Intelligent Methods in Engineering Sciences 2(4): 102-114 2023

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International Open Access

December, 2023

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2023.70

# Predicting Future Demand Analysis in the Logistics Sector Using Machine Learning Methods

***Erhan HAYTA*** [a], ***Bunyamin GENCTURK*** [b], ***Cuneyt ERGEN*** [a], ***Murat KOKLU*** [c, *]

[a] *ALISAN Logistic Inc., Istanbul, Türkiye*
[b] *Technology Faculty, Selcuk University, Konya, Türkiye*
[c] *Department of Computer Engineering, Selcuk University, 42250 Selcuklu, Konya, Türkiye*

ARTICLE INFO

*Article history:*
Received 10 November 2023
Accepted 28 December 2023

*Keywords:*
Data analysis and prediction,
Forecasting future needs,
Logistics sector,
Machine learning,
Time series forecasting

ABSTRACT

In this study, the potential of machine learning methods for analyzing future needs in the logistics sector was investigated. The research is conducted using the MATLAB platform. Numeric pallet demand data obtained from a logistics company are employed to train MLP, LSTM, and CNN models. Data security and confidentiality take priority during the data collection process. This dataset, comprising a total of 3,062 daily records, serves as the primary data source for the study. In the data preprocessing phase, missing or erroneous data is rectified, and outliers are detected and corrected. The models are tested to predict pallet quantities over periods of 25 days and 4 weeks. The results are evaluated by comparing the model predictions with actual data. Model performances are assessed using metrics such as MSE, RMSE, NRMSE, MAE, ESD, and RC. The outcomes of the last 25 days demonstrate that the LSTM model exhibits the lowest MSE (6,410.5571) and RMSE (80.0660) values. For the MLP model, the MSE value is calculated as 20,536.5564, and the RMSE value is 143.3058. Performance evaluations for the CNN model yield an MSE of 8,492.4297 and an RMSE of 92.1544. Furthermore, it is observed that the MLP model provides the best results for the 4-week forecasts. The results of this study indicate the success of the models used for predicting pallet transportation quantities in the logistics sector. In addition to this study, a contribution is made toward enabling logistics companies to make more informed and strategic decisions.



## 1. INTRODUCTION

The logistics sector is an industry that manages the delivery of goods and services from the source to the consumer. This process involves the procurement, storage, transportation, and distribution of materials. Logistics aims to deliver the right product to the right place at the right time and at the right cost to customers. In today's context, the logistics sector operates within a rapidly changing, innovation-driven, and competitive landscape. The rapidly growing e-commerce sector, increasing consumer demands, and the complexity of global supply chains pose challenges for logistics companies to effectively manage their operational activities. For this reason, logistics firms require new and advanced analytical tools to efficiently utilize resources, accurately forecast demand, and predict future needs. The dynamic nature of the industry underscores the importance of implementing these tools for effective decision-making and strategic planning [1-3].

Machine learning possesses the capability to analyze vast amounts of data, identify trends, recognize patterns, and make predictions about the future. It holds significant potential for predicting future needs in the logistics sector. For instance, within the logistics industry, companies can make better decisions in critical areas such as demand forecasting, inventory management, and optimizing logistics operations by leveraging machine learning techniques. When machine learning is applied to demand forecasting, it enables accurate fulfillment of customer demands and optimization of inventory management. Furthermore, machine learning methods are utilized to efficiently plan logistics operations, improve route optimization, and optimize delivery timelines, enhancing overall operational efficiency. Through these applications, machine learning plays a pivotal role in enabling logistics companies to meet demands effectively and utilize resources efficiently in an ever-evolving industry [4-7].

Support vector machines, decision trees, artificial neural networks, and genetic algorithms are among the widely used machine learning methods in the logistics sector. Each method comes with its own set of advantages and disadvantages, and its applications are tailored to meet the diverse needs of the industry. Tailoring the choice of methods to fit the specific requirements and data

*** Corresponding Author:** mkoklu@selcuk.edu.tr

conditions of each logistics operation is crucial. By doing so, logistics companies can make more informed and strategic decisions, facilitating the prediction of future demands. Utilizing machine learning methods for forecasting future demand analysis holds significant importance in the logistics sector, enabling companies to gain a competitive advantage and enhance operational efficiency. Through these methods, the prediction of future demand analysis becomes easier, contributing to the overall success of logistics businesses [6, 8-10].

## 2. Related Works

Gils et al. (2017) focus on situations that facilitate the delivery of belated orders to warehouses promptly, contributing to the differentiation of warehouse customer services from competitors. Warehouses are facilities where order receipt, product stocking, order picking, and shipping activities take place. The study underscores the significance of workload forecasting in companies. Time series models are employed to predict the number of daily order lines. Three different evaluation metrics, namely Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD) are utilized in the study [11].

Chan et al. (2019) presented several time series forecasting methods, including machine learning-based approaches like Support Vector Regression, among others. These prediction methods are employed with concurrent secondary data to forecast the container shipping capacity of a port. Data from the previous year's transportation capacity is used to predict the transportation capacity of the year following the year in which the study is conducted. The study employs Support Vector Regression (SVR) and Artificial Neural Network (ANN), both of which are popular data mining techniques today. Five time series methods are utilized to predict the container shipping capacity of a port, and their performances are compared [12].

Talupula (2019) employed machine learning methods to enhance the long-term volume forecasts of logistics service providers. In an experiment conducted on the acquired dataset, the performance of Convolutional Neural Network (CNN), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) models was evaluated. The dataset comprises 98 months of data with 14 variables, including time and total demand characteristics of these variables. 75% of the data was used for training, 10% for validation, and 15% for testing. Upon examining the results, it is observed that these models exhibit similar predictive performance. The CNN method is stated to be more effective in predicting the demands of products distributed towards the output [13].

Sohrabpour et al. (2021) underline the significance of sales forecasting in production and supply chain management. Due to the limitations of traditional forecasting methods, the use of causal forecasting methods is recommended. The accuracy of predictions made using real data is also examined within the scope of this study. Artificial intelligence techniques have been favored in the study to enhance the effectiveness of these forecasts. To evaluate the quality of the causal forecasting model, four different error measurement metrics have been employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared, and the Correlation Coefficient [14].

Deng & Liu (2021) emphasizes the significance of emerging intelligent applications such as artificial intelligence and Internet of Things (IoT) in supply chain management and communication. The study highlights that effective and predictable inventory management throughout the lifecycle of the supply chain plays a critical role in cost reduction. The research aims to optimize inventory management using deep learning methods. By employing a mathematical model, the inventory management (IM) process is formulated, with the goal of minimizing logistic costs. Based on this model, a deep inventory management (DIM) method utilizing Long Short-Term Memory (LSTM) is proposed. It's noted that this method achieves high accuracy in inventory demand forecasting compared to other approaches [15].

