Journal of Infrastructure, Policy and Development 2024, 8(8), 6639.
https://doi.org/10.24294/jipd.v8i8.6639

Article

Leveraging variational autoencoders and recurrent neural networks for demand forecasting in supply chain management: A case study

Khaoula Khlie[1,*], Zoubida Benmamoun[2], Widad Fethallah[3], Ikhlef Jebbor[4]

[1] Industrial Management Department, Faculty of Business, Liwa College, Abu Dhabi PB 41009, UAE
[2] Industrial Engineering Department, Faculty of Engineering, Liwa College, Abu Dhabi PB 41009, UAE
[3] National School for Applied Sciences, Abdelmalek Essaadi University, Tangier PB 1818, Morocco
[4] ENSA Kenitra, Ibn Tofail University, Kenitra 14000, Morocco
*** Corresponding author:** Khaoula Khlie, Khaoula.khlie@lc.ac.ae

CITATION

Khlie K, Benmamoun Z, Fethallah W, Jebbor I. (2024). Leveraging variational autoencoders and recurrent neural networks for demand forecasting in supply chain management: A case study. Journal of Infrastructure, Policy and Development. 8(8): 6639. https://doi.org/10.24294/jipd.v8i8.6639

ARTICLE INFO

Received: 24 May 2024
Accepted: 12 June 2024
Available online: 26 August 2024

Abstract: Accurate demand forecasting is key for companies to optimize inventory management and satisfy customer demand efficiently. This paper aims to Investigate on the application of generative AI models in demand forecasting. Two models were used: Long Short-Term Memory (LSTM) networks and Variational Autoencoder (VAE), and results were compared to select the optimal model in terms of performance and forecasting accuracy. The difference of actual and predicted demand values also ascertain LSTM's ability to identify latent features and basic trends in the data. Further, some of the research works were focused on computational efficiency and scalability of the proposed methods for providing the guidelines to the companies for the implementation of the complicated techniques in demand forecasting. Based on these results, LSTM networks have a promising application in enhancing the demand forecasting and consequently helpful for the decision-making process regarding inventory control and other resource allocation.

Keywords: generative AI; demand forecasting; inventory management; Recurrent Neural Networks (RNN); Long Short-Term Memory (LSTM); Variational Autoencoder (VAE); comparative analysis

1. Introduction

Supply Chain Management (SCM) is a broad and complex academic discipline that has a task of planning, analyzing, coordinating, implementing and monitoring logistic activities that are manufacturers, sellers, buyers, and users of materiel, supplied materials, money and information through the complete spectrum of organic functioning from raw materials to customers, organizational performance enhancement and value addition (Lummus et al., 2017). This paper finds that SCM is relevant to all business organization because it emphasizes on efficient management of the supply chain network so that cost of operation is reduced, while the needs of the customers are well addressed hence, being able to gain a competitive edge. As noted above, the outcomes of the effective SCM practices are critical for the flows to run smoothly in the business processes and for the products to be delivered to the right place at the right time in the right condition.

Over the last couple of years, Artificial Intelligence (AI) has gained much attention as a tool to improve SCM processes by analyzing large volumes of data, learning from existing patterns and making effective decisions. AI techniques have been increasingly incorporated into various dimensions of SCM to enhance efficiency, reliability and adaptability (Hangl et al., 2022). Despite intensive research on the application of AI in SCM, there is still a gap in understanding the comparative effectiveness of different Generative AI Models in specific SCM functions, such as demand forecasting. The existing literature focuses on individual Generative AI methods without providing a comprehensive comparison of their performance in real-world scenarios. Moreover, while various Generative AI techniques have been proposed and evaluated, there is a lack of empirical evidence on their applied implementation and impact on SCM deliverables. This gap in the literature underlines the need for expanded research efforts to compare the performance of different Generative AI models in enhancing demand forecasting accuracy within SCM.

In this context, this paper aims to address the identified research gap by exploring the practical implementation of two prominent Generative AI techniques—Variational Autoencoders (VAEs) and Recurrent Neural Networks (RNNs)—in SCM, focusing on their application specially in demand forecasting. By conducting a real case study, this research seeks to determine which of these techniques best enhances forecasting accuracy. The innovative aspect of this study lies in its comparative approach, providing empirical evidence on the effectiveness of VAEs and RNNs in a practical SCM context. Through this work, we aim to contribute to the existing body of knowledge by offering insights into the practical implications and benefits of using Generative AI techniques in SCM. The main objective of this paper is to determine which of these techniques best enhances forecasting accuracy through a real case study.

For that, the rest of the paper is divided as follows: Section 2 elaborates on the existing literature and identifies the gap. Section 3 provides the proposed methodology of the study followed by the case study and major findings in section 4. The implications and conclusion are given in section 5.

