JJMIE

Volume 19, Number 2, June. 2025
ISSN 1995-6665
Pages 415– 426
https://doi.org/10.59038/jjmie/190213
Editor-in-Chief: Prof. Ali M. Jawarneh

Jordan Journal of Mechanical and Industrial Engineering

Optimizing Sales Forecasting in Dairy Industry Using Neural Networks

Samer Abaddi

IDentity Research, Amman, Jordan

Received 8 Sep 2024 *Accepted 12 Mar 2025*

Abstract

This study attempts to design a more accurate and reliable sales forecasting model for the dairy business, specifically to mitigate challenges in Al-Youm Food Company's case. The nature of the dairy business, volatile demand, and the perishable nature of their products demand accurate forecasting to help restrict waste and maximize profits. The company is employing classic forecasting methods, which failed to capture their sales patterns' complexity. In response to this, a neural network-based forecasting model was developed and trained through Python machine learning packages using 43 weeks of sales history (January - October 2023). The neural network was compared to classic forecasting methods, i.e., exponential smoothing and moving averages. The neural network model was found to be more accurate in prediction, having a Root Mean Squared Error of 19.68 compared to 20.52 for exponential smoothing and 21.98 for moving average. Similarly, the Mean Absolute Percentage Error for the neural network was found to be 27.76%, better than that of exponential smoothing (29.33%) and moving average (30.06%). The stability of the model was also verified using training loss and validation loss analysis. An extended forecast was developed, forecasting sales between week 44 to week 80. With limitations such as short periods of historical data and lack of extraneous variables, it is found that neural networks can be employed to improve sales forecasting in the dairy business significantly. The application of this model in Al-Youm Food Company is likely to lead to better control of inventory, minimized waste, and higher profits, and even be used as a model for other companies in the business to be more efficient and more sustainable in their business in the dairy business.

Keywords: Neural network models, Sales forecasting, Dairy industry, Exponential smoothing, Moving average.

List of Abbreviations

ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
CBR	Case-Based Reasoning
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ES	Exponential Smoothing
KNN	K – Nearest Neighbor
LSTM	Long Short-Term Memory
MA	Moving Average
MLP	Multi-Layer Perceptron
RNN	Recurrent Neural Network
SVM	Support Vector Machines
YFC	Al Youm Food Company

1. Introduction

Any company plan must include sales forecasting as it enables organizations to make well-informed decisions on resource allocation, revenue estimates, and inventory management [1]. However, predicting sales accurately can be challenging due to the complexity and variability of influencing factors [2].In recent years, sales forecasting has been a perfect application area for neural networks since it can analyze past sales data along with several variables such as seasonality, promotions, economic indicators, and consumer behavior. Neural Nets come up with very accurate forecasts of future sales volumes using data to search for trends and relationships [3].

The use of neural networks in forecasting sales is a relatively new field gaining popularity due to the availability of advanced machine learning techniques and massive datasets. Neural Networks have been particularly effective at being responsive to recent trends, which bodes well in many use cases, including those about revenue forecasting models [4]. Several types of Neural Networks can be applied to sales forecasting, each with its strengths and weaknesses:

- Multi-Layer Perceptron (MLP): The most common and simple type of Neural Network [5].
- Recurrent Neural Network (RNN): Because they consider data from past and present observations, they are especially well-suited for time-series forecasting [6].

* Corresponding author e-mail: eng.sameralabbadi@gmail.com.

- Convolutional Neural Network (CNN): While CNNs were primarily used for computer vision, they are now being used for time-series forecasting [7].

Neural Networks offer several advantages over traditional forecasting methods. They are capable of adapting to changing sales variables and can consider a vast dataset of variables. They reveal patterns and relationships that traditional methods often overlook, resulting in more accurate sales prediction models [8]. This study will explore the use of neural networks in forecasting sales within an industrial engineering context: their training, the factors these networks must consider, and how they can be optimized to provide the most precise predictions.

Al Youm Food Company (YFC) relies on conventional forecasting methods, such as moving averages, and simple exponential smoothing. The methods do not capture YFC's sales patterns, which are highly dependent and often non-linear in nature. From interviews of YFC staff, this deficiency results in high waste ratios for a number of its products, frequent stock outs, and poor planning. The perishable nature of dairy foods, having short shelf lives and strict temperature requirements, highly aggravates the forecasting problem. In other words, YFC in Jordan encounters significant challenges in accurately forecasting its sales due to the intricate and fluctuating factors that influence consumer behavior and market trends. Conventional forecasting methods often fall short in capturing the dynamic nature of sales data, resulting in inefficient resource allocation, inadequate inventory management, and imprecise revenue projections. This gap is therefore filled by the present study, which will explore neural networks in sales forecasting at YFC. With the inclusion of advanced machine learning, this study will review how to develop a more precise and responsive sales forecast that will adapt to changes within the food industry market. Thus, based on this background, the following are the objectives this study aims to achieve:

- To evaluate the current sales forecasting methods used by YFC and identify their limitations
- To develop a sales forecasting model based on neural networks that is tailored to the specific requirements of YFC.
- Additionally, to conduct a comparative analysis of the effectiveness of the neural network model as opposed to traditional forecasting methods in terms of accuracy and responsiveness.

The research findings are laid out in five all-inclusive chapters. Section 1 commences by introducing the research, backed by the background of the study, the problem statement, and the objectives. Section 2 presents a detailed literature review, exploring existing sales forecasting methods, the role of neural networks in forecasting, and relevant case studies. Section 3 describes the methodology, including the research design, data collection processes, and the specific neural network models and techniques employed. Section 4 focuses on the analysis and discussion of results, comparing the performance of the neural network model against traditional forecasting methods. Finally, Section 5 concludes the study, summarizing the findings, and providing recommendations for implementing the neural network model.

2. Literature review

2.1. Forecasting types

The most famous types of forecasting are short-, medium- and long-term forecasting. Each type differs according to the time, the methodology used, and the amount of detail as shown in Table 1. The short forecast covers up to one year (12 months), while the medium forecast covers from one year to ten years, and the long-term forecast usually provides forecasts for more than ten years. The details of each methodology differ as short-term predictions are usually more accurate compared to medium- and long-term predictions. The medium forecasts use different patterns focusing on the economic and demographic impact, while the long-term patterns consider the economic and climatic dimensions over the decades [9].