Ribeiro et al. (2022) compare the performance of three machine learning models (SVR, Random Forest, and XGBoost), three deep learning models (RNN, LSTM, and GRU), and a classical time series model ARIMA for predicting daily energy consumption. The study utilizes a dataset comprising 8040 records over an 11-month period from an unrefrigerated logistics facility in Ireland. The best configurations for each model were determined using a grid search method. XGBoost models are noted to exhibit superior performance in very short-term and short-term load forecasting compared to other models [16].

## 3. Materials and Method

Machine learning methods have been employed for forecasting future needs in the logistics sector. In this study, the developed predictive model will contribute to more efficient decision-making in operational processes of logistics companies. MATLAB R2022b has been chosen for data analysis and model training. The flowchart depicting the overall workflow of the study is presented in Figure 1. The generated dataset, the utilized methods, and the performance evaluation metrics of the models are detailed in this section.

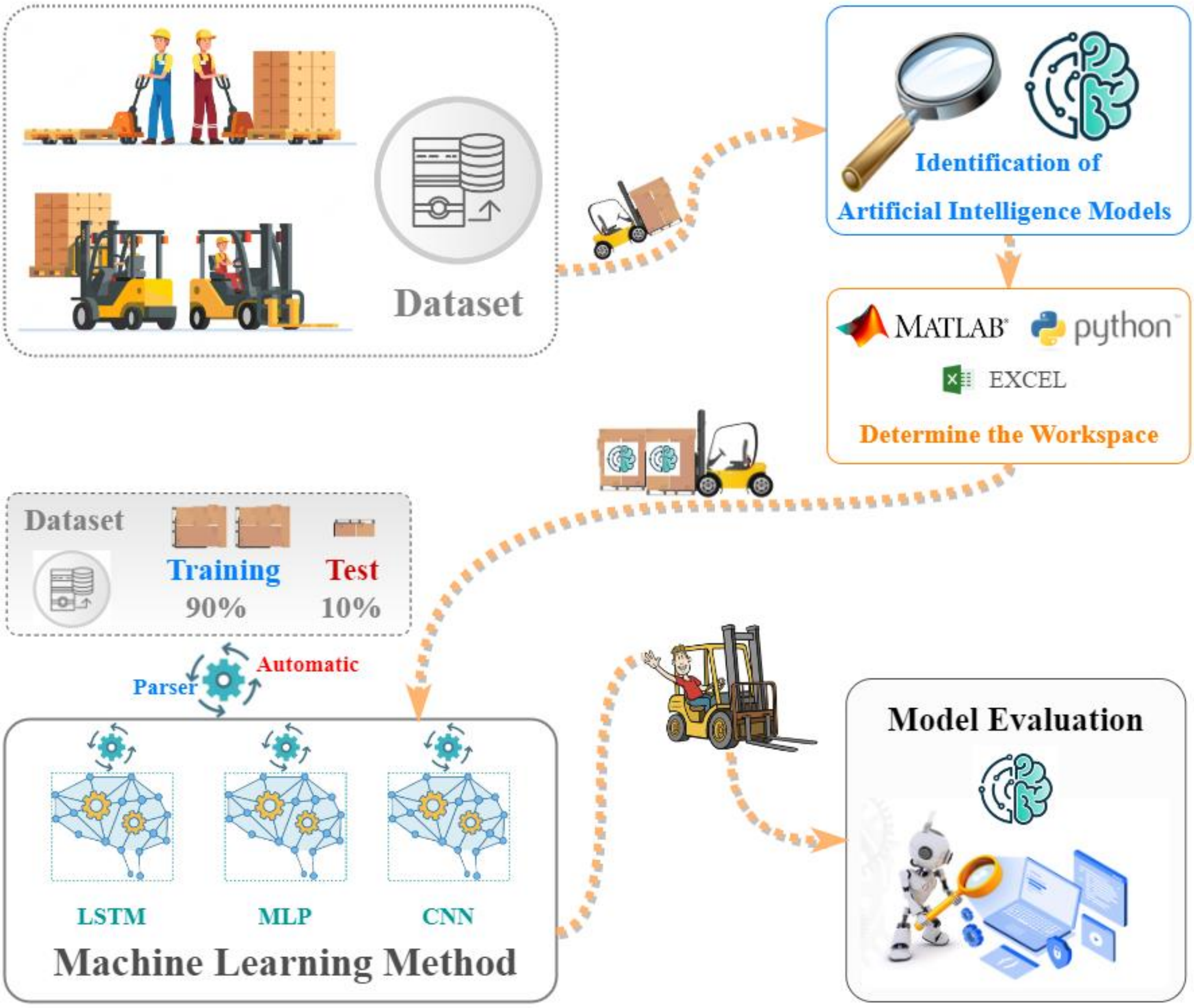


**Figure 1.** General Workflow Diagram

### *3.1. Dataset and Data Preprocessing*

The data obtained from the operational databases of the logistics company has been utilized as the dataset for this study. The dataset encompasses various variables, including customer demands, inventory levels, lead times, shipment details, geographic and weather conditions. During the data collection process, utmost care has been taken regarding data security and privacy. The dataset contains a total of 3,062 daily records spanning from September 5, 2012, to September 27, 2022. Daily data includes information such as the total number of pallets transported in a day. Despite having approximately 10 years of data available from the supplying firm, 3,062 daily records are in a usable format for this study. This is due to data integrity issues arising from the COVID-19 pandemic and the inability to conduct transportation operations during official and religious holidays. The presence of missing data on days where linearity is not maintained adversely affects the prediction performance of the models.

***Data Preprocessing:*** The collected data has been prepared for analysis. During the data preprocessing phase, missing or erroneous data has been corrected, outliers have been identified, and variable types have been transformed when necessary. The dataset has been structured in a suitable format for automatic allocation into 90% training and 10% testing subsets before model training. Furthermore, the dataset has been formatted appropriately for time series analysis. Figure 2 provides some summary information about the dataset used in the study.

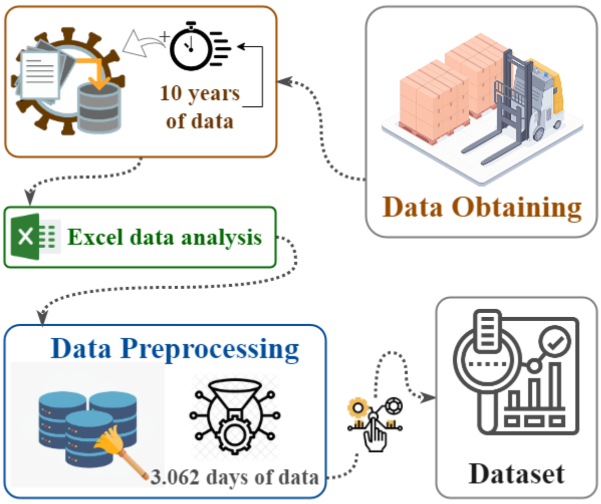


**Figure 2.** Dataset Creation Process

### *3.2. Long short-term memory (LSTM)*

Long Short-Term Memory (LSTM) is an artificial neural network utilized in fields such as machine learning, natural language processing, and deep learning. It boasts a more advanced architecture compared to traditional Recurrent Neural Networks (RNNs). LSTM is particularly successful in handling time series data or datasets with text-based dependencies. It was designed to address issues arising in situations where utilizing historical information from the past is challenging.