2. Literature review

2.1. Variational Autoencoders (VAEs)

VAEs is a generative model that amalgamates the things found in autoencoders and variational inference. They were originally proposed by Kingma and Welling in (Kingma and Welling, 2014), but later became famous among the researchers from a paper one year later (Germain et al., 2015). Here’s how VAEs work:

Encoder: Like autoencoders, VAEs have an encoder network that takes input data and maps in a low-dimensional representation. Nevertheless, rather than directly outputting the input as a single point in the latent space, the encoder decodes the parameters of a probability distribution (typically a Gaussian one) that describes the latent space.

Decoder: VAEs, furthermore, possess a decoder network that, given a point from the latent space, which is sampled from the distribution that is parameterized by the encoder, recreates the initial input data.

Objective Function: VAEs are trained by learning to maximize a lower bound on the log-likelihood of the data under the model. This objective function consists of two parts: an evaluation (reconstruction) term which measures how well the expert can recreate data and a normalization term usually equal to the Kullback-Leibler (KL) divergence between the distribution created by the encoder and a prior distribution, usually a standard Gaussian.

The primary innovation of VAEs is the transition to probabilistic models for producing random data in the latent space (Singh and Ogunfunmi, 2021). By obtaining the distribution over the latent space instead of merely a single point, VAEs create the generated data points with smoother transitions between different data points. Instead of the conventional autoencoders, GAN produces a more consistent and organized data generation. VAEs have proven to be a tool with many applications, including image generation, image interpolation, semi supervised learning, and so on. Various extensions and improvements to VAEs have been proposed to address these limitations, such as Conditional VAEs (CVAEs) and Adversarial Autoencoders (AAEs) (Baskin, 2020).

2.2. Recurrent Neural Networks (RNNs)

RNNs are created to deal with the vector sequences by adding the loops that allow the information through different elements (Almaleh et al., 2023; Salehinejad et al., 2018). These models are not typical Neural networks as they are designed for those problems, where input and output data contain sequences, for example, NLP, speech recognition, time series predicting, etc. The Key characteristics of RNNs include There are various key characteristics of RNNs, which include the following;

Recurrent Connections: The nodes of the RNNs are inter-connected and go through feed through the previous network information and back and forth connections. This method makes them able to minimize sensors by extracting features based on time characters.

Hidden State: Supporting their hidden state vector, RNNs use it to represent the remembering part of their series of lightweight operations. The hidden state is re-updated at each time step when the network takes input sequences into account and gets information about the context of the sequence.

Variable Length Inputs/Outputs: RNNs can deal with unpredictable lengths of an input sequence which is what makes them ideal equivalents in situations where an input size can range (e.g., sentences can be of different lengths in NLP). In the same fashion, they can generate output sequences that are specially designed to be either long or short.

Training: In the case of the BPTT algorithm, RNNs are iteratively trained using an extension of the backpropagation algorithm to manage sequential tasks. BPTT calculated forward the gradients along the recurrent connections of the network to sum them up and then updated network parameters based on the summation.

Together with the vanishing gradient difficulty, which limits RNN's ability to fully capture long-term sequences (Chen et al., 2023), this deep architecture addresses the end-of-sentence prediction task. To address this issue, several advanced RNN architectures have been developed, including Long Short-Term Memory (LSTM): LSTMs were the first neural network architecture proposed by Hochreiter and Schmidhuber (1997) in 1997. These networks employ memory cells and gating schemes that make it possible to overcome the vanishing gradients. In particular language models are capable of memorizing long-term dependencies and they are often used for sequence model learning. GRU: The authors Cho et al. (2014) propose in their work GRU which is similar, yet simpler, to LSTM. Furthermore, they couple the gating mechanisms to direct the flow of information in the network that are more accurate than the LSTMs. Phyu and May introduce a Two-Tier LSTM model that enhances caption quality and efficiency for Image captioning by combining image and language processing to generate descriptive captions using CNNs for image tasks and RNNs, including LSTMs, for language tasks (de Castro Moraes et al., 2023). Early healthcare prediction using Big Data and machine learning is crucial for timely treatment. This study introduces an RST-RNN model that outperforms traditional methods in disease prediction accuracy (Talasila et al., 2020).

Along with RNN and its various kinds, they have been implemented to many other things, including language modeling, machine translation, sentiment analysis, speech recognition, and so on. Such networks serve as the basis of deep learning and are useful for professions that involve data whose order is important.

2.3. Overview of supply chain management

SCM is the process of managing the flow of goods, services, information, and finances as they move from the supplier to the manufacturer, wholesaler, retailer, and ultimately to the end consumer (Raja Santhi and Muthuswamy, 2022). It involves the coordination and integration of various activities, including procurement, production, inventory management, logistics, and distribution, to ensure the efficient and effective movement of products or services from the point of origin to the point of consumption. Key components of SCM include (as shown in **Figure 1**):

Figure 1. Key components of SCM.