Table 1. Comparison between short, medium, and long-term forecasts [10]

Category	Short-term	Medium-term	Long-term
Time horizon	Up to 12 months	1-10 years	More than 10 years
Level of detail	High details	Medium	Low
Factors	Short-term trends	Economic Demographic Technological	Global trends
Forecast methods	Time series Regression	Scenario analysis Time series Regression	Global system models AI and neural networks Scenario analysis
Uncertainty	Low degree	Medium degree	High uncertainty

2.2. Time-series forecasting methods

The methodologies of forecasting are widely used in various sciences, such as economics, engineering, and finance. Time series is a pivot method in forecasting. Time series is defined as a regular sequence of recorded values within specific periods. The methodology is divided into two main parts: the first part seeks to understand the pattern of the data, while the second section aims to develop a future prediction with the help of the best-fit curves [11]. There are two types of time series analyses: the univariate, which usually includes a single pattern of data recorded over time, such as the hourly energy consumption, and the second multivariate, which includes a group of variables [12].

According to [13] nine popular time-series forecasting techniques were used in the field of energy consumption which are Artificial Neural Network (ANN), Autoregressive Integrated Moving Average (ARIMA), Support Vector Machines (SVM), Case-Based Reasoning (CBR), Fuzzy time series, grey prediction model, Moving Average and Exponential Smoothing (MA & ES), K –

Nearest Neighbor prediction method (kNN) and the hybrid models. Each of the techniques above gives appropriate results according to the nature of the data and the period covered.

Each of the previous time-series techniques has its pros and cons that drive researchers to choose it. One of the most prominent advantages of the ANN technique is its ability to draw the input and output relationships accurately, and it can be used in non-linear time series analysis. However, the ANN depends on the weight values and has a problem with local minimum limits and the inability to generalize the results [14]. Regarding the ARIMA technique, its advantages include relying on the shift of historical data, the ability to improve the efficiency of the model with the help of regression, and the existence of confidence intervals, on the other hand, generalizing the results of the model is difficult, and the ARIMA is not suitable for long-term expectations. It does not support non-linear data [15]. The SVM is a popular forecasting technique that provides a general picture and is suitable for long periods. However, the results are not transparent in this technique [16]. The CBR is close to human mental simulation and does not need to find rules between the problem parameters. However, this technique requires huge amounts of data and the definition of new aspects. The fuzzy algorithm is like the CBR technique in terms of closeness to human experience. Moreover, the first can help solve uncertainty problems, but it is complex and has low stability levels compared to other techniques [17].

ARIMA forecasting was applied to various settings. In Jordan, [18] recently utilized the approach to forecasting energy consumption in Jordan up to 2030. The approach had varying MAPE results between 6-22%. However, [19] utilized the approach to forecasting social media engagement rates based on 2-year data from Facebook and Instagram. The approach was effective in predicting the behavior of non-linear data with multiple variables

The grey forecasting approach was developed in the 1980s and is based on the premise that a small sample of data can predict future behaviour. One of the main advantages of this model is its ability to predict effectively when there is a limited amount of data and to handle it with ease. Additionally, this approach is suitable for both short-term and long-term analysis. On the downside, the gray technique assumes that the data is linear and may not yield good results in the presence of seasonal fluctuations and trends. It was used to predict the amount of CO2 emissions in China [20]. In Jordan, [21] utilized system dynamics to predict CO2 emissions reduced post-shifting to electrification in heavy-duty vehicles. The study highlighted the importance of incorporating more factors for better long-term forecasting, such as ICT infrastructure, technology maturity and social factors.

The MA approach is one of the oldest forecasting techniques that rely on calculating the average of a set of data to obtain future forecasts. One of the advantages of the MA method is its ease of computing and understanding; furthermore, it can be applied in various sciences and used in short- and long-term prediction. On the other hand, the MA method is sensitive according to the data. It also assumes that the behavior of the past is similar to the future, and it ignores the impact of other variables in its predictions. ES is another popular forecasting technique that depends on weighted averages of previous data. The ES is used for short and long forecasts, and it allows the scheduling and modification of parameters easily for quarterly and seasonal data. The disadvantage of the technique includes the ignorance of other factors, and it also requires the selection of smoothing parameters [22].

2.3. Examples of previous work

Numerous pieces of literature have been presented to predict production demand using various models, including time series.

A study titled "Exploring the Use of Deep Neural Networks for Sales Forecasting in Fashion Retail", although focused on fashion retail, provides valuable insights applicable to dairy production forecasting. They applied a Deep Neural Network model to predict sales volumes using past sales data and other features, such as sales promotions and economic indicators. The approach was to train a multilayer DNN, letting it model complex patterns in the data. The results showed that the method, based on the DNN, exceeds the classical approaches, such as ARIMA and exponential smoothing, in accuracy and adaptability to changes in the market. The study exemplifies the practical use of the neural network to improve forecasting precision in dynamic and variable environments, such as dairy production [3].

Another study focused on the dairy industry and was titled "Application of Neural Networks to Explore Manufacturing Sales Prediction." The authors have adapted the RNN model due to its efficiency when the forecast is made over a time series and can take into account the past and present information simultaneously. This study follows the approach by collecting past production and sales data from the dairy firm, preparing the data with respect to the seasonality and trend effect in the data, and training the RNN model. The results obtained from this study show that the RNN model improved forecasting accuracy significantly compared with the classical methods in cases dealing with common seasonal fluctuations in the dairy production process. According to the obtained results, the study concludes about the appropriateness of using neural networks, and in this case, especially RNNs, as applicable tools for predicting industries with complex and variable production processes [4].

Another research paper titled, "Predicting the Number of Customer Transactions in the Dairy Sector Using Stacked LSTM Recurrent Neural Networks" [2], extends to the dairy sector. The selected method is through a Stacked LSTM network in the RNN framework. This method solves the vanishing gradient problem and can learn long-term dependencies. Previous background information relates to historical sales data, promotional activities, and external economic factors. The results showed how a stacked LSTM network could provide more precise and stable demand forecasts than the traditional models, efficiently capturing the seasonal and promotional effects that are typical in the dairy sector. This study emphasizes the advantages of advanced neural network architectures in improving forecasting reliability and precision within dairy production.

A comparative analysis titled "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions" included applications in the dairy production sector. They compared forecasting effectiveness among various methods: ARIMA, exponential smoothing, and some neural nets like RNNs and CNNs. The developed models were used to check forecast accuracy and computational efficiency. This is from the past, meaning the production and sales data from different dairy companies when checking on the models. Their work showed the comparison of RNN and CNN models with traditional methods, especially in capturing non-linear patterns and adapting to changes in demand that come at an unexpected instant. Another work further brought evidence supporting the importance of neural networks in enhancing demand forecasting in the dairy sector [8].

Neural networks have been used in Jordan for forecasting practices for the major fields of finance, other than the peripheral ones of agriculture and production industries. This subsection presents the pioneering studies and research conducted in this country using neural networks for forecasting purposes. A study by [23] aimed to explore the application of neural networks in forecasting the financial performance of Jordanian banks. The essence was to develop a model that could allow the application of credits in loan applications. The neural network provided the approach used by the researchers. The model was trained using historical financial data of various banks in Jordan. The results showed that the applied neural network model brought significant improvement in the traditional statistical method with more accurate and timely predictions.