LSTM is especially preferred in applications like time series forecasting, text generation, language models, speech recognition, and translation. Its key feature lies in utilizing a cell state and three distinct gate mechanisms. These mechanisms, as depicted in Figure 3, are termed the forget gate, input gate, and output gate. The forget gate determines which pieces of information will be discarded from the cell state. The input gate manages the addition of new information to the cell state. The output gate, on the other hand, dictates how much of the updated cell state will be used as output. LSTM can effectively track long-term dependencies and retain important information through these mechanisms [13, 17, 18].

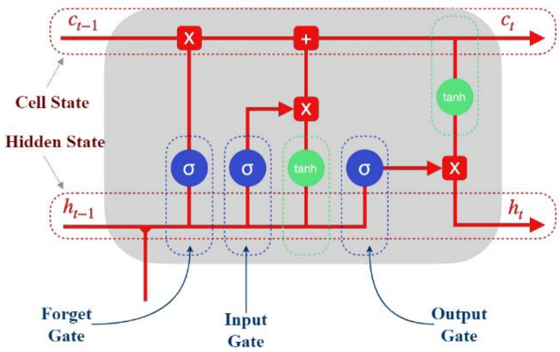


**Figure 3.** The general working structure of LSTM

### *3.3. Multilayer Perceptron (MLP)*

Multilayer Perceptron (MLP) is a widely used artificial neural network-based machine learning model and forms one of the fundamental building blocks of deep learning. MLP consists of at least one input layer, one or more hidden layers, and an output layer. As seen in Figure 4, each layer of the multi-layer perceptron contains one or more neurons. Input data passes through the layers via weights and activation functions, ultimately producing an output. It can be employed for various machine learning tasks, including classification, regression, pattern recognition, and time series forecasting [19-22]. Some advantages of MLP include:

***Universal Approximation:*** In theory, MLP can approximately compute any function with a sufficient number of hidden layers and neurons.

***Depth and Complexity:*** By adding multiple hidden layers, depth and complexity can be introduced, enabling the modeling of more intricate patterns and relationships.

***Learning and Generalization:*** Through the backpropagation algorithm, MLP learns its weights and threshold values, enabling it to recognize patterns in the dataset and generalize to new data.

***Feature Extraction:*** MLP can automatically extract essential features from input data. This reduces the need for extensive feature engineering, allowing the model to learn more complex features.

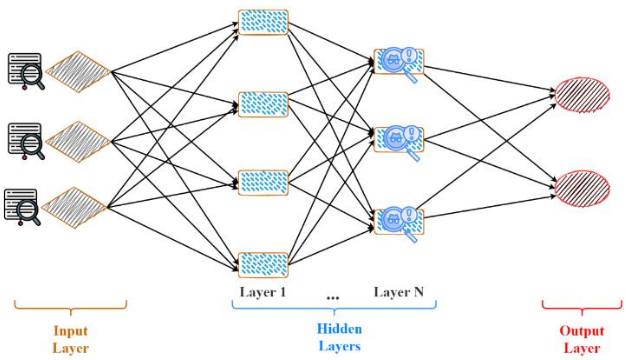


**Figure 4.** The General Structure of MLP

MLP is effective for modeling nonlinear relationships in general. It is commonly used in fields such as natural language processing, image processing, audio processing, and financial analysis. Additionally, due to being a fundamental building block in deep learning model designs, it is also preferred for solving complex tasks and utilizing large deep neural networks [23-25].

### *3.4. Convolutional Neural Network (CNN)*

CNN is a deep learning model based on artificial neural networks. It is particularly effective for processing and analyzing visual data. It is commonly used for tasks like image classification, object detection, face recognition, and image segmentation. Based on the idea that data can create local patterns and features together, CNN takes into account the structure and local dependencies of the data. CNN has a specialized structure, especially suitable for processing images, with convolution and pooling layers that play a key role in its architecture [21, 23, 26-30]. The structure of the CNN model is provided in Figure 5.

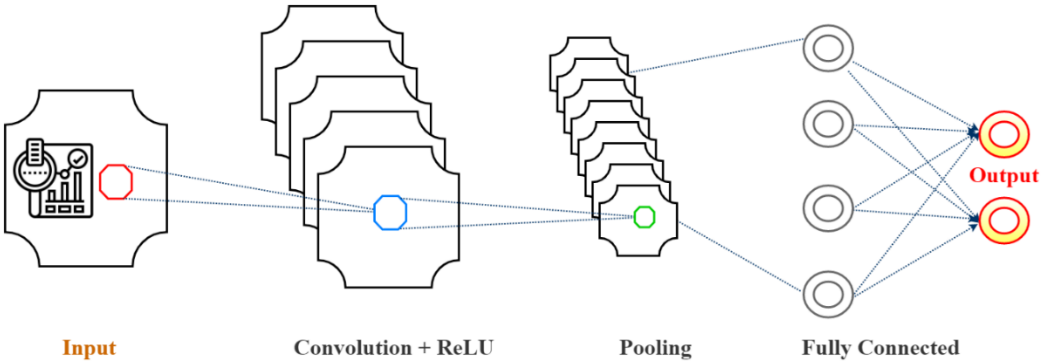


**Figure 5.** The general working structure of CNN

The main components of CNN are as follows: [29, 31, 32]:

Convolutional Layers: Feature maps are created by applying a series of filters to the input data. Each filter is defined by learned weights to detect specific patterns. The convolution process involves filtering over local regions of the data, extracting local features.

**Activation Functions:** After each convolutional layer, an activation function (usually Rectified Linear Unit - ReLU) is applied to the obtained feature maps. The activation function compresses outputs, aiding the network in forming more general and learnable representations.

**Pooling Layers:** Following the convolutional layers, pooling is applied to reduce dimensions and preserve the originality of feature maps. Max pooling is commonly used, selecting the largest features to summarize data and reduce dimensionality.

**Fully Connected Layers:** CNN concludes with one or more fully connected layers, transforming feature maps into higher-level features for predicting output classes.

CNN focuses on detecting local patterns and features in data while progressively learning higher-level features. Convolutional layers create feature maps by filtering data, activation functions are applied to these maps for more general representations, and pooling layers reduce dimensions by selecting key features. Finally, fully connected layers transform feature maps into higher-level features for class predictions. With these features, CNN excels in processing and analyzing visual data. Convolution and pooling layers, along with activation functions and fully connected layers, form the core components of CNN. This enables the extraction of features from visual data and the creation of higher-level representations [30, 31, 33-38]. In the study, the ResNet50 architecture, commonly used and stable in literature, was chosen for training the machine learning methods.