The constituents of SCM classical include procurement, production, stock, transportation, distribution, and good information flow (Trivedi and Negi, 2023). The procurement process includes gathering of raw materials, concluding agreements, and administering supplier dealings. The manufacturing activities are planned and executed considering demand forecasts of goods and quality control. Inventory management is an important instance to avoid stock shortages but always can hold down costs with various activities like demand forecasting and safety stock handling. Logistics is focused on the organizing of movements of merchandise from a supplier to a customer, including transportation and storage (Benmamoun et al., 2024). The distribution channel ensures that the product reaches the end consumers via multiple associated channels, while the information flow helps in communication with the various stakeholders. Successful SCM minimizes costs, enhances customer service, achieves timely responses and handle risks effectively, by optimizing processes and being in synchronization with the market and disruptions (Benmamoun et al., 2024).

2.4. Applications of VAEs supply chain management

VAEs have gained a broad attention in SCM for their ability to generate realistic data samples and streamline operations in various areas such as demand forecasting, inventory management, and quality control. Several studies have shown the effectiveness of VANs in generating synthetic data for predictive training models, leading to an improvement of forecasting accuracy and a decrease of supply chain variabilities (Jackson et al., 2024; Yilmaz and Korn, 2022).

VAEs have been used in SCM mainly for tasks such as demand forecasting, inventory management, and anomaly detection. Thus, VAEs enable efficient data compression and reconstruction, by learning latent representations of complex data, leading to improved forecast accuracy and inventory management (de Bruijne et al., 2021). Moreover, VAEs can help with detecting anomalies in Supply Chain by identifying deviations from normal patterns and thus, help enhance risk management strategies.

The use of VAEs in SCM covers multiple functions that aim to increase operational efficiency and value creation in a decision-making process as well. According to the VAEs, there are two main effects: First, forecasting based on historical demand distribution, which helps in generating probabilistic forecasts with accommodating for uncertain factors, which further leads to dynamic inventory policy changes based on demand distribution (Luleci and Catbas, 2023). On top of these, Anomaly Detection and Quality Control are also facilitated by VAEs in an extremely efficient way since VAEs can be used to identify deviations from the normal distribution of data obtained via manufacturing processes and product attributes, which helps in discriminating defects and ensuring that product quality is of high standards (Jebbor et al., 2023). AI models are evident in Supply Chain Risk Management as AI models in VAEs are the ones that distribute the critical risk factors which in turn aid in the identification and assessment of potential risks and finally enable scenario planning to mitigate the disruptions (Pan et al., 2023). The VAE technique of simulating different scenarios mimics real-life situations to evaluate an optimization strategy based on the distribution of perturbations of the input variable, particularly in driving efficiency and logistics planning (Benmamoun et al., 2023) where VAEs optimize transportation variables such as loading time and shipment time. Furthermore, VAEs help in Customer Segmentation and personalization in a way that they show the latent customer segments and individual preferences; thereby the marketing strategies and demand based SCM processes can be customized to the eye of the consumer. Using these applications, VAEs provide possibilities to increase and improve the effectiveness of SCM and make decisions quickly.

2.5. Application of Recurrent Neural Networks (RNNs) in SCM

RNNs are widely used in SCM for sequential data analysis, time series forecasting, and NLP tasks (Schroeder and Lodemann, 2021). In SCM, RNNs have been applied for demand forecasting, supply chain optimization, and predictive maintenance. The ability of RNNs to capture temporal dependencies in sequential data ensures accurate predictions of future trends and patterns in supply chain processes.

In SCM, RNNs offer a multitude of applications that enhance operational efficiency and risk response strategies. At first, in Demand Forecasting, RNNs excel by analyzing historical sales data to detect temporal patterns and trends, providing more accurate forecasts by integrating factors such as seasonality and promotional events. Their proficiency in handling sequential and time-series data enables granular forecasts at the SKU level, store level, or regional level (Schroeder and Lodemann, 2021). Furthermore, RNNs play a key role in Inventory Management by predicting future inventory levels and optimizing replenishment policies, in the aim of minimizing shortages or excess inventory costs (Wang and Hong, 2023). They dynamically adjust safety stock levels based on real-time demand fluctuations and lead time variations. Moreover, in Supply Chain Risk Management, RNNs analyze historical data to identify patterns indicating potential disruptions, enabling proactive risk mitigation by monitoring factors like supplier performance and transportation delays. The initial analysis of the information is based on the monitoring of the risk which is exposed in real time and can detect inconsistencies. Furthermore, RNNs contribute to the Predictive Maintenance where the sensor data is used to anticipate the maintenance requirements and avoid the early downtime, and for determining the proper schedule for maintenance to reduce the impact of downtimes (Molęda et al., 2023). Finally, in Transportation Optimization, RNNs predict the demand of transport, estimate the delivery time and plan the routes of the transport including traffic situation and possible capacity, thereby decreasing the transportation expenses. The current route optimization enables effectiveness because of the flexibility in satisfying the real-time demand and delivery shifts. In the above various usages, RNNs have a huge role of managing the SCM as much as possible and containing uncertainties.