Recent literature indicates a boost in energy demand across the globe, predominantly in the industrial sector, due to improvements in technologies. The study [24] was focused on Jordan's industrial sector, applying artificial neural networks (ANNs) to simulate, analyze, and forecast electricity demand. The work analyzed the effects of different key variables, such as the number of industrial establishments, number of employees, price of fuel and electricity, gross output, structural effects, and capacity utilization on energy demand. Multiple models of ANN were created and compared using different metrics such as RMSE, MAPE, and R-squared. The work found that number of establishments and number of employees, followed by gross output, are the primary drivers of electricity demand in Jordan's industrial sector. The proposed ANN model was found to be better in prediction compared to previous results in the literature.

Another study utilized ANNs to predict the porosity and hardness of novel Al-glass composite materials produced via powder metallurgy. Results showed that sintering time and percentage of glass influenced these properties most. An outstanding predictive accuracy was found in a Levenberg-Marquardt-trained ANN, with three input variables and one output (up to R-squared of 99.99%, up to RMSE of 0.007141). The work indicates that ANNs can capture sophisticated relationships in composite material properties to a great extent, offering a significant time and cost-benefit compared to experimental approaches [25].

The study of [26] aimed to project wheat production in Irbid region of Jordan by constructing forecasting models. Given the high variability of domestic wheat production, the objective was to establish a system for providing early information about expected production levels. The researchers have formulated linear and log-linear prediction functions using accumulated rain in the growing season, including special consideration of rain in November, December, and January. It has been found that early monthly rainfall and the cultivated areas have influenced wheat production. At the same time, temperature and the number of rainy days have a slight influence on wheat production. The findings showed the role being played by rainfall; beyond a single millimeter of rain in December and January, wheat production increased, and the excess amount of wheat production contributed by 120 and 111 tons, respectively. In addition, the forecast showed that stable wheat production projections might be obtained by the end of January, with numerous months for economic decisions.

Newer studies highlighted the role of other technologies similar to neural networks, such as artificial intelligence (AI), to increase efficiency, produce added value, and improve work safety in factories. In the study of [27], machine learning and image processing techniques were investigated using four architectures of Convolutional Neural Network (CNN) (such as ResNet-152 and Inception) to classify marble slabs. Similarly, [28] noted that to remain competitive in the face of rapid technological advancements, manufacturing companies must adopt cutting-edge technologies. The study specifically investigated the impact of integrating supply chain partners in the implementation of Industry 4.0 (I4.0) technologies on overall competitiveness, showing a significant relationship.

The literature review highlights some applications of neural networks and advanced optimization techniques in forecasting in various sectors. This study illustrates the potential for neural networks to deal with variability and uncertainty in sales forecasting and optimize resource allocation and strategic planning in dairy production.

3. Methodology

3.1. Roadmap

The study approach involves a structured and methodical process designed to address the outlined objectives. Figure 1 illustrates the study's step-by-step approach, detailing the process from defining the current problem to suggesting future improvements to sales forecasts. For neural networks, Python was used. However, the exponential smoothing and the moving average forecasting were performed using Excel. The combination of both software facilitates the delivery and clarification of the outcome.

3.2. Data collection and validation

The primary data source includes UFC's sales figures, gathered from various stores in Jordan over 44 weeks (between January and October 2023). This longitudinal data collection approach allows for a comprehensive analysis of sales patterns and the factors influencing them.

The study begins with the collection of sales data, ensuring that it encompasses a diverse range of store locations and product categories. This data is compiled into a central database, enabling detailed statistical analysis and modelling. Table 2 provides an overview of the key variables to be collected and their descriptions.

Data preprocessing steps are implemented to guarantee the integrity and consistency of the dataset with regard to handling missing values, normalizing sales data, and encoding categorical variables, considering data collection added to the research design. We will go on and perform exploratory data analysis to bring out the underlying pattern and trend of the sales data. Descriptive statistics and visualization will be done in order to derive insight into the characteristics of the data, showing the various relationships that may exist among these variables. We will then go on to develop a neural network-based model for sales forecasting after we have completed the exploratory analysis. This model will be specifically designed to capture the complex, non-linear relationships present in the sales data. The architecture of the neural network will be customized to meet the specific needs of YFC, integrating layers and activation functions tailored for time-series forecasting.

Traditional forecasting methods, like exponential smoothing with alpha factors of 0.2, 0.35, and 0.5, will also be developed to compare the performance of the neural network model. Each model will be evaluated in terms of accuracy and responsiveness using performance metrics such as RMSE, MAPE, and sMAPE." Performance metrics are summarized for all models within the following table, Table 3.

Table 2. Key variables and descriptions

Variable	Description
Store ID	Unique identifier for each store
Product ID	Unique identifier for each Al Youm product
Week	The week number, ranging from 1 to 44
Gross Sales	The total/ cumulative sales in each store per week
Returns	The number of returned products due to expiry or other conditions
Net Sales	The net sales (without the returns) for each product in each store per week

Table 3. Performance metrics for model evaluation

Metric	**Description**
RMSE	The square root of the average of squared errors
MAPE	Average of absolute percentage errors between predicted and actual values
sMAPE	Average of absolute percentage errors between predicted and actual values, normalized by the sum of the actual and predicted values

