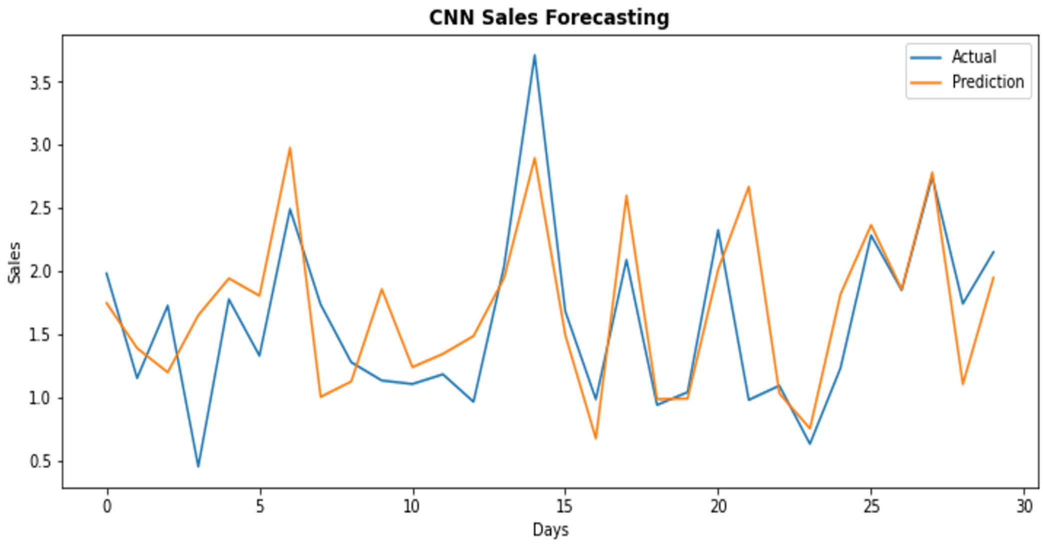

Figure 6. CNN sales forecasting.

4.1.3 CNN-LSTM Model

The proposed model integrates the CNN for spatial feature extraction with the LSTM for temporal feature analysis where the CNN's 64 filters are fused with LSTM neurons of 100 units. This model was trained over more than 200 epochs and yielded RMSE of 0.06514 and MAE of 0.29392, it seems that this combined model performed better than CNN and LSTM models. This implies that the hybrid model is best suited to capture both spatial and temporal patterns, hence making it our best choice of solution in this forecasting problem.

4.1.4 MLP Model

Based on the MLP model with 500 nodes in the hidden layer, it is more centered on the effective and fully connected layer. After training for 100 cycles of training and validations, the created model had an RMSE of 0.11608 and an MAE of 0.52888.

4.2 Convolutional Neural Network (CNN) Results

A convolutional neural network trained on more than one layer is a perfect method for deep learning. Applications of this can be anything, from predicting time series in forecasting to analyzing medical and satellite images, to detect and predict anomalies. To predict the sale for the next month, we use a Convolutional Neural Network (CNN) model over the data. We ran this Model four times with extra dense layers to improve the results and lower the inaccuracy.

In order to extract features, we build a CNN model with four convolutional layers. Using the rectified linear activation function, each layer has a customizable number of kernel sizes and filters. The number of steps is defined by the kernel size and the number of filters employed to determine the number of parallel fields on which the weighted input is projected and read. Following the convolutional layer, we additionally employ a max pooling layer to refine the weighted input feature.

For the purpose of sales forecasting, our CNN model is configured as follows.

- N_input: 36 (3 Years Dataset) this is the number of observation in order to feed into a model.
- N_filters: 256 the amount of filters that were incorporated into our model.
- N_kernel: 3 The amount of steps read from each input sequence.
- N_epochs: 100 This model's training dataset is revealed a certain number of times.
- N_batch: 100 made use of the sample size for every iteration following the weight updates.

Figure 6 shows the actual sale and the yellow line shows the expected sale for the entire month, as obtained by the CNN model.