### *3.5. ResNet-50*

ResNet50, a commonly used deep learning architecture, is a convolutional neural network (CNN) consisting of a total of 50 layers. It was proposed by Kaiming and his colleagues in 2015. This architecture includes 48 convolutional layers, 1 max pooling layer, and 1 medium pooling layer. The ResNet architecture enables CNNs to work with multiple layers. Deep neural networks with multiple layers have higher training error percentages compared to models with fewer layers. The relatively moderate number of layers in the ResNet50 architecture offers significant advantages during model training. The ResNet architecture is typically used in object detection, segmentation, and data labeling domains [39-42].

### *3.6. Performance Metrics*

Performance metrics are criteria used to evaluate how well a model is performing. This feature allows for the objective measurement, comparison, and improvement of the quality of a trained artificial intelligence model. It also enables the assessment of the accuracy, precision, predictive power, and consistency of the model's predictions. These metrics are preferred in model selection, hyperparameter tuning, and comparing different models [43, 44]. Some common performance metrics used in prediction with artificial intelligence models include:

MSE (Mean Squared Error): is used to measure how well predictions align with actual values. It calculates the average of the squared differences between predicted values and actual values. The formula for calculating MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

**n:** is the total number of data points,
**$y_i$:** represents the actual values,
**$\hat{y}_i$:** represents the predicted values.

The Mean Squared Error (MSE) is a commonly used performance metric that measures the average of the squared differences between predicted and actual values. It tends to emphasize larger error values and can be affected by outliers. As the MSE value approaches zero, it indicates that the predictions are closer to the actual values. MSE is often used as a loss function in optimization algorithms, aiming to find a model or predictor that minimizes its value [45, 46].

***RMSE (Root Mean Squared Error):*** The Root Mean Squared Error (RMSE) is a metric used to measure the agreement between predictions and actual values. Similar to MSE, it indicates how far the predictions are from the actual values. However, RMSE is obtained by taking the square root of MSE, making RMSE values interpretable in the original data units. The formula for calculating RMSE is given by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_{i)}{}^2}{n}} \qquad (2)$$

RMSE measures how far predictions are from the actual values and places greater emphasis on larger error values. It's particularly useful for calculating the standard deviation of error values or comparing the performance of different models. A lower RMSE value indicates better model performance. As the result value approaches zero, it signifies that predictions are closer to the actual values. One disadvantage of RMSE is that it can be influenced by outliers, impacting the RMSE value [45-47].

***NRMSE (Normalized Root Mean Square Error):*** NRMSE, or Normalized Root Mean Squared Error, is a normalized version of RMSE. It's used to prevent the RMSE metric from being affected by data-dependent scaling. NRMSE is preferred to assess how well prediction errors perform based on the variation of the actual values. NRMSE is calculated by dividing RMSE by the range or standard deviation of the actual values. This allows for a fairer comparison of models with different ranges or scales of actual values in different datasets. NRMSE values are often expressed as percentages (%) [45, 48, 49].

here is the formula for calculating Normalized Root Mean Squared Error (NRMSE):

$$RMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{i)}{}^2}}{y_{max} - y_{min}} \times 100 \qquad (3)$$

**$y_{max}$:** is the maximum value of the actual values,

**$y_{min}$:** is the minimum value of the actual values.

***MAE (Mean Absolute Error):*** MAE is a commonly used error metric in the fields of artificial intelligence, machine learning, and statistics. It is used to measure how far predictions are from the actual values. It is often preferred when evaluating the performance of regression models. Mean Absolute Error (MAE) is calculated by taking the average of the absolute values of errors. This metric treats the magnitudes of errors equally and balances the impact of outliers. The MAE value represents the average distance between predictions and actual values.

The calculation formula for MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|y_i - \hat{y}_i| \qquad (4)$$

During the training process, the model's parameters are adjusted to minimize this loss function. This minimization process is typically achieved using gradient descent algorithm. Gradient descent computes the derivatives of the loss function with respect to the parameters and makes updates to reduce the amount of error. As the training progresses, a decrease in the MAE value indicates that the model is making more accurate predictions [47, 50, 51].

***ESD (Error Standard Deviation):*** It represents the standard deviation of the errors generated by the artificial intelligence model's predictions. In the fields of machine learning and statistics, it's a concept used to measure how much deviation the predictions of a model exhibit from the actual values. Standard deviation signifies how much the values in a dataset spread out from the mean. When ESD (Error Standard Deviation) is calculated by computing the differences between a model's predictions and the actual values, it represents the standard deviation of these differences. A higher error standard deviation indicates that the predictions have greater variability from the actual values, while a lower error standard deviation indicates that the predictions are closer to the actual values. A low ESD result suggests that the model's predictions are generally closer to the actual values and make more consistent predictions. On the other hand, a higher error standard deviation may indicate that the model's predictions are more variable and make more errors [52-54].

***RC (Rank Correlation):*** It's a concept used in statistical analysis. Rank correlation is used to measure how similar or different the relationship between two variables is in terms of ranking. This measurement is based on the order of rankings rather than the exact values of the variables. Rank correlation is used instead of the parametric correlation measure, Pearson correlation coefficient. While Pearson correlation is based on the assumption that variables exhibit a linear relationship, rank correlation does not make this assumption. Therefore, rank correlation is preferred to evaluate the relationship between variables in a more flexible manner. Popular rank correlation coefficients like Spearman's Rho and Kendall's Tau are used to measure different types of relationships. Spearman's Rho is used when there is a monotonic relationship between variables, while Kendall's Tau measures the ranking-level relationship between variables and doesn't require the assumption of monotonicity. Rank correlation is often involved when dealing with ordered data, such as exam scores, survey rankings, or grading data [55-58].

## 4. Experimental Results

In the training of time series classification models, LSTM, MLP, and CNN, all the available data was utilized. For the training of LSTM and CNN models, the ResNet50 architecture was employed. The training process was executed with a maximum of 100 epochs and a mini-batch size of 16. The model's learning rate was set to 0.001, and the Adam optimization method was employed. The model consists of 2 hidden layers and employs a dropout value of 0.5. These parameters encapsulate the values utilized in configuring an LSTM and CNN-based model. The ResNet50 architecture constitutes the fundamental structure of the model, while the training process was carried out with a specific number of epochs and a designated mini-batch size. The learning rate and optimizer guide the updating and optimization process of the model. The parameter values for the LSTM and CNN models are presented in Table 1.

**Table 1.** The required parameter values for LSTM and CNN models

| Parameter | Value |
|---|---|
| Model | ResNet50 |
| Max. Epochs | 100 |
| Mini Batch Size | 16 |
| Learning Rate | 0.001 |
| Optimizer | adam |
| Hidden Layers | 2 |
| Dropout Value | 0.5 |

The hidden layers within the model represent higher-level features of the data, while the dropout value is employed to mitigate the issue of overfitting. These parameters can significantly impact the model's performance and learning capability. Consequently, the selection of accurate parameters is a crucial factor that can influence the success of the model. Information regarding the values used for configuring the MLP model is provided in Table 2. The number of layers and neurons in each layer has been determined for the model. The training function governs how the model is updated and optimized, while the maximum iteration count controls the duration of the training process.