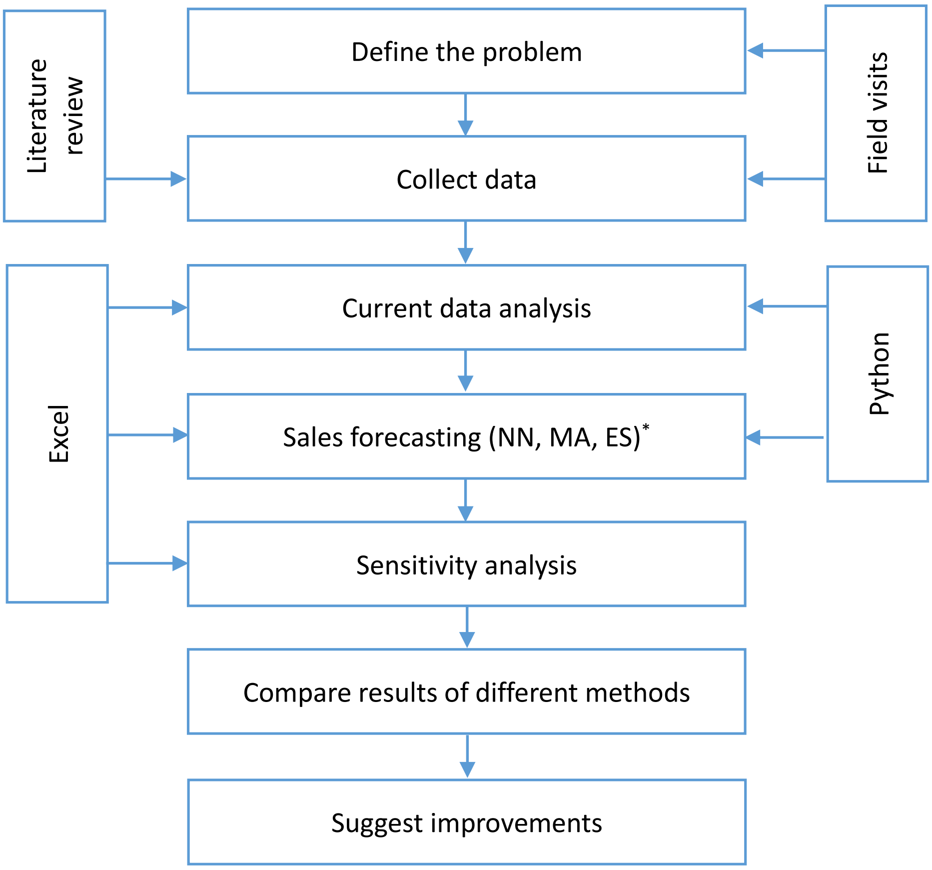

Figure 1. The roadmap of the study

* NN: Neural network, MA: Moving average, ES: Exponential smoothing

3.3. Neural network architecture

The neural network model designed for this study is a multi-layered feedforward network. The architecture comprises the following components [29]:

1. Input Layer: The input layer comprises neurons that correspond to the dataset features, such as net sales, promotional indicators, seasonal indices, and relevant economic indicators.
2. Hidden Layers: Several hidden layers capture nonlinear relationships between data points. Each of the hidden layers has an activation function like ReLU (Rectified Linear Unit) to give it non-linearity. The number of hidden layers and the number of neurons in each layer will be determined through hyper parameter tuning.
3. Output Layer: The output layer consists of a single neuron representing the predicted sales value for a given week and store-product combination.

The activation function for the hidden layers is the ReLU function (Equation 1), defined as:

$$f(x) = \max(0, x) \quad (1)$$

The Rectified Linear Unit (ReLU) outputs zero for any negative input, and outputs the input itself for any positive input [30]. This straightforward computation technique makes ReLU a popular choice because it helps address the vanishing gradient problem that can occur with other activation functions such as the sigmoid or tanh. By allowing the network to approximate any arbitrary function, ReLU enables the learning of complex patterns in sales data. Forward propagation refers to the process in which the input data passes through the layers of the network to eventually output a prediction. At each of the layers, the output from any neuron goes through an intermediate calculation by the summation of its weighted inputs and an activation function applied to the result [31].

During forward propagation, the weighted sum of inputs for each neuron is calculated, and the activation function is applied (as shown in Equations (2 and (3). For a neuron j in layer l, the output is given by:

$$z_j^{(1)} = \sum_{i=1}^{nl-1} w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)} \quad (2)$$

$$a_i^{(l)} = f(z_i^{(l)}) \quad (3)$$

Where:

$w_{ij}^{(l)}$ is the weight connecting neuron $i$in layer $l-1$ to neuron j in layer l

$a_i^{(l-1)}$is the activation of neuron $i$in layer $l-1$

$b_j^{(l)}$is the bias term for neuron $j$in layer l

f is the activation function (ReLU in this case)

The loss function serves to quantify the disparity between predicted and actual values, thus guiding the training process by indicating the model's performance [32]. This enables the adjustment of the neural network's weights and biases to reduce errors and enhance prediction accuracy. Backpropagation is a crucial algorithm in training neural networks. It minimizes the loss function by iteratively adjusting the weights and biases. It involves computing the gradient of the loss function with respect to each weight and bias through the chain rule of calculus [33]. This gradient indicates the direction and magnitude of the adjustments needed, as shown in Equations (4) and (5).

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta \frac{\partial \text{MSE}}{\partial w_{ij}^{(l)}} \quad (4)$$

$$b_j^{(l)} = b_j^{(l)} - \eta \frac{\partial \text{MSE}}{\partial b_j^{(l)}} \quad (5)$$

Where:

$\eta$is the learning rate

$\frac{\partial \text{MSE}}{\partial w_{ij}^{(l)}}$is the partial derivative of the loss function with respect to the weight

$\frac{\partial \text{MSE}}{\partial b_j^{(l)}}$ is the partial derivative of the loss function with respect to the bias

4. Analysis and discussion of results

4.1. Conducting initial neural network forecasting

Utilizing Python, the forecasting model was meticulously developed and tested. Initially, the forecasting was confined to 13 weeks, from week 31 to week 43, serving as a controlled test phase. This timeframe was selected based on its recent occurrence, ensuring that the model could be validated against known outcomes to assess its predictive accuracy rigorously. The forecast for the last known 13 weeks is shown in Figure 2.

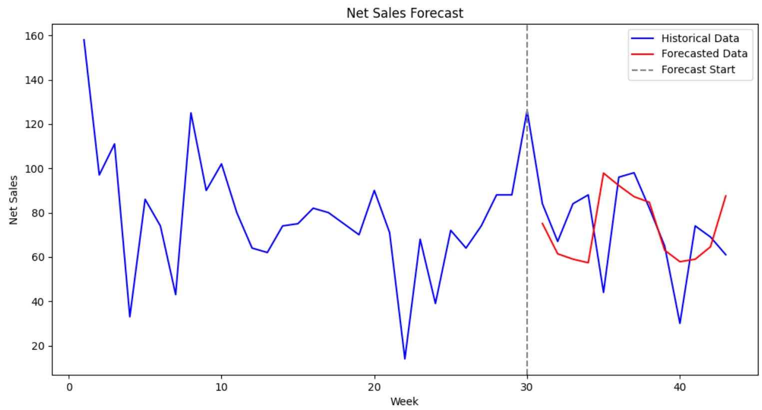

Figure 2. The initial forecast conducted for training the model and testing its accuracy

The historical data, depicted in blue, shows significant fluctuations, typical of dairy sales data, due to varying factors such as demand changes and supply issues. The forecasted data, shown in red, begins at week 31, indicated by the dashed line, and closely follows the trends and movements of the historical sales data. The initial analysis of the alignment suggests that the neural network has effectively captured the underlying patterns in the sales data. To verify this, a sensitivity analysis was conducted before digging into the long-term forecast.

The Sum of Squared Errors (SSE), totaling 5032.9558, and the average squared error per week, calculated as 387.1504, highlight the aggregate and average discrepancies between the forecasted and actual sales figures, respectively. The RMSE of 19.6761 further quantifies the average magnitude of the forecasting errors, as shown in Table 4, providing a more interpretable measure of forecast accuracy by adjusting for the scale of the data.