4.3 Long Short-term Memory (LSTM) Results

LSTM is employed in time series and sales forecasting. In order to make predictions, it keeps track of the inputs made before. All four layers of the Long Short-Term Memory algorithm communicate with one another. We create a multi-layer LSTM model with a hidden LSTM layer that makes use of several hyperparameters. Additionally, our model incorporates four more dense layers. For the purpose

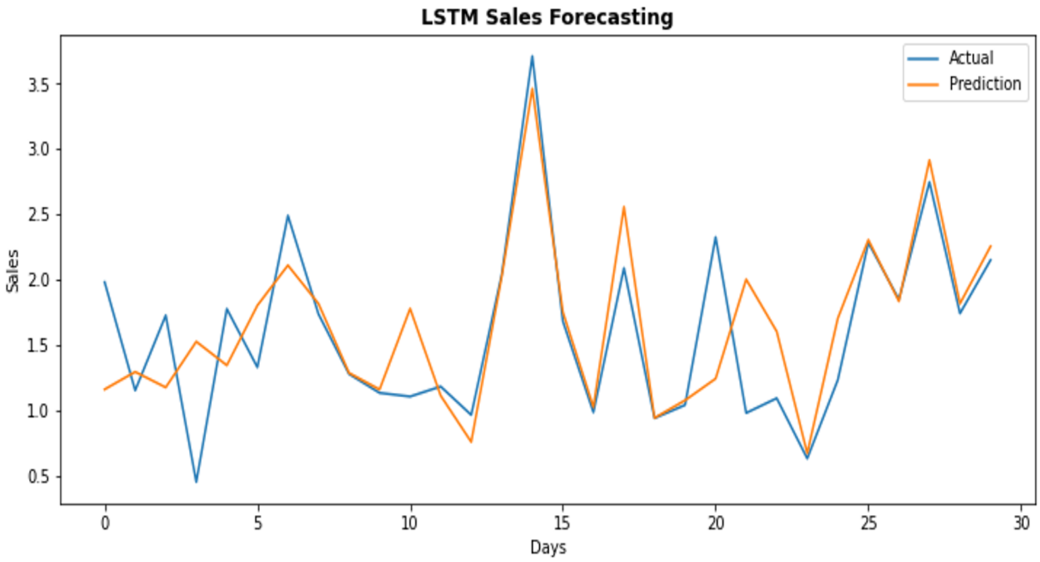

Figure 7. LSTM sales forecasting.

of sales forecasting, our LSTM model is configured as follows:

- N_input: 36 (3 Years Dataset) it is the number of observation for a model to use as input.
- N_units: 50 Number of LSTM units used in the hidden layer.
- N_epochs: 100 Number of times the full training dataset is made available.
- N_batch: 100 Number of samples per epoch after the weights are updated.
- N_diff: 12 Differencing order.

Figure 7 shows us the blue line that represents the actual sale, while the yellow line reveals the expected sale, for the entire month as obtained by the long LSTM model.

4.4 Convolutional Neural Network with Long short-term memory (CNN-LSTM) Results

The CNN-LSTM architecture combines the strengths of Convolutional Neural Networks for automatic feature extraction with Long Short-Term Memory networks for the analysis of sub-sequences in the input data. Each element serves a specific function: the CNN is responsible for extracting features, whereas the LSTM addresses the temporal relationships within the sub-sequences. Important configurations consist of 36 sub-sequence samples, 12 steps for each sub-sequence, and the incorporation of four dense layers, all fine-tuned to improve predictive accuracy.

Automatic feature extraction and learning from raw data is within the CNN model's capabilities. Here, we merge the CNN model's capabilities with that of the LSTM model, with the former handling feature extraction and the latter applying sub-sequences to input data. The CNN model uses LSTM to read the many sub-sequences gradually.

Each layer of our model CNN and an LSTM have its own unique set of parameters and purpose. The LSTM model makes predictions based on the CNN model's analysis of the input sub-sequences. In our proposed hybrid CNN-LSTM model, we have included an extra set of four dense layers.