**Table 2.** The necessary parameter values for the MLP model

| Parameter | Value |
|---|---|
| Number of Feed Forward Layers | 2 |
| Number of Neurons in First Layer | 15 |
| Number of Neurons in Second Layer | 10 |
| Number of Neurons in Third Layer | 10 |
| Train Function | trainlm |
| Max. Iteration | 300 |

According to the information in Table 2, the parameters of the MLP model are as follows: The model consists of a total of 3 feedforward layers. The first layer contains 15 neurons, the second layer contains 10 neurons, and the third layer also contains 10 neurons. The training process employs the "trainlm" training function, with a maximum iteration limit set to 300. These parameters significantly influence the behavior and performance of the MLP model. The number of layers and neurons contribute to determining the model's representational capacity and learning capability.

Following the training process, predictions were made using two different approaches: for the last 25 days and for a span of 4 weeks. The original data for the last 25 days and the model predictions are presented in Table 3.

**Table 3.** The prediction of the pallet count for the last twenty-five days

| Number of Data | Last 25 Days | MLP | LSTM | CNN |
|---|---|---|---|---|
| 1 | 501 | 377.5 | 431.5 | 383.0 |
| 2 | 707 | 486.5 | 446.0 | 663.0 |
| 3 | 507 | 399.0 | 354.0 | 625.0 |
| 4 | 427 | 411.5 | 415.0 | 480.0 |
| 5 | 110 | 197.0 | 121.5 | 307.5 |
| 6 | 591 | 427.0 | 521.5 | 482.5 |
| 7 | 586 | 526.0 | 522.0 | 616.5 |
| 8 | 463 | 366.0 | 473.0 | 413.0 |
| 9 | 611 | 444.5 | 622.0 | 713.5 |
| 10 | 361 | 319.0 | 356.5 | 646.5 |
| 11 | 110 | 223.5 | 191.5 | 424.5 |
| 12 | 513 | 471.5 | 454.0 | 438.0 |
| 13 | 614 | 540.5 | 619.5 | 746.5 |
| 14 | 580 | 400.5 | 536.5 | 685.5 |
| 15 | 991 | 548.0 | 632.0 | 850.5 |
| 16 | 559 | 521.0 | 430.0 | 728.5 |
| 17 | 253 | 349.5 | 206.0 | 536.0 |
| 18 | 621 | 502.5 | 551.5 | 689.0 |
| 19 | 523 | 425.0 | 494.5 | 647.5 |
| 20 | 691 | 532.0 | 585.0 | 732.0 |
| 21 | 810 | 596.5 | 477.5 | 851.0 |
| 22 | 644 | 512.0 | 555.5 | 743.0 |
| 23 | 154 | 244.5 | 300.0 | 476.0 |
| 24 | 536 | 426.0 | 480.0 | 692.0 |
| 25 | 413 | 402.5 | 480.0 | 658.5 |

According to Table 3, the LSTM model has shown the closest predictions to the actual results for the days 1, 4, 6, 8, 9, 10, 11, 13, 14, 17, 18, 19, 22, and 24. On days 3, 5, 12, 16, 23, and 25, the MLP model has provided predictions closest to the actual results. On days 2, 7, 15, 20, and 21, the CNN model has shown the closest predictions to the actual results. The graph generated based on the model predictions is presented in Figure 6.

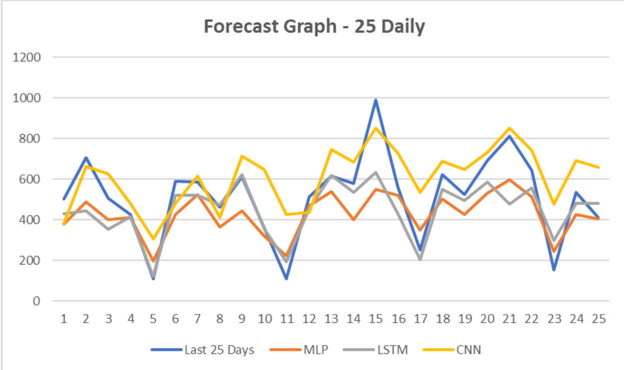


**Figure 6.** Twenty-five days forecast graph

According to Figure 6, it can be said that the graph of the LSTM model closely resembles the actual data graph. When examined overall, the ups and downs of the ceiling and floor values mostly occur on a day-to-day basis in a similar manner. Due to the fact that the amount of pallets to be transported per day does not follow a regular pattern, predicting it poses a challenging problem. For instance, when predicting the solar exposure of solar panels, similar values can be obtained when looking at years. However, in logistic data, the amount of pallets to be transported both yearly and daily does not exhibit variability based on a specific criterion. Therefore, predicting such data is considerably challenging. Nevertheless, upon examining the data and graphs, it can be stated that highly accurate prediction values have been achieved. Performance metric calculations conducted over the last 25 days on the entire dataset to evaluate the performance of machine learning methods are provided in Table 4.

**Table 4.** Performance evaluation results for the last 25 days using all data

| METHOD | MSE | RMSE | NRMSE | MAE | ESD | RC |
|---|---|---|---|---|---|---|
| LSTM | 6,410.5571 | 80.0660 | 0.2165 | 19.1275 | 0.8080 | 0.8990 |
| MLP | 20,536.5564 | 143.3058 | 0.3875 | 38.9471 | 0.1342 | 0.5683 |
| CNN | 8,492.4297 | 92.1544 | 0.2492 | 9.5310 | 0.7308 | 0.8619 |

According to the data in Table 4, the LSTM method has the lowest MSE (6,410.5571) and RMSE (80.0660) values, indicating that LSTM is capable of making more accurate predictions compared to other methods. The MLP method, on the other hand, has higher MSE (20,536.5564) and RMSE (143.3058) values, indicating a higher margin of error in predictions. The CNN method performs better than LSTM but is less successful than the MLP method. These results suggest that the LSTM method is a more effective choice for demand analysis in the logistics sector. The graphs illustrating the MAE and NRMSE values obtained from this data are provided in Figure 7.