The ANN model's MPE of 11.9179, indicating a mean underestimation of sales, is consistent with potential limitations of ANNs, discussed in [14], in that weight values used in the model and potential local optima can lead to skewed estimates. Underestimation is a serious supply chain problem that can lead to shortages. The MAPE of 27.7591% shows a high variance in model accuracy, a finding contrary to results in certain studies of better-than-standard ANN performance compared to traditional methods ([24], [25]) though in different applications (electricity demand forecasting and material properties, respectively). The sMAPE of 11.74% gives a fair reflection of model performance. As high precision of ANNs has been found in certain studies (for instance, up to R-squared of 99.99% in forecasting material properties [25]), results here point to the challenge in accurately forecasting sales in a dynamic system, in keeping with complications discussed in studies of using neural networks in dairy and retail sales ([3], [2], [4], [8]). The variance in precision of forecast here may, as in using time series in social media [19], display complexity using a single model.

4.2. Comparison with forecasting methods

Following the neural network forecasting, exponential smoothing forecasting was performed utilizing a smoothing constant α = 0.2 (The most optimum with comparison to α = 0.35, 0.5, and 0.7). This value represents the weight given to the most recent observation in the exponential smoothing model, balancing between the historical data and the new observations. By applying this method, the forecasting model aims to smooth out short-term fluctuations and highlight longer-term trends or cycles in the dairy sales data. The results of the sensitivity analysis for the exponential smoothing forecasting with α =0.2 are presented in Table 4.

The neural network model is more accurate compared to exponential smoothing in most of the error metrics, in accordance with findings in a number of studies that affirm the use of neural networks in forecasting ([3], [4], [8], [25]). The neural network's smaller Sum of Squared Errors (SSE) and Average SSE per Week point to a more accurate and credible forecast over the observed period, in accordance with neural networks' ability to learn sophisticated patterns, such as in studies applying ANNs in sales forecasting and demand forecasting ([3], [4], [8]). Especially the neural network's smaller RMSE of 19.676 compared to that of exponential smoothing's 20.520 also bears witness to its higher predictive power, making its forecast more accurate to actual sales data. This conforms with high accuracy findings in studies applying ANNs in forecasting material properties [25] and electricity demand [24], though in different application contexts.

The neural network also exhibits a substantially lower MPE, suggesting less bias in the forecast. However, the MAPE is slightly lower in the neural network model, indicating that, on average, its predictions are closer to actual figures compared to those of the exponential smoothing model.

Following the application of neural network and exponential smoothing forecasting methods, a moving average forecasting method was performed for a 5-period interval to further evaluate the sales trends of Al-Youm Food Company. The moving average method is a simple yet effective technique that smooths out short-term fluctuations and highlights longer-term trends by averaging sales data over a specified number of periods. This method was employed to provide a comparative baseline against the more sophisticated forecasting techniques previously used. Table 4shows the accuracy results of this model.

Compared to the neural network and exponential smoothing methods, the moving average method yields higher SSE and average SSE per week, indicating larger discrepancies between the forecasted and actual sales figures. The RMSE of 21.9847 is higher than the revised neural network RMSE but lower than the initial neural network and exponential smoothing RMSE values, suggesting moderate forecasting accuracy.

The moving average approach reveals an MPE of 22.9559, a higher bias towards underestimation compared to that of neural networks and exponential smoothing models. The underestimation inclination is likely a reflection of the moving average's natural deficiency in handling sudden changes or complex patterns in the data, compared to neural networks in studies [3], [4], and [8]. The moving average's MAPE of 30.0628 is higher compared to that of a tuned neural setup yet lower compared to that of a less-tuned neural setup, a reflection of a relatively better absolute percentage error compared to a less-tuned setup, yet short of a highly tuned setup as indicated in studies [25] and [24]. The symmetric Mean Absolute Percentage Error (sMAPE) of 0.1089, slightly higher compared to that of exponential smoothing, is a fairly balanced performance yet short of that of the neural network, affirming the overall superiority of the neural network approach in this specific forecasting application. The moving average approaches are better compared to that of time-series application in the energy sector in line with [18] with MAPE results between 6-22%.

4.3. Long-term forecasting using neural networks

Building on the superior accuracy and consistency demonstrated by the neural network model in the comparative analysis, long-term forecasting was conducted to project sales trends from week 44 to week 80. This decision was predicated on the neural network's ability to capture complex patterns and fluctuations in the sales data, as evidenced by its lower error metrics compared to both exponential smoothing and moving average methods.

After conducting thorough training of the neural network model with historical sales data and validating its performance against the known 13-week period (Weeks 31 to 43), the model was deployed for extended forecasting. The aim of the long-term forecasting (Figure 3) was to provide Al-Youm Food Company with a reliable projection of future sales, facilitating more strategic decision-making and improved inventory management.

The parameters of the neural network model were precisely adjusted during the process of training to be highly accurate and resilient, a process in keeping with achieving high performance as in other applications of ANNs [25, 24]. The model's performance was compared to that of the conventional methods (exponential smoothing and moving average) using tried metrics: Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and symmetric Mean Absolute Percentage Error (sMAPE). 1 In all of these metrics, better performance was shown by the neural network model, beating the usual methods in all instances. The consistency of this better performance across different error metrics, just as in comparative studies in [3], [4], and [8], was a compelling reason to use the neural network model in long-term sales forecasting, owing to its capability to capture underlying patterns better.

The blue line illustrates the historical sales data, demonstrating notable fluctuations and volatility characteristic of the dairy industry's sales patterns. The red line depicts the forecasted sales data for the validation period (Weeks 31 to 43), while the green line represents the extended forecast from week 44 to week 80. The forecast indicates sustained fluctuations, aligning with expected historical sales patterns. The extended forecast also shows a cyclic pattern, with peaks and troughs occurring on a regular basis, suggesting that sales are seasonal. This would be expected in the dairy industry, for example, where the seasonality could depend on holidays, school terms, or other reasons. By anticipating future sales trends, the company can optimize its operations to meet expected demand, reducing the risk of overproduction or stockouts.