For sales forecasting, we utilize the following configuration in our CNN-LSTM model:

- N_input: 36 (3 Years Dataset) Number of sub-sequence samples.
- N_steps: 12 (for 1 year) Number of step in each sub-sequence.
- N_filters: 64 Number of parallel filters used.
- N_kernel: 3 Number of steps that read in each input sequence.
- N_nodes: 100 Number of LSTM units used in the hidden layer.
- N_epochs: 200 Number of times make available the full model training dataset.
- N_batch: 100 used the number of samples for each epoch after updated the weights.

The number of filters, size of filters, and the number of layers on CNN as well as the type and parameters

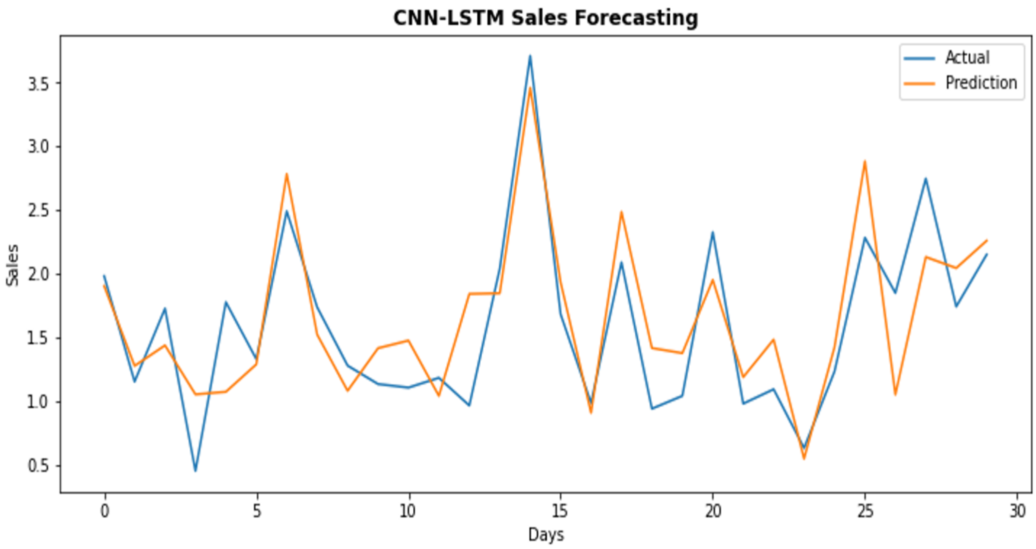

Figure 8. CNN-LSTM sales forecasting.

of LSTM layer these choices were informed by prior studies, prior experience and optimization. The number of sub-sequences was chosen to be 36 and the number of steps per sub-sequence was chosen to be 12 so as to indirectly capture certain temporal patterns that could exist in the sales data for multiple years. The number of filters N_filters = 64 and the size of the kernel, N_kernel = 3 were set with an intention to enable the CNN to learn features while not compromise the general computational resource requirements of the architecture. The LSTM units (N_nodes = 100) were established to address the temporal dependencies that sequence data comprises following other LSTM applications in time series forecasting. We therefore fixed N_epochs as 200 and N_batch as 100 after testing for the optimal number of epochs and batch size. These hyperparameters were optimized in order to-have the minimum RMSE and MAE, as presented in the output above.

Figure 8 represents the Convolutional Neural Network with Long Short-Term Memory's actual and expected sales for the entire month, with blue line representing actual sales and the yellow line representing the forecast sales.

4.5 Multilayer Perceptron (MLP) Results

MLP is a prediction method for time series that makes use of multiple observations. The output and input layers that make up an MLP are completely interconnected. Hidden layers in an MLP might have the same amount of inputs and outputs as the main layers. Many applications can benefit from MLPs, including image processing, machine translation, speech recognition, and time series forecasting.

We create MLPs with a single hidden layer that contains a certain amount of nodes. Additional performance enhancement is achieved by applying the rectified linear function to the hidden layer. For continuous value prediction in the output layer, we employ a linear activation function; similarly, the input layer contains a number of prior data. For an algorithmic speed boost, we employ four extra dense layers.