Both models, namely, VAEs and RNNs have their distinct features when it comes to SCM. VAEs are particularly good at learning encode-conditional distributions of data and generating new samples from them. Concerning their use in SCM, they can be employed in tasks like feature extraction, identifying anomalies within the supply chain, and risk evaluation –endeavors for which comprehending the structure of the underlying data and modeling uncertainties are crucial. Last but not the least, RNNs are designed specifically for sequential data and are well-suited for the type of problems that have dynamic characteristics such as time series analysis, decision making in sequences, and text analysis. In SCM the usage of RNNs can be defined in many applications like demand forecasting, production planning and scheduling, inventory control and management and logistics and material management. Therefore, the type of model or a combination thereof will depend on the nature of SCM issue/problem and the/its goals or objectives of the organization, where organizations embrace the strength in each of the models to improve decision making and drive operational excellence on their supply chain management operations.

2.6. Use of VAE and RNN models in demand forecasting

VAEs are selected for demand forecasting in the context of supply chain management because of the capacity to identify non-linear and temporal structures in a dataset. VAEs are proven to be good at learning the encoded representation while on the other hand RNNs are good at handling the sequential data. This makes it possible to derive informative features from demand history data and make accurate forecasts. Based on the literature review, VAEs and RNNs are capable of handling different forms of demand data such as seasonality, trends and irregular pattern which are well suitable for volatile supply chain context. Utilizing these models leads to superior forecast errors and complex data patterns and temporal dependencies, the models' resistance to noise or gaps in the data, and superior long-term forecasting. However, the archive stored by VAEs and the modeling by RNNs need substantial computational power and knowledge in deep learning approaches; moreover, interpretability of the models is rather intricate. Nevertheless, given the enhancement of forecast performance as well as the model performance and robustness, VAEs and RNNs are efficient tools to develop the field of demand forecasting in Supply Chain Management.

3. Methodology

In this study, we adopted an integrated model to forecast the demand using recurrent neural networks (RNNs). The dataset contained 1000 demands having the column of date, item ID, and demand as well. First, primary data was collected which was done by obtaining historical demand data from the records of the company. Encoding of the data also involved eliminating unwanted features, data cleansing and dealing with missing values in the data, and normalization of the data for further use with the aim of making the values fine-tuned and reasonable as shown in **Figure 2**. This specific characteristic of VAE model states the output shape of the model is equal to the input shape to yield data which can be re-inserted into the initial data set. For predictive modeling to be used in our work, we used RNN architecture since the latter is suited in the analysis of time series data. First, we employed the Long Short-Term Memory (LSTM) networks, RNN that is famous for the capability of long-term dependencies learning in sequences. The model is trained on 19,301 row of the given dataset and is checked on another phase checking the generality of the model. Mean square root error (MSRE), mean absolute percentage error (MAPE) and mean absolute error (MAE) were applied to assess the accuracy of the generated forecast by every model. By employing this particular approach, our study seeks to establish an accurate prognosis for demand that will improve the strategic planning and executives' decision-making for the retail business industry.

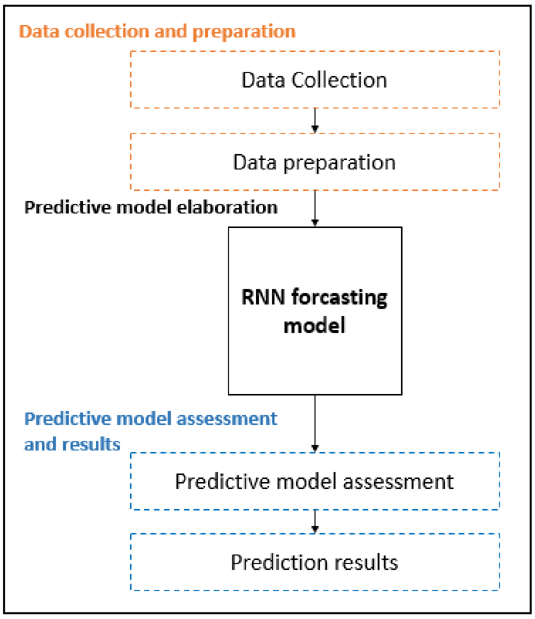

Figure 2. Methodology illustration.