Table 4. The results of the sensitivity measures for the three forecasting methods (Neural networks, exponential smoothing, and moving average)

Metric	Symbol	Neural Network	Exponential Smoothing	Moving Average
Sum of Squared Errors	SSE	**5032.955754**	5473.695679	6283.28
Avg. SSE per Week	$\overline{SSE}$	387.1504426	421.0535138	483.3292308
Root Mean Square Error	RMSE	**19.67613891**	20.51958854	21.98474996
Mean Percentage Error	MPE	**11.91785361**	20.18251651	22.9559206
Mean Absolute Percentage Error	MAPE	27.75905891	29.32795618	30.06280181
Symmetric Mean Absolute Percentage Error	sMAPE	0.117425645	0.107285835	0.108881446

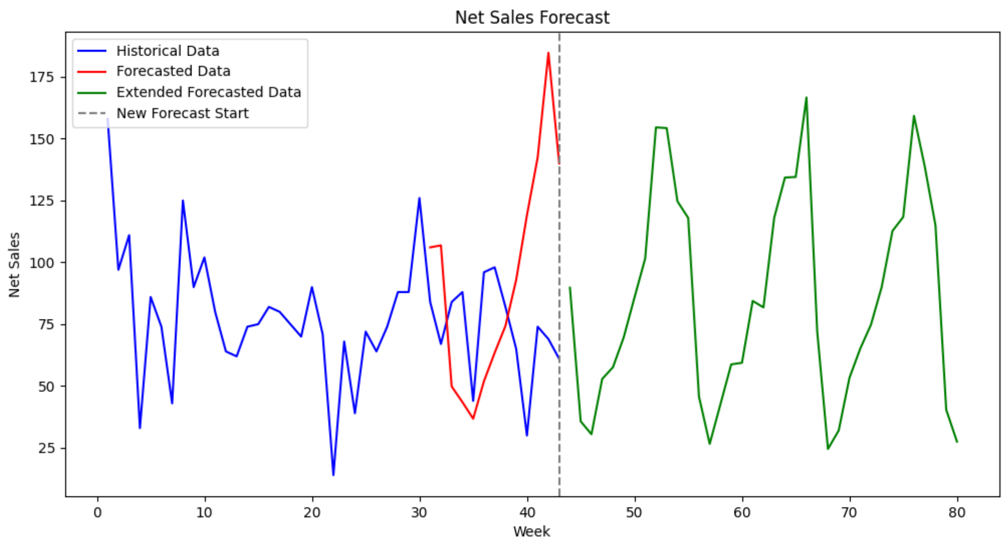

Figure 3. The long-term forecast using neural networks

4.4. Training and validation loss analysis

Python's extensive libraries, such as Tensor Flow and Keras, facilitate the development and fine-tuning of complex neural network architectures. They enable us to visualize and analyze the training and validation loss over multiple epochs. Figure 4 shows the training and validation loss analysis for the neural network forecasting model.

The loss is high at the beginning of the training and validation. This is expected since the model is barely learning the underlying pattern in the data. The rapid decline in both losses within this phase indicates that the model adjusts its parameters very fast to suit the data better. As the training progresses, both the losses continue to decrease, although at a slower rate. Around epoch 200, the losses stabilize, indicating that the model is approaching an optimal set of parameters. The training loss is slightly lower than the validation loss, suggesting that the model fits the training data slightly better, but still generalizes well to the validation data.

In the final phase, both the training and validation losses are consistently low and stable with minimal fluctuations. The stable low validation loss suggests that the model has not overfitted the training data and is performing well on unseen data. This stability is essential for the reliability of the forecasts produced by the model.

4.5. Discussion of Results

The results from the neural network model show that advanced machine learning techniques are effective in predicting complex sales patterns at Al-Youm Food Company. The neural network was more accurate and reliable than traditional methods like exponential smoothing and moving average in both short-term (Weeks 31 to 43) and long-term (Weeks 44 to 80) forecasting scenarios.The training and validation loss analysis underscores the neural network's ability to learn and generalize from the data without overfitting, as evidenced by the convergence of the loss values and their stability over numerous epochs. This reliable performance highlights the model's capacity to capture underlying sales trends and seasonal variations, making it a powerful tool for strategic planning and operational management.

Compared to earlier work that used more fundamental linear models or more traditional time series methods such as ARIMA [18], exponential smoothing, or moving averages [22], the neural network's non-linear nature is a significant strength, in line with findings in [3], [4], and [8]. While more traditional methods can identify overall trends, even cyclical patterns such as those discussed in their descriptions, they generally fail to capture the complex, multifaceted nature of sales data in the dairy business, as highlighted by challenges in applying such methods in various forecasting applications [20, 21, 26]. The neural network's ability to identify such complex relationships, as shown in its better-than-alternative performance in various error metrics in this work and also in applications such as material property prediction [25] and electricity demand forecasting [24], greatly enhances forecast precision and credibility. Such ability is a fundamental improvement over the linear and less-dynamic nature of the more traditional methods discussed.

Applications of neural networks to forecasting have been done in other fields in Jordan, thus setting a very useful benchmark for this study. For example, a study by [23] on the financial performance of Jordanian banks presented marked improvements in the accuracy of neural network-derived forecasts compared to using traditional statistical methods. The study also illustrates that neural networks can outperform more traditional methods, such as exponential smoothing and moving averages in sales forecasting in the dairy industry. Both these

studies further highlight the excellent capability of the neural network in handling complex data patterns and providing timely and accurate forecasts, which is crucial for any decision-making process.

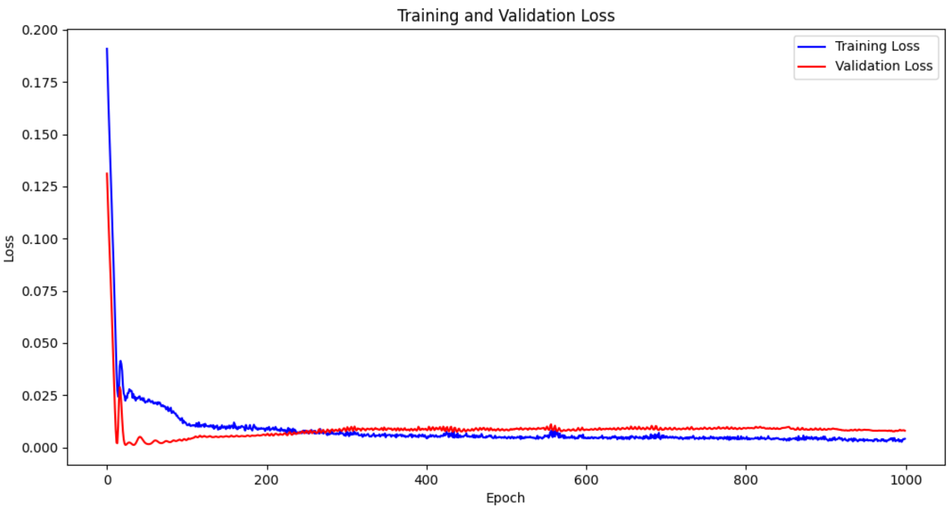

Figure 4. Training and validation loss calculation

A pertinent study [26] examined the projection of wheat production in the Irbid region through the application of linear and log-linear prediction functions based on accumulated rainfall data. Although the study utilized straightforward linear models, it underscored the necessity of incorporating various influential factors, such as rainfall and cultivated areas, into the forecasting model. In contrast, our neural network model implicitly captures such complex relationships and interactions within the sales data, leading to more precise forecasts. This ability to handle non-linear relationships and multiple influencing factors positions neural networks as a more versatile and powerful tool for forecasting in diverse applications.