The following setup is used for sales forecasting in our MLP model:

- N_input: 36 (36 Months) Number of input employed in the model to account for data.
- N_nodes: 500 Number of nodes used in the hidden layer.
- N_epochs: 100 Number of times make available the full model training dataset.
- N_batch: 100 after revising the weights utilized the sample size for every epoch.

Figure 9 represents both the actual and predicted sales are depicted by the blue and yellow lines, respectively for the entire month, as obtained by the Multilayer Perceptron (MLP).

This study investigates various models for predicting sales, including CNN, LSTM, CNN-LSTM and MLP. These models have shown effectiveness in forecasting sales. In our future research, we plan to explore the use of predictive models like Random Forest and XGBoost to strengthen our results and offer a more comprehensive analysis of retail sales predictions.

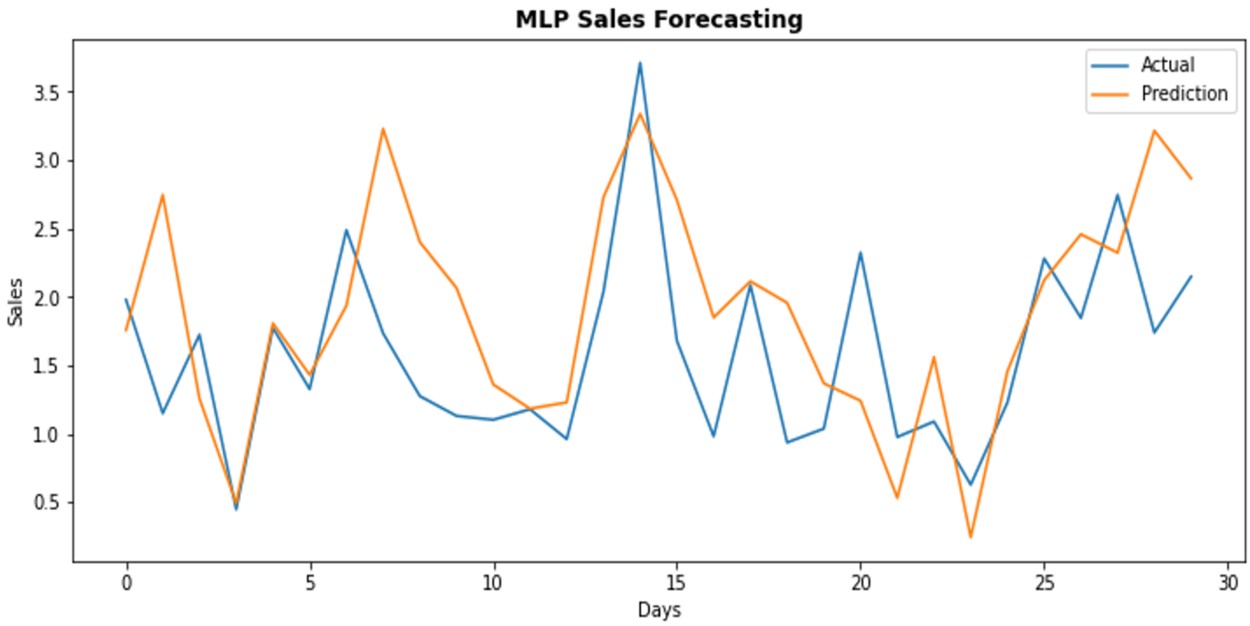

Figure 9. MLP sales forecasting.

Figure 10. Comparison of models for sales forecasting.

Table 4. Evaluation of models using MAE and RMSE metrics.

Index	RMSE	MAE
CNN	0.084	0.360
MLP	0.116	0.528
LSTM	0.073	0.325
CNN-LSTM	0.065	0.293

4.6 Comparison of all Models

We run the experiment on the sales dataset using various deep learning methods. Each model has its own unique setup, and we add four dense layers to each of them. For this, we run two tests: such as MAE and RMSE. After evaluating the performance of each deep learning model for sales forecasting, we determined that CNN-LSTM was the most effective and appropriate model for our retail sales dataset.