3.1. VAE model: Architecture, training strategy and tuning

3.1.1. Model architecture

In the creation of AI models, the architecture of the models is central in defining the capacities and efficiency of the resultant models, where VAE architecture rises most prominently. VAEs possess the capability of mapping input data into a compressed latent space and can reconstruct the input data since it contains encoder network, decoder network, and latent space. This makes VAE models flexible for many fields (Zietlow et al., 2021). In SCM, VAE architectures are trained for specific tasks such as demand forecasting or quality control, with adjustments made to network architecture, layer sizes, and activation functions to capture relevant features of the supply chain data. Moreover, in the Variational Recurrent Autoencoder (VRAE), VAEs' functions are developed to handle sequential data relevant in SCM due to temporal dependencies (Lin et al., 2024). In VRAE, either RNN, LSTM, or GRU is used in both encoder and decoder so that the dependency with time can be captured. This architecture is involved in functions of the supply chain management, for instance in the demand forecasting that largely involves temporal characteristics in data. Consequently, the design flexibility and complexity of the architecture of VAE and VRAE aids in enhancing operations for SCM as these models can also learn and examine primary features and concerning temporal characteristics of supply chain data (Cheng et al., 2019).

3.1.2. Training strategies

Variational Inference and Stochastic Gradient Variational Bayes (SGVB) play crucial roles in training VAEs efficiently and effectively, particularly in the context of SCM (Simian et al., 2022). Variational inference approximates posterior distributions in probabilistic models like VAEs by optimizing a variational objective function. SGVB, a variant of variational inference, utilizes stochastic optimization techniques such as stochastic gradient descent (SGD) to efficiently optimize the variational objective function. These techniques are useful in training VAEs for SCM and can make the VAEs learn the posterior distribution of latent variables compiled with the observed data and learn rich representation of supply chain data. Additionally, Furthermore, Reconstruction Loss and Kullback-Leibler (KL) Divergence are used elements of the VAE objective function (Giannakopoulou et al., 2022). Reconstruction loss calculates the difference between input data and the reconstructed data produced through the VAE's decoder; the KL divergence looks at the discrepancy between the probability distributions and is often used to control the distribution of the latent space in VAEs. In the context of SCM applications, reconstruction loss and KL divergence bear considerable importance while formulating the trade-off between reconstructing actual data with high accuracy and preventing entanglement of the latent space representation with irrelevant characteristics of the supply chain data with regards to original supply chain phenomena, on the one hand, and, on the other hand, providing VAE with meaningful and interpretable latent space representation.

3.1.3. Hyperparameter tuning and optimization

Hyperparameter tuning for VAEs in SCM is essential for optimizing model performance. This process involves fine-tuning parameters like learning rate, batch size, latent space dimensionality, and regularization parameters to enhance VAEs' effectiveness in SCM tasks (Akkem et al., 2024). Different methods like grid search (El Filali et al., 2022), random search, and Bayesian optimization are used to optimize hyperparameters which are beneficial for VAE models that are required for SCM. Also, the proper choice of optimization algorithm plays an important role in effective training of VAEs. During training of VAEs, methods such as the stochastic gradient descent (SGD), Adam, RMSprop or adaptive moment estimation (Adamax) can typically be used to optimize the variational objective (Liao et al., 2022). Due to these dependencies between the optimization algorithm and its hyperparameters and the performance of VAEs in SCM tasks, it is suggested that: Therefore, it can be deduced that the selection and tuning of optimization algorithms play a crucial role for getting better results in SCM applications. Thus, if Appropriate VAE models and architectures are adopted and trained, using the above methodologies and techniques, the concepts and features of the supply chain data gets captured more precisely and these models improves the decision-making processes in SCM.

3.2. Recurrent Neural Network (RNN) model: Architecture, training and tuning

A recurrent neural network (RNN) is a type of artificial neural network distinguished by its ability to retain information from previous time steps. This feature makes RNNs particularly well suited for processing sequential data, which is prevalent in various scientific and practical applications, including forecasting demand in SCM (Abbasimehr et al., 2020). In supply chain management, accurate demand forecasting is essential for optimizing inventory levels, production schedules, and distribution strategies. RNNs have emerged as powerful tools for forecasting demand in this context.

3.2.1. Model architecture

LSTM and GRU were selected from the family of the RNN structures, which are often used in SCM due to their utility in modeling sequential data (Schroeder and

Lodemann, 2021). LSTM which is a type of Recurrent Neural Network specifically addressed the vanishing gradient problem in sequences and long dependencies, is widely used in SCM activities like demand forecasting, inventory control, and anomalous behavior detection. It is equally suited in applications that require it to memorize information over long sequences because of temporal dependencies involved in the application decision-making process. On the other hand, GRU which stands for Gated Recurrent Unit structures another subclass of LSTM but with more basic architecture and less parameters for training, is another middle ground between model complexity and the effectiveness of the model. In SCM, GRU networks are also applied to, for example, demand forecasting, inventory management, or predictive maintenance processes, in which proper modeling of sequential data is critical (Serrou et al., 2016). The main advantage of using GRU is that it has relatively simple structure than LSTM and it is computationally efficient to learn from the supply chain data while it is able to capture temporal patterns as well. In general, LSTM and GRU networks are useful in SCM, as they improve modeling and forecasting of sequential data, as well as enrich decision-making across multiple fields in SCM.