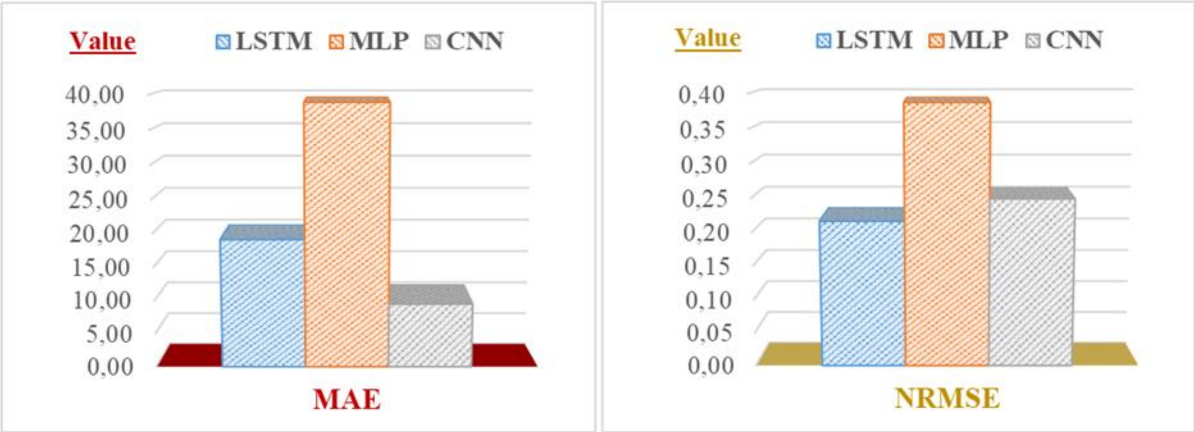


Figure 7. The result graphs of the MAE and NRMSE values for the last 25 days of data.

Table 5 presents the performance metrics calculated for the LSTM, MLP, and CNN methods regarding the training and testing data. Among the metrics, MSE, RMSE, NRMSE, MAE, ESD, and RC are included. Based on calculations on the training data, the LSTM method has the lowest MSE (2,867.3645) and RMSE (53.5478) values, while the MLP method is observed to have the highest MSE (17,372.6876) and RMSE (131.8055) values. The CNN method exhibits a moderate level of performance. In terms of calculations on the testing data, the LSTM method has the highest MSE (38,311.1055) and RMSE (195.7322) values, while the MLP method stands out as the method with the highest prediction errors. The CNN method, on the other hand, demonstrates better performance compared to the other two methods.

**Table 5.** Last 25 Days Performance Metric Calculations Training and Test

| Method | MSE | RMSE | NRMSE | MAE | ESD | RC |
|---|---|---|---|---|---|---|
| Performance Metric Calculations of Training Data - last 25 days | | | | | | |
| LSTM | 2,867.3645 | 53.5478 | 0.1494 | 14.6842 | 0.9119 | 0.9469 |
| MLP | 17,372.6876 | 131.8055 | 0.3678 | 32.0389 | -0.1443 | 0.5744 |
| CNN | 4,002.3105 | 63.2638 | 0.1765 | 6.7935 | 0.8644 | 0.9153 |
| Performance Metric Calculations of Test Data - last 25 days | | | | | | |
| LSTM | 38,311.1055 | 195.7322 | 0.4138 | 59.1318 | -0.2485 | 0.3713 |
| MLP | 49,021.9216 | 221.4089 | 0.4680 | 101.1437 | -0.0855 | 0.3452 |
| CNN | 48,918.4805 | 221.1752 | 0.4675 | 34.1773 | -0.1683 | 0.3191 |

Based on the data in Table 5, it can be inferred that there is a difference between the training and testing data, indicating the models' generalization capabilities. Methods that perform better on the training data may exhibit higher error values on the testing data.

For weekly predictions, data from the dataset was used once again. Within an approximately 10-year period, a total of 520 weekly data points is available. However, the quantities of pallets transported on a weekly basis vary significantly. Due to inconsistencies caused by official or religious holidays within these weeks, not all data points were utilized. In order to perform weekly predictions, the models were trained and tested. The 4-week prediction data is presented in Table 6.

**Table 6.** Prediction of pallet quantities for four weeks

| Number of Data | Last 4 Weeks | MLP | LSTM | CNN |
|---|---|---|---|---|
| 1 | 2722 | 3168.5 | 2954.5 | 3102.0 |
| 2 | 3510 | 2429.1 | 3018.5 | 3121.0 |
| 3 | 3443 | 3200.0 | 3195.5 | 3292.0 |
| 4 | 949 | 1874.0 | 1898.5 | 1929.0 |

When examining Table 6, it can be observed that the MLP model is the one that comes closest to the actual value in the 4th week. The LSTM model is the one that comes closest to the actual values in the first week. In the 2nd and 3rd weeks, the CNN model is the one that comes closest to the actual values. The graph depicting the predictions for the 4-week period based on the data in the table is presented in Figure 8.

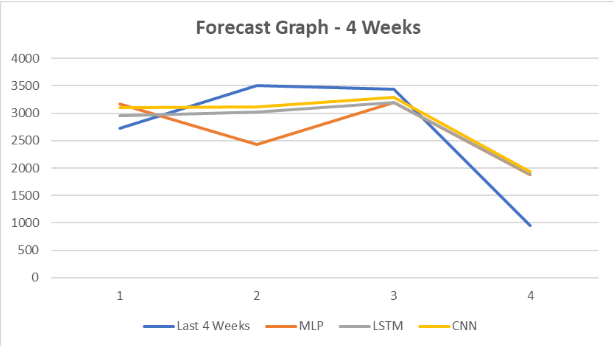


Figure 8. The prediction of palette quantities for the last 4 weeks

According to Figure 8, LSTM and CNN models, except for the MLP model, have generated similar graphs for the last 4 weeks of data. These graphs might exhibit even closer resemblance to the original data graph when more data points are available. When reviewing the literature, it becomes evident that predictions for monthly and yearly data require daily data spanning over at least 20 years. Upon examining the data predicted by the models in this project and the graphs drawn based on these predictions, it can be observed that the models are capable of making successful predictions. Logistic data doesn't typically exhibit consistent daily upward or downward trends. Due to these factors, predicting the future in such data sets is quite challenging. However, using approximations, daily, weekly, monthly, and yearly estimates for the number of palettes to be transported can be made. It's worth noting that during the Covid-19 pandemic, life came to a standstill, leading to a significant slowdown in the logistics sector as well. The data used in this study also includes information from the pandemic period.

The models used in the study have the potential to be improved by retraining them with continuously added data. With the last 4 weeks of data, performance metrics have been calculated for the LSTM, MLP, and CNN methods, and the results are provided in Table 7. The metrics include MSE, RMSE, NRMSE, MAE, ESD, and RC values.

**Table 7.** Performance Evaluation Results for the Last 4 Weeks of Data

| METHOD | MSE | RMSE | NRMSE | MAE | ESD | RC |
|---|---|---|---|---|---|---|
| LSTM | 75,683.7812 | 275.1068 | 0.1282 | -66.9752 | 0.7121 | 0.9092 |
| MLP | 243,934.4332 | 493.8972 | 0.2302 | 60.9471 | -0.7174 | 0.5386 |
| CNN | 82,456.3203 | 287.1521 | 0.1338 | -23.2919 | 0.7075 | 0.9049 |

When examining Table 7, it can be observed that the LSTM method has higher MSE (75,683.7812) and RMSE (275.1068) values compared to the other methods. The MLP method stands out as the one with the highest prediction errors with MSE (243,934.4332) and RMSE (493.8972) values. The CNN method, on the other hand, has MSE (82,456.3203) and RMSE (287.1521) values. The graphs representing the MAE and NRMSE values obtained from these data are provided in Figure 9.