To summarize, the "Net Sales Forecast" graph (Figure 3) depicts the neural network model's forecasting strength in forecasting dairy sales, comparing forecasted data to historical data of the model. Such precision in forecasting, such as the ability to capture short-term patterns, is of immense significance in the dairy business, owing to its perishable nature of commodities and dynamic demand. In comparison to existing attempts in forecasting in Jordan, there is a striking contrast. While studies such as [18] employed ARIMA models to forecast energy consumption, resulting in MAPE between 6-22%, such models are intrinsically incapable of grasping non-linear patterns. Other studies in Jordan used ANN in other areas, such as forecasting wheat production [26], banking financial performance [23], material science [25], and electricity demands [24], providing avail of neural networks' strength in handling non-linear patterns. The application of ANNs in forecasting electricity demand in Jordan's industrial sector [24] was promising, and our work extends such success to the dairy business, proving again that neural networks hold the potential to make forecasting more accurate in various sectors of the Jordanian economy. This is a clear departure from the results of more fundamental models, such as system dynamics employed to forecast CO2 [21].

The findings of this work have practical implications for the dairy industry. Proper forecasting of sales is not a mere theory; it has a direct bearing on more efficient operations, waste reduction, and increased profitability. By keeping overstock (which leads to wasteful spoilage of perishable dairy products) to a minimum and preventing understock (which leads to missed sales and consumer dissatisfaction), companies like YFC can streamline their supply chain, increase their bottom line, and become more competitive in general. Moreover, the approach developed in this work, though particular to YFC, is transferable and can be applied to other dairy companies that also face a similar forecasting issue, making a more efficient and more sustainable dairy business in general. Finally, while neural networks have been applied to forecasting problems in various domains, their application to the specific challenges of the dairy industry remains low.

5. Conclusion

This study has proved that a neural network-based approach is capable of improving sales forecasting accuracy in the dairy industries, specifically to address the requirements of YFC. The neural network model was trained and tested against 43 weeks of historical sales data and was found to be a significant improvement over the current methods employed by YFC. Quantitatively, the model yielded an RMSE of 19.68 compared to 20.52 using exponential smoothing and 21.98 using the moving average method. Similarly, MAPE using the neural network was found to be 27.76% compared to exponential smoothing's 29.33% and moving average's 30.06%. This is a gain in MAPE of 5.4% over exponential smoothing and a gain of 7.7% over the moving average. Such statistically significant improvement in forecast accuracy shows that the neural network is capable and has future in modeling nonlinear, complex patterns in the sales data that existing methods failed to capture accurately.

Thus, the analysis, through a learned and tested model, confirmed that seasonality and promotional activities were strong factors in affecting sales trends in dairy products, which the neural network model argued for well by making the forecasts more accurate and actionable. Based on findings and insights identified within this present study, the following recommendations are made for further improvement in YFC's sales forecasting processes:

- Given the neural network model's demonstrated superior performance in forecasting accuracy and responsiveness, it is recommended that YFC adopt this model as the primary tool for sales forecasting.
- In order to maintain the accuracy and reliability of the neural network model, it is crucial to regularly update and retrain it with new sales data.
- Additionally, incorporating external factors such as economic indicators, competitor activities, and market trends could further enhance forecasting accuracy.
- YFC should conduct advanced sensitivity analyses regularly to better understand the impact of various factors on sales forecasts.

While this study has demonstrated the effectiveness of neural network models in improving sales forecasting for YFC, several limitations need to be acknowledged:

- The main limitation of this study is the relatively small dataset, which only includes 43 weeks of sales data. This limited data may affect the reliability and applicability of the neural network model, especially for long-term forecasting purposes.
- Additionally, the study mainly concentrated on internal sales data and did not take into account external factors such as economic indicators, market trends, or competitor activities.
- Neural network models are often seen as "black boxes" due to their complex architectures. This complexity can make it challenging to interpret the model's decision-making process and understand the specific factors driving the forecasts.

A significant avenue of work in the future is to perform a broad comparative analysis of various neural network architectures to forecast dairy sales. As this work was focused on a specific neural network realization and its comparison to existing methods in YFC, comparing other architectures like various settings of RNNs, LSTMs, GRUs, and CNNs would provide further improvement in performance. The comparison would also need to be done in terms of predictive precision metrics (RMSE, MAPE) but also in terms of computational efficacy, time taken to train, and resilience to noisy data.

References

[1]Mentzer J T, Moon M A. Sales Forecasting Management: a Demand Management Approach. SAGE Publications, 2004.

[2]M. V. Sebt, S. H. Ghasemi, S. S. Mehrkian, "Predicting the Number of Customer Transactions Using Stacked LSTM Recurrent Neural Networks," Social Network Analysis and Mining, Vol. 11, No. 1, 2021.doi: https://doi.org/10.1007/s13278-021-00805-4.

[3]A. L. D. Loureiro, V. L. Miguéis, L. F. M. da Silva, "Exploring the Use of Deep Neural Networks for Sales Forecasting in Fashion Retail," Decision Support Systems, Vol. 114, 2018, pp. 81–93.doi: https://doi.org/10.1016/j.dss.2018.08.010.

[4]P.H. Wang, G.-H. Lin, Y.C. Wang, "Application of Neural Networks to Explore Manufacturing Sales Prediction," Applied Sciences, Vol. 9, No. 23, 2019, p.5107.doi: https://doi.org/10.3390/app9235107.

[5]Kruse R, Mostaghim S, Borgelt C, Braune C, Steinbrecher M, Multi-layer Perceptrons.In Texts in Computer Science, Cham: Springer, 2022, pp. 53–124.

[6]I. Sutskever and G. Hinton, "Temporal-Kernel Recurrent Neural Networks," Neural Networks, Vol. 23, No. 2, 2010, pp. 239–243.doi: https://doi.org/10.1016/j.neunet.2009.10.009.

[7]Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," IEEE Transactions on Neural Networks and Learning Systems, Vol. 33, No. 12, 2021, pp.1–21.doi: https://doi.org/10.1109/tnnls.2021.3084827.

[8]H. Hewamalage, C. Bergmeir, K. Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions," International Journal of Forecasting, Vol. 37, No. 1, 2021, pp. 388–427.doi: https://doi.org/10.1016/j.ijforecast.2020.06.008.

[9]J. V. Ringwood, D. Bofelli, F. T. Murray, "Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks," Journal of Intelligent and Robotic Systems, Vol. 31, 2001, pp. 129–147.doi: https://doi.org/10.1023/a:1012046824237.