Using trends of past demand, other attributes such as seasonal variations and promo events, as well as economic variables, RNNs thus produce accurate forecasts that enable organizational to better predict future demands. **Figure 3** represents the general overview of the RNN model mainly focusing on the Recurrent connections that permits information to pass from one time-step to the next.

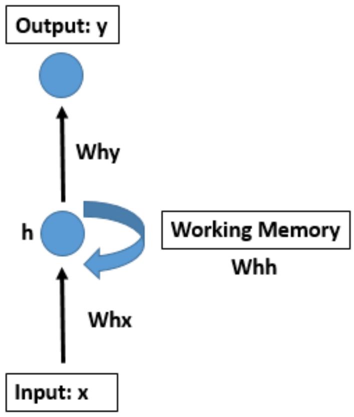

Figure 3. A recurrent neural network illustration.

where:

$X = \{x_1, x_2, x_3 \ldots x_t\}$

$H = \{h_1, h_2, h_3 \ldots h_t\}$

$Y = \{y_1, y_2, y_3 \ldots y_t\}$

A sequence $X = \{x_1, x_2, x_3 \ldots x_t\}$ as input, the RNN computes both a hidden state sequence,

$H = \{h_1, h_2, h_3 \ldots h_t\}$, and an output sequence, $Y = \{y_1, y_2, y_3 \ldots y_t\}$, using Equations (1) and (2). This mechanism allows the RNN to capture temporal dependencies within the data and make predictions or classifications based on the context provided by previous elements in the sequence.

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$y_t = g(W_{yt}h_t + b_y) \quad (2)$$

In Equations (1) and (2), W_{hx}, W_{hh}, and W_{yt} represent the input-to-hidden, hidden-to-hidden, and hidden-to-output weight matrices, respectively. The vectors b_h and by denote the bias vectors for the hidden layer and output layer. The activation functions $f(\cdot)$ and $g(\cdot)$ are applied to the hidden layer and output layer, respectively. The hidden state of each time step is passed to the hidden state of the next time step (Nguyen, 2019).

Figure 4 present a visual representation of how an RNN processes sequential data in an unfolded manner. When applied to demand forecasting in supply chain management, the RNN computes both hidden state sequences and output sequences, leveraging temporal context to generate accurate forecasts of future demand. These forecasts can then be used to inform inventory management decisions, production planning, and other aspects of supply chain optimization, ultimately leading to improved operational efficiency and customer satisfaction.

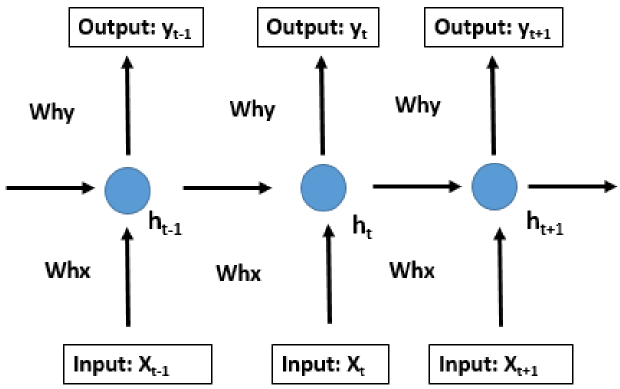

Figure 4. An unrolled recurrent neural network illustration.

3.2.2. Training strategies

In training RNNs for SCM tasks, several strategies are employed to enhance training efficiency and effectiveness. Backpropagation Through Time (BPTT) remains a significant training algorithm unrolling the network over a temporal horizon to compute gradients via backpropagation (Sefati et al., 2023). In SCM, BPTT is used to train RNNs to accomplish activities such as demand forecasting for Inventory management, where such vectors are crucial in delivering good results. Moreover, there is the constant use of such methods as Teacher Forcing and Scheduled Sampling to enhance training stability and convergence. Teacher Forcing is a training strategy that feed the model with ground truth output instead of the one generated by the model while the Scheduled Sampling is a training strategy in which the model gradually transitions from using the ground truth to its output fed to it during training. Such techniques are particularly helpful to SCM applications as it improves the RNN training's reliability and convergence; crucial in time-sensitive features such as in demand forecasting and inventory management where temporal patterns are vital to optimize model performance (Bassiouni et al., 2023). By training the RNNs in such a manner, temporal features in the supply chain data can be captured hence enhancing the decision-making processes in Supply Chain Management the in such training approaches.