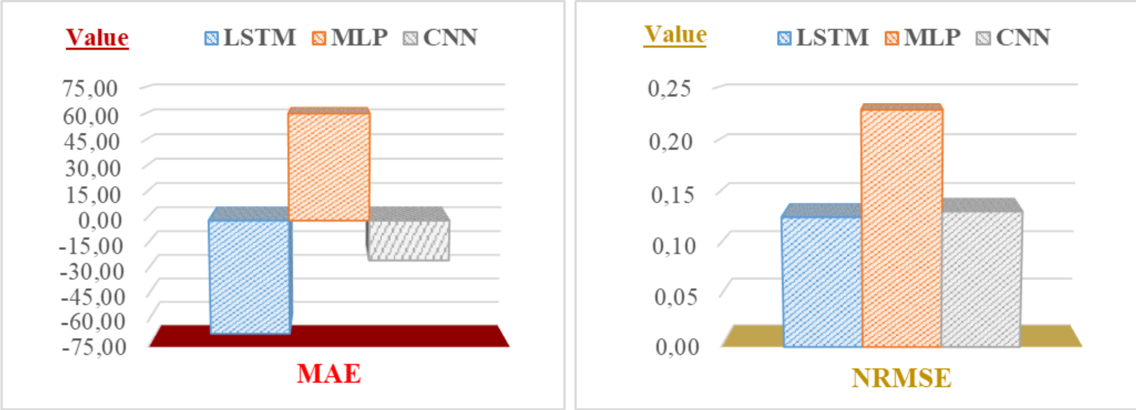


**Figure 7.** MAE and NRMSE Result Graphs for the Last 4 Weeks of Data

In Table 8, performance metrics related to LSTM, MLP, and CNN methods have been calculated for training and test data. The metrics include MSE, RMSE, NRMSE, MAE, ESD, and RC. According to the calculations on the training data, the LSTM method has the lowest MSE (21,158.7227) and RMSE (145.4604) values, while the MLP method is observed to have the highest MSE (191,272.0793) and RMSE (437.3455) values. The CNN method exhibits a performance close to the other two methods. On the test data, the LSTM method has the highest MSE (571,973.1875) and RMSE (756.2891) values, while the MLP method stands out as the method with the highest prediction errors. The CNN method, on the other hand, demonstrates better performance compared to the other two methods.

**Table 8.** Last 4 Weeks Performance Metric Calculations Training and Test

| Method | MSE | RMSE | NRMSE | MAE | ESD | RC |
|---|---|---|---|---|---|---|
| Performance Metric Calculations of Training Data - last 4 weeks | | | | | | |
| LSTM | 21,158.7227 | 145.4604 | 0.0699 | -67.3789 | 0.9219 | 0.9360 |
| MLP | 191,272.0793 | 437.3455 | 0.2102 | 57.6191 | -0.7110 | 0.5092 |
| CNN | 20,972.3359 | 144.8183 | 0.0696 | -22.9950 | 0.9178 | 0.9468 |
| Performance Metric Calculations of Test Data - last 4 weeks | | | | | | |
| LSTM | 571,973.1875 | 756.2891 | 0.2758 | -63.3008 | -6.2240 | -0.0295 |
| MLP | 723,278.4302 | 850.4578 | 0.3102 | 267.3233 | -2.4642 | 0.0535 |
| CNN | 642,086.0625 | 801.3027 | 0.2923 | -25.9943 | -3.4594 | 0.0032 |

The results of these experiments demonstrate the potential application of machine learning methods in various areas of the logistics sector, such as demand forecasting, route optimization, and inventory management. Additionally, the outcomes emphasize the need for careful evaluation during the training and testing phases of the models. Analyses conducted using different performance metrics can guide the process of selecting the right model and assessing prediction accuracy.

In conclusion, this study highlights that machine learning methods can be a valuable tool for future needs analysis in the logistics sector. However, it is important to consider the performance of these methods across different datasets and to conduct a thorough evaluation when selecting models.

## 5. Conclusion and Discussion

The conducted experimental analysis and evaluation of performance metrics have demonstrated that LSTM, MLP, and CNN models can be effectively utilized for predicting the substantial pallet transportation quantities in the logistics sector. The results have highlighted that the LSTM model has the lowest error values on training data. However, a decrease in its performance has been observed on test data. Despite having higher errors on training data, the MLP model exhibited better performance on test data. The CNN model, on the other hand, has generally shown a balanced performance. The findings of this study support that machine learning methods are effective tools for predicting pallet transportation quantities in the logistics sector. Decision-makers can utilize the information provided by these models to make more informed and strategic decisions. Nevertheless, factors like model selection and performance metrics should be considered. In future research, more comprehensive datasets should be employed to thoroughly examine the models' performance. Furthermore, comparing different machine learning algorithms and developing more advanced prediction models should also be explored. Such studies could contribute significantly to enhancing efficiency in the logistics sector and improving decision-making processes. As authors of this article, we aim to achieve more comprehensive results in future research by employing broader datasets and different machine learning algorithms.

## Compliance with ethics requirements

This study does not contain any studies with human participants or animals performed by any of the authors.

## Availability of Supporting Data

The data used to support the findings of this study are available from the corresponding author upon request.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Fundings

This publication is sponsored by ALISAN Logistics Inc.

## Acknowledgements

In this project, ALISAN Logistics Inc. is a stakeholder that contributed to the emergence of the working idea and creation of data. We would like to express our gratitude to ALISAN Logistics Inc., which operates in Turkey and provides expertise and knowledge in the logistics industry.

## References

[1] Gutelius, B. and N. Theodore, The future of warehouse work: Technological change in the US logistics industry. 2019.

[2] Gubán, M. and G. Kovács, Industry 4.0 Conceptions. Acta Technica Corviniensis-Bulletin of Engineering, 2017. 10(1).

[3] Karagulle, A.O., Green business for sustainable development and competitiveness: an overview of Turkish logistics industry. Procedia-Social and Behavioral Sciences, 2012. 41: p. 456-460.

[4] Stough, R.R., New technologies in logistics management, in Handbook of Logistics and Supply-Chain Management. 2017, Emerald Group Publishing Limited.

[5] Zhao, J. and F. Xie, Cognitive and artificial intelligence system for logistics industry. International Journal of Innovative Computing and Applications, 2020. 11(2-3): p. 84-88.

[6] Tsolaki, K., et al., Utilizing machine learning on freight transportation and logistics applications: A review. ICT Express, 2022.

[7] Han, Z., et al., A review of deep learning models for time series prediction. IEEE Sensors Journal, 2019. 21(6): p. 7833-7848.

[8] Dingli, A. and K.S. Fournier, Financial time series forecasting-a deep learning approach. International Journal of Machine Learning and Computing, 2017. 7(5): p. 118-122.