[10]K. G. Boroojeni, M. H. Amini, S. Bahrami, S. S. Iyengar, A. I. Sarwat, O. Karabasoglu, "A Novel multi-time-scale Modeling for Electric Power Demand forecasting: from short-term to medium-term Horizon," Electric Power Systems Research, Vol. 142, 2017, pp. 58–73. doi: https://doi.org/10.1016/j.epsr.2016.08.031.

[11]Dannecker L. Energy Time series forecasting: Efficient and accurate forecasting of evolving time series from the energy domain. Springer, 2015.

[12]C. Faloutsos, J. Gasthaus, T. Januschowski, Y. Wang, "Classical and Contemporary Approaches to Big Time Series Forecasting," International Conference on Management of Data, Amsterdam, Netherlands, 2019.

[13]C. Deb, F. Zhang, J. Yang, S. E. Lee, K. W. Shah, "A review on time series forecasting techniques for building energy consumption," Renewable and Sustainable Energy Reviews, Vol. 74, 2017, pp. 902–924. doi: https://doi.org/10.1016/j.rser.2017.02.085.

[14]Q. Liu, M. F. Iqbal, J. Yang, X. Lu, P. Zhang, M. Rauf, "Prediction of chloride diffusivity in concrete using artificial neural network: Modelling and Performance Evaluation," Construction and Building Materials, Vol. 268, 2021, pp. 121082–121082. doi: https://doi.org/10.1016/j.conbuildmat.2020.121082.

[15]MohitBeniwal, A. Singh, N. Kumar, "A Comparative Study of Static and Iterative Models of ARIMA and SVR to Predict Stock Indices Prices in Developed and Emerging Economies," International Journal of Applied Management Science, Vol. 15, No. 4, 2023, pp. 352–371.doi: https://doi.org/10.1504/ijams.2023.134452.

[16]M. Guermoui, F. Melgani, K. Gairaa, M. L. Mekhalfi, "A comprehensive review of hybrid models for solar radiation forecasting," Journal of Cleaner Production, Vol. 258, 2020, p. 120357.doi: https://doi.org/10.1016/j.jclepro.2020.120357.

[17]T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, R. Jin, "FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting." The 39th International Conference on Machine Learning, 2022.

[18]B. Alsardi, "Jordan Energy Demand Forecast Using ARIMA Model until 2030," Jordan Journal of Mechanical and Industrial Engineering, Vol. 18, No. 03, 2024, pp. 471–483. doi: https://doi.org/10.59038/jjmie/180303.

[19]S. Abaddi, "Forecasting Social Media engagement: an ARIMA-based Approach," International Journal of Agile Systems and Management, Vol. 1, No. 1, 2024. doi: https://doi.org/10.1504/ijasm.2024.10067382.

[20]S. Ding, H. Zhang, "Forecasting Chinese Provincial CO2 emissions: a Universal and Robust new-information-based Grey Model," Energy Economics, Vol. 121, 2023 pp. 106685–106685. doi: https://doi.org/10.1016/j.eneco.2023.106685.

[21]S. Abaddi, "Modeling e-Truck adoption for sustainable transport in Jordan: A system dynamics approach," Natural Resources Forum, 2024. doi: https://doi.org/10.1111/1477-8947.12452.

[22]M.D. Liu, L. Ding, Y.L. Bai, "Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction," Energy Conversion and Management, Vol. 233, 2021, p.113917. doi: https://doi.org/10.1016/j.enconman.2021.113917.

[23]S. Eletter, S. Yaseen, "Applying Neural Networks for Loan Decisions in the Jordanian Commercial Banking System," International Journal of Computer Science and Network Security, Vol. 10, No. 1, 2010, pp. 209–214.

[24]A. Alkhatatbeh, M. A. Gharaibeh, R. Al-Jarrah, A. M. Jawarneh, "Artificial Neural Networks Modeling of the Electricity Demands in the Jordanian Industrial Sector," Jordan Journal of Mechanical and Industrial Engineering, Vol. 18, No. 02, 2024, pp. 297–309. doi: https://doi.org/10.59038/jjmie/180204.

[25]A. Abu Sleem, M. Arafeh, S. Al-Shihabi, R. Obiedat, Y. Al-Zain, "Glass-Reinforced Aluminum Matrix Composite: Synthesizes, Analysis, and Hardness and Porosity Modeling Using Artificial Neural Networks," Jordan Journal of Mechanical and Industrial Engineering, Vol. 18, No. 03, 2024, pp. 509–519. doi: https://doi.org/10.59038/jjmie/180306.

[26]E. Al-Karablieh , A. Salman, F. M. Fisher, "Forecasting wheat production: The case of the Irbid region of Jordan," Journal of International Agriculture, Vol. 41, No. 3, 2022, pp. 191–206.

[27]A. Abu Hanieh, A. A. Hasan, S. AbdElall, and M. Alhanjouri, "Investigating Artificial Intelligence and Modern Technologies Enhancement in Stone and Marble Cutting in Palestine," Jordan Journal of Mechanical and Industrial Engineering, Vol. 18, No. 01, 2024, pp. 207–218.doi: https://doi.org/10.59038/jjmie/180116.

[28]C. A. A. Rashed, Mst. N. Bagum, Md. M. H. Kibria, R. A. Chowdhury, Md. A. Islam, "Integrating Supply Chain Partners through Implementing Industry 4.0 Technologies to Enhance Competitiveness," Jordan Journal of Mechanical and Industrial Engineering, Vol. 18, No. 02, 2024, pp. 351–363. doi: https://doi.org/10.59038/jjmie/180208.

[29]M. Uzair, N. Jamil, "Effects of Hidden Layers on the Efficiency of Neural Networks." 23rd International Multitopic Conference, Bahawalpur, Pakistan, 2020.

[30]G. Lin, W. Shen, "Research on Convolutional Neural Network Based on Improved Relu Piecewise Activation

Function," Procedia Computer Science, Vol. 131, 2018, pp. 977–984.doi: https://doi.org/10.1016/j.procs.2018.04.239.

[31]P. Asteris, P. Roussis, M. Douvika, "Feed-Forward Neural Network Prediction of the Mechanical Properties of Sandcrete Materials," Sensors, Vol. 17, No. 6, 2017, p. 1344. doi: https://doi.org/10.3390/s17061344.

[32]C. Tantithamthavorn, S. McIntosh, A. E. Hassan, K. Matsumoto, "An Empirical Comparison of Model Validation Techniques for Defect Prediction Models," IEEE Transactions on Software Engineering, Vol. 43, No. 1, 2017, pp. 1–18.doi: https://doi.org/10.1109/TSE.2016.2584050.

[33]Rojas R, The Backpropagation Algorithm.In: Neural Networks, Berlin, Heidelberg: Springer, 1996, pp. 149–182.