Figure 10 shows the comparison of several deep learning models in this study. Each line above as given in Figure 10 reflects a different model's output: in this case, blue for the actual sale, yellow for the CNN model, green for the LSTM model, red for the CNN-LSTM model, and purple for the MLP model. The CNN-LSTM model emerged victorious in our evaluation of deep learning options for use in sales forecasting. Model performance indicators, such as MAE and RMSE are used for evaluation and their values are presented in the following Table 4.

Table 4 displays how well the deep learning models performed when set up with the default parameters and with minimal tweaking. We can see that CNN-LSTM, or Convolutional Neural Network with Long Short-Term Memory, did well on both the RMSE and MAE metrics as shown in this table. When compared to models using Convolutional Neural Networks, Multilayer Perceptron, and Long Short-Term Memory, CNN-LSTM produces the most accurate sales forecasts.

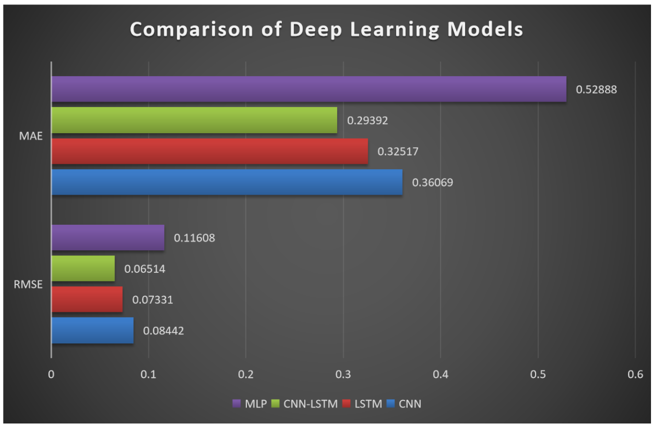

Figure 11. Comparison of deep learning models.

Figure 11 and Table 5 represents a variety of deep learning models' average absolute and root-mean-squared mistakes (e.g., CNN, LSTM, CNN-LSTM, and MLP) produce. This shows that the CNN-LSTM model does a good job of predicting future sales. This is what the performance evaluation metrics came up with. Based on our comparisons with other deep learning algorithms, we found that CNN-LSTM produced the lowest error rate.

4.7 Comparison with Other Studies

This study evaluates deep learning models for pharmaceutical retail sales forecasting. Direct numerical comparison with previous studies is challenging due to differences in datasets, scales, and normalization methods. Nevertheless, the proposed CNN-LSTM hybrid model demonstrates strong performance on the pharmaceutical dataset. As opposed to the studies of [23–25] that employed the individual models, the proposed model integrated CNN and LSTM, and produced superior feature learning and temporal property learning. This integration helped our proposed CNN-LSTM hybrid model achieved the lowest RMSE (0.06514) and MAE (0.29392) values among all models evaluated in this study, demonstrating superior accuracy for pharmaceutical retail sales forecasting.

This table represents the breakthroughs and superior outcomes of our technique by comparing our findings with three important researches in the area of sales forecasting.

The findings of the present study show that the proposed CNN-LSTM model improves retail sales forecasting in Pharmaceutical industry to a large content compared to other deep learning models. Both CNN and LSTM as a hybrid model overcome the issue of overfitting and training errors in the sequential data analysis, and therefore predicted values are very close to the actual values. Thus, the proposed CNN-LSTM model helps professionals and decision makes in Pharmaceutical industry in more accurate sales forecasting, reducing supply chain cost, capacity building, and improving happiness of customers.