[9] Goyal, A., et al. A solution to forecast demand using long short-term memory recurrent neural networks for time series forecasting. in Proceedings of the Midwest Decision Sciences Institute Conference, Indianapolis, IN, USA. 2018.

[10] Lolli, F., et al., Machine learning for multi-criteria inventory classification applied to intermittent demand. Production Planning & Control, 2019. 30(1): p. 76-89.

[11] Van Gils, T., et al., The use of time series forecasting in zone order picking systems to predict order pickers' workload. International Journal of Production Research, 2017. 55(21): p. 6380-6393.

[12] Chan, H.K., S. Xu, and X. Qi, A comparison of time series methods for forecasting container throughput. International journal of logistics research and applications, 2019. 22(3): p. 294-303.

[13] Talupula, A., Demand forecasting of outbound logistics using machine learning. 2019.

[14] Sohrabpour, V., et al., Export sales forecasting using artificial intelligence. Technological Forecasting and Social Change, 2021. 163: p. 120480.

[15] Deng, C. and Y. Liu, A Deep Learning-Based Inventory Management and Demand Prediction Optimization Method for Anomaly Detection. Wireless Communications and Mobile Computing, 2021. 2021: p. 1-14.

[16] Ribeiro, A.M.N., et al., Short-and very short-term firm-level load forecasting for warehouses: a comparison of machine learning and deep learning models. Energies, 2022. 15(3): p. 750.

[17] Tang, Y.-M., et al., Forecasting economic recession through share price in the logistics industry with artificial intelligence (AI). Computation, 2020. 8(3): p. 70.

[18] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. 9(8): p. 1735-1780.

[19] Widiasari, I.R. and L.E. Nugroho. Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. in 2017 International Conference on Innovative and Creative Information Technology (ICITech). 2017. IEEE.

[20] Noriega, L., Multilayer perceptron tutorial. School of Computing. Staffordshire University, 2005. 4: p. 5.

[21] Bishop, C.M., Neural networks for pattern recognition. 1995: Oxford university press.

[22] Cinar, I. and M. Koklu, Classification of rice varieties using artificial intelligence methods. International Journal of Intelligent Systems and Applications in Engineering, 2019. 7(3): p. 188-194.

[23] Desai, M. and M. Shah, An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). Clinical eHealth, 2021. 4: p. 1-11.

[24] Ahmed, N.K., et al., An empirical comparison of machine learning models for time series forecasting. Econometric reviews, 2010. 29(5-6): p. 594-621.

[25] Koklu, M. and K. Sabanci, Estimation of credit card customers payment status by using kNN and MLP. International Journal of Intelligent Systems and Applications in Engineering, 2016. 4(Special Issue-1): p. 249-251.

[26] Chauhan, R., K.K. Ghanshala, and R. Joshi. Convolutional neural network (CNN) for image detection and recognition. in 2018 first international conference on secure cyber computing and communication (ICSCCC). 2018. IEEE.

[27] Singh, D., et al., Classification and analysis of pistachio species with pre-trained deep learning models. Electronics, 2022. 11(7): p. 981.

[28] Dogan, M., et al., Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. Computers and Electronics in Agriculture, 2023. 204: p. 107575.

[29] Kursun, R., et al. Flower Recognition System with Optimized Features for Deep Features. in 2022 11th Mediterranean Conference on Embedded Computing (MECO). 2022. IEEE.

[30] Taspinar, Y.S., et al., Computer vision classification of dry beans (Phaseolus vulgaris L.) based on deep transfer learning

techniques. European Food Research and Technology, 2022. 248(11): p. 2707-2725.
[31] Koklu, M., et al., A CNN-SVM study based on selected deep features for grapevine leaves classification. Measurement, 2022. 188: p. 110425.
[32] Alzubaidi, L., et al., Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 2021. 8: p. 1-74.
[33] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.
[34] Kishore, B., et al., Computer-aided multiclass classification of corn from corn images integrating deep feature extraction. Computational Intelligence and Neuroscience, 2022. 2022.
[35] Koklu, M., I. Cinar, and Y.S. Taspinar, Classification of rice varieties with deep learning methods. Computers and electronics in agriculture, 2021. 187: p. 106285.
[36] Butuner, R., et al., Classification of deep image features of lentil varieties with machine learning techniques. European Food Research and Technology, 2023. 249(5): p. 1303-1316.
[37] Cinar, I. and M. Koklu, Identification of rice varieties using machine learning algorithms. Journal of Agricultural Sciences, 2022: p. 9-9.
[38] Yasin, E.T., R. Kursun, and M. Koklu, Deep Learning-Based Classification of Black Gram Plant Leaf Diseases: A Comparative Study.
[39] Mascarenhas, S. and M. Agarwal. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. in 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). 2021. IEEE.
[40] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[41] Tian, X. and C. Chen. Modulation pattern recognition based on Resnet50 neural network. in 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP). 2019. IEEE.
[42] Gencturk, B., et al., Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. European Food Research and Technology, 2023: p. 1-14.
[43] Qi, J., et al., On mean absolute error for deep neural network based vector-to-vector regression. IEEE Signal Processing Letters, 2020. 27: p. 1485-1489.
[44] Botchkarev, A., Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. arXiv preprint arXiv:1809.03006, 2018.
[45] Chicco, D., M.J. Warrens, and G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 2021. 7: p. e623.
[46] Wang, W. and Y. Lu. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. in IOP conference series: materials science and engineering. 2018. IOP Publishing.
[47] Chai, T. and R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE). Geoscientific model development discussions, 2014. 7(1): p. 1525-1534.
[48] Bishop, C.M. and N.M. Nasrabadi, Pattern recognition and machine learning. Vol. 4. 2006: Springer.
[49] Géron, A., Hands-on machine learning with scikit-learn and tensorflow: Concepts. Tools, and Techniques to build intelligent systems, 2017.
[50] Willmott, C.J. and K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 2005. 30(1): p. 79-82.
[51] Frías-Paredes, L., et al., Dynamic mean absolute error as new measure for assessing forecasting errors. Energy conversion and management, 2018. 162: p. 176-188.
[52] Lee, D.K., J. In, and S. Lee, Standard deviation and standard error of the mean. Korean journal of anesthesiology, 2015. 68(3): p. 220-223.
[53] Gao, J., Machine learning applications for data center optimization. 2014.
[54] Mousavi, S.M. and G.C. Beroza, A machine-learning approach for earthquake magnitude estimation. Geophysical Research Letters, 2020. 47(1): p. e2019GL085976.
[55] Kendall, M.G., Rank correlation methods. 1948.
[56] Zar, J.H., Spearman rank correlation. Encyclopedia of biostatistics, 2005. 7.
[57] Kendall, M.G., A new measure of rank correlation. Biometrika, 1938. 30(1/2): p. 81-93.
[58] Shi, H., M. Drton, and F. Han, On the power of Chatterjee's rank correlation. Biometrika, 2022. 109(2): p. 317-333.