3.2.3. Hyperparameter tuning and optimization

Among the standards, hyperparameters tuning and optimization are especially important for enhancing the performance of RNNs after a specific application of SCM. May techniques like Grid Search and Random Search, discover the hyperparameters in a structured and methodical manner, in a quest for a combination that produces the most optimal model outcome (Rimal et al., 2024). On the other hand, Bayesian Optimization inductively builds a model of the objective function and helps to find better hyperparameters in consecutive steps, using probabilistic models. These methods are used instrumentally in the context of SCM for tuning of hyperparameters of RNN models: learning rate, batch size, number of layers, and the number of hidden units. Moreover, learning rate scheduling and early stopping are the two considered methods commonly used to improve the training of RNNs. Learning rate scheduling enables the tuning of the learning rate over a period in training to avoid overshooting or stalling during the optimization process; early stopping on the other hand halts the training process when the model's performance on the validation set starts to worsen thereby avoiding overfitting. In SCM, these techniques are mostly used for tasks including demand forecasting and inventory management where overfitting should be avoided, and the accuracy of the model should be improved. Therefore, this paper's hyperparameter tuning and optimization approach can effectively improve the RNN model in SCM and improve the decision-making and operational performance of SCM.

3.3. Long-Short Term Memory model

Long-Short Term Memory (LSTM) networks are a type of recurrent neural network (RNN) that can learn long-term patterns. They were first introduced by Hochreiter and Schmidhuber (1997) in 1997 and have since been refined by many researchers for various applications. LSTMs are specifically designed to overcome the problem of losing important information over time, which is common in traditional RNNs.

Compared to simple RNNs, LSTMs have a more complex structure, including three key components: the input gate, the forget gate, and the cell state gate. These gates help control the flow of information within the LSTM cell as shown in **Figure 5**.

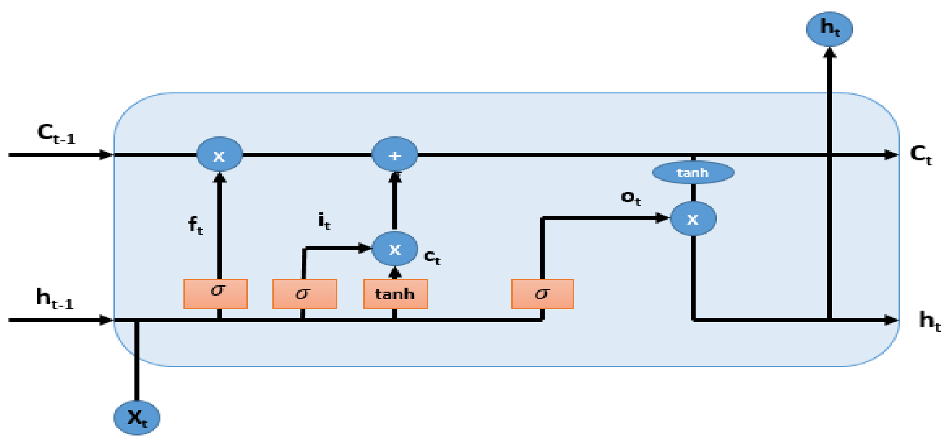

Figure 5. A Long-Short Term Memory cell illustration.

The forget gate operates after receiving the output from the previous state, denoted as $h(t-1)$. Its purpose is to decide what information should be discarded from the previous state $h(t-1)$, retaining only the relevant information. It uses a sigmoid function to scale the input between 0 and 1, determining how much of the information should be forgotten and retained.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3}$$

In the input gate, we determine whether to incorporate new information from the current input into our current cell state. This decision is based on how much importance we assign to these new pieces of information. The input gate utilizes a sigmoid layer to decide which values should be updated and a hyperbolic tangent (tan$h$) layer to generate a vector of new candidates to be added to the current cell state. Equation (4) describes the process of determining which values to update, while Equation (5) outlines how the new candidate values are calculated using the tan$h$ function.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$\hat{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_c) \tag{5}$$

Next, the cell state is calculated using Equation (6).

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \tag{6}$$

The output gate is the final step in determining what information to output from the cell state. Equation (7) shows how a sigmoid function helps make this decision. We first use the hyperbolic tangent (tan$h$) function to compress the values between −1 and 1.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

Then, we multiply this result with the output of the sigmoid function to filter out only the information we want to output. Equation (8) further illustrates this process.

$$h_t = o_t \tanh(C_t) \tag{8}$$

Supply Chain Management is one of the many application areas of the recurrent neural network model described in the text. This model can be highlighted by this feature as its main advantage, as the sequence of variables is typically used when predicting demand. As opposed to most models, RNNs can learn temporal relations in the data, which makes the forecasts more accurate adapted on the past behavior. Also, specifically, input gates, forget gates, and cell state gates in LSTM networks appropriately regulate information in the long term and improve the model's performance in discovering the patterns at different time horizons. Altogether, the RNN model and LSTM networks are considered as a breakthrough achievement within the field of demand forecasting, as they give more accurate and effective tools for inventory management, production planning, and other links of the Supply Chain Management.