4.8 A Study Case

As an example of a real-world application of the proposed hybrid CNN-LSTM model, we consider an actual dataset from a pharmaceutical company, where daily sales are predicted based on historical data. CNN is used for feature extraction in this case we extract features like sales trends and patterns while the LSTM handles the time series aspect of the data. The model also detects cycles like daily sales and weekly sales, then filters out noises that are irregular sales activities. This paper shows that this hybrid model is more accurate in forecasting daily sales over a one-year period than the traditional forecasting methods, with a significant decrease in the forecasting error rate. Since the temporal dynamics of sales data differ, the CNN-LSTM model was able to respond to the temporal variability and generate accurate sales predictions, which demonstrated the model's flexibility in real-time sales settings.

5 Conclusion

Our study highlights the critical role of accurate sales forecasting for effective inventory management, customer service, and production planning. By evaluating various deep learning models, including CNNs, MLPs, LSTMs, and CNN-LSTMs, this research found that the CNN-LSTM model delivered the most precise predictions with the lowest error rates. It is observed that by combining CNN and LSTM models, the predictive accuracy of the model improves and the

Table 5. Comparative analysis with other studies.

Study	Model	RMSE	MAE	Key Features and Limitations
[23]	LSTM	0.145	0.112	LSTM for e-commerce sales demand forecasting on Walmart dataset. Different dataset and scale.
[24]	Multilayer Perceptron (MLP)	0.152	0.478	We adopted this MLP method and conducted prediction experiments on pharmaceutical retail sales data.
[25]	CNN	0.08442	0.36069	CNN for sales forecasting (different time horizons). Results reported in absolute sales units, not normalized.
Our Study	CNN-LSTM Hybrid	**0.06514**	**0.29392**	Hybrid CNN-LSTM with additional dense layers on pharmaceutical retail daily sales data.

error rate decreases, due to that predictive ability of the proposed method improves and vulnerability to overfitting is minimized. Metrics such as MAE and RMSE are used in this study. Results of these metrics as (RMSE=0.06514, MAE=0.29392) for CNN-LSTM are better than other models, namely CNNs, MLPs, and LSTM.

For future works, we recommend focusing on enhancing model performance through advanced techniques like data augmentation, attention mechanisms, and ensemble learning. Integrating external data sources and automating model updates will further improve forecasting accuracy and relevance. Emphasizing scalability, efficiency, and user-friendliness will ensure that these advanced forecasting solutions are practical and beneficial for retail management.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Gandhi, M. A., Maharram, V. K., Raja, G., Sellapaandi, S. P., Rathor, K., & Singh, K. (2023, July). A novel method for exploring the store sales forecasting using fuzzy Pruning LS-SVM approach. In *2023 2nd International Conference on Edge Computing and Applications* (*ICECAA*) (pp. 537-543). IEEE. [CrossRef]

[2] de Castro Moraes, T., Yuan, X. M., & Chew, E. P. (2024). Hybrid convolutional long short-term memory models for sales forecasting in retail. *Journal of Forecasting*, *43*(5), 1278-1293. [CrossRef]

[3] Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, *163*, 120480. [CrossRef]

[4] Eglite, L., & Birzniece, I. (2022). Retail sales forecasting using deep learning: Systematic literature review. *Complex Systems Informatics and Modeling Quarterly*, (30), 53-62. [CrossRef]

[5] Hasan, M. R., Kabir, M. A., Shuvro, R. A., & Das, P. (2022). A comparative study on forecasting of retail sales. *arXiv preprint arXiv:2203.06848*. [CrossRef]

[6] Rohaan, D., Topan, E., & Groothuis-Oudshoorn, C. G. (2022). Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. *Expert systems with applications*, *188*, 115925. [CrossRef]

[7] Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283-1318. [CrossRef]

[8] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889. [CrossRef]

[9] Giri, C., & Chen, Y. (2022). Deep learning for demand forecasting in the fashion and apparel retail industry. *Forecasting*, *4*(2), 565-581. [CrossRef]

[10] Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C., & Oyewole, A. T. (2024). Real-time data analytics in retail: A review of USA and global practices. *GSC Advanced Research and Reviews*, *18*(3), 059-065. [CrossRef]

[11] Chopra, M., Chopra, R., Reddy, R., & Chopra, S. (2023). Leveraging LSTM Neural Networks and ARIMA Models for Enhanced Real-Time Sales Forecasting in Dynamic Retail Environments. *Journal of AI ML Research*, *12*(2), 1-25. https://joaimlr.com/index.php/v1/a

rticle/view/19

[12] Kaneko, Y., & Yada, K. (2016, December). A deep learning approach for the prediction of retail store sales. In *2016 IEEE 16th International conference on data mining workshops (ICDMW)* (pp. 531-537). IEEE. [CrossRef]

[13] Ahmadov, Y., & Helo, P. (2023). Deep learning-based approach for forecasting intermittent online sales. *Discover Artificial Intelligence, 3*(1), 45. [CrossRef]

[14] Xu, W., Cao, Y., & Chen, R. (2024). A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce. *Decision Support Systems, 177*, 114104. [CrossRef]

[15] Li, Q., & Yu, M. (2023). Achieving sales forecasting with higher accuracy and efficiency: A new model based on modified transformer. *Journal of Theoretical and Applied Electronic Commerce Research, 18*(4), 1990-2006. [CrossRef]

[16] Rao, F. A., Muneer, A., Almaghthawi, A., Alghamdi, A., Fati, S. M., & Ghaleb, E. A. A. (2023). BMSP-ML: big mart sales prediction using different machine learning techniques. *IAES International Journal of Artificial Intelligence*, 12(2), 874.

[17] Ratre, S., & Jayaraj, J. (2023, January). Sales prediction using arima, facebook's prophet and xgboost model of machine learning. In *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021* (pp. 101-111). Singapore: Springer Nature Singapore. [CrossRef]

[18] Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications, 201*, 116993. [CrossRef]

[19] Joseph, R. V., Mohanty, A., Tyagi, S., Mishra, S., Satapathy, S. K., & Mohanty, S. N. (2022). A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting. *Computers and Electrical Engineering, 103*, 108358.. [CrossRef]

[20] Loureiro, A. L., Miguéis, V. L., & Da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems, 114*, 81-93. [CrossRef]

[21] Burinskiene, A. (2022). Forecasting model: the case of the pharmaceutical retail. *Frontiers in Medicine, 9*, 582186. [CrossRef]

[22] Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2018, December). Sales-forecasting of retail stores using machine learning techniques. In *2018 3rd international conference on computational systems and information technology for sustainable solutions (CSITSS)* (pp. 160-166). IEEE. [CrossRef]

[23] Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019, December). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In *International conference on neural information processing* (pp. 462-474). Cham: Springer International Publishing. [CrossRef]

[24] Karoński, A., Hernes, M., Walaszczyk, E., & Rot, A. (2024, April). Forecasting Sales at Fuel Stations Using a Multilayer Perceptron. In *Asian Conference on Intelligent Information and Database Systems* (pp. 206-218). Singapore: Springer Nature Singapore. [CrossRef]

[25] Amir, W. K. H. W. K., Soom, A. B. M., Jasin, A. M., Ismail, J., Asmat, A., & Rahman, R. A. (2023). Sales forecasting using convolution neural network. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 30*(3). [CrossRef]

Rana Shakeel is a Master's student in the Department of Computer Science at Lahore Leads University. He is currently working in a private-sector organization in Lahore, Pakistan. His areas of interest include machine learning, deep learning, sales, and marketing.

Muhammad Hasnain is currently serving as an Associate Professor in Computer Science department, Lahore Leads University, Pakistan. He finished in PhD from Monash University Australia. His areas of interest include, machine learning, deep learning, blockchain technology, performance testing, and Large language models. He has published several research articles in reputed journals and conferences.

Muhammad Hamza Mahmood is currently working as a lecturer in computer science department at Lahore Leads University, Pakistan. His areas of interest include as machine learning, deep learning, and computer vision.

Muhammad Abid Mehmood is currently working as an IT administrator and teaching in Computer Science department at Lahore Leads University, Pakistan. His areas of interest include as networking, machine learning, and blockchain technology.