4. Case study

Every company must stay on top of planning activities to meet the demand for goods based on customer needs. An accurate demand forecast is crucial for predicting which products are required at each location or within the organization. This ensures high availability for customers while minimizing stock risk and supporting capacity management, labor force planning, and other operational aspects.

This paper utilizes Long Short-Term Memory (LSTM) networks, which are well suited for handling time-series data and widely employed for forecasting purposes. These networks will help the company predict future demand trends accurately, allowing for better inventory management and resource allocation to meet customer needs effectively. In addition, the results obtained from the LSTM forecasting model will be compared with that gene rated by Variational Autoencoder (VAE) models. This comparison will provide insights into the effectiveness and accuracy of each approach in predicting demand for the company's products. By evaluating both LSTM and VAE models, the company can determine which method best suits its forecasting needs and supports informed decision-making regarding inventory management and resource allocation.

4.1. Data collection

For this study, data collection involved gathering daily statistics from a selected company. These statistics encompassed 1000 raw data points on demand volume. The data collection process aimed to capture a comprehensive overview of demand dynamics over time as shown in **Figure 6**. The dataset includes information such as the date, ID item, and demand quantity for each observation. Here is an example of the overall daily demands.

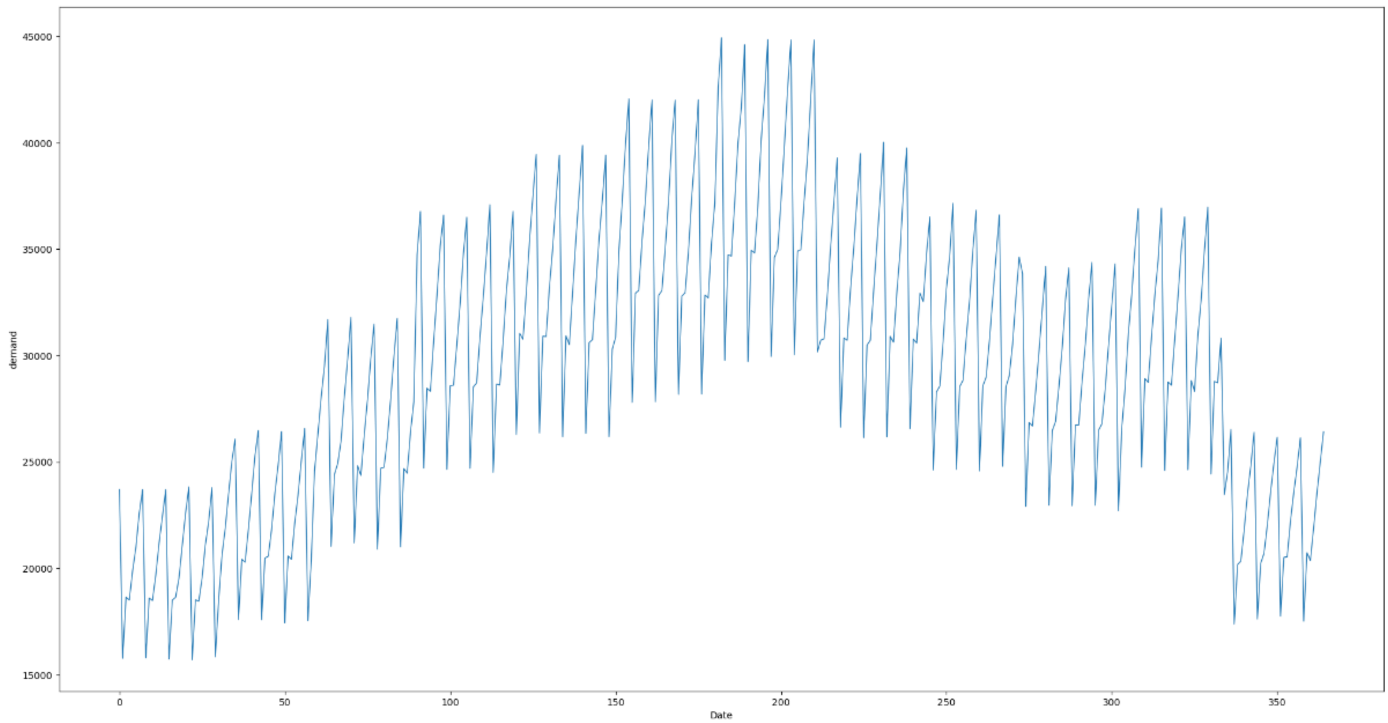

Figure 6. Overall daily demands illustration.

4.2. Data preparation

In this data-preprocessing phase, we initially cleaned the dataset obtained from previous steps, ensuring its suitability for further analysis. Specifically, we subsampled the training set to exclusively include data from last year, a strategy aimed at reducing training time while retaining recent information for modelling purposes.

Our training dataset consisted of 913,001 entries, with columns for date, store, item ID, and demand. Following the sub-sampling, the dataset underwent transformation into a time series problem, a crucial step enabling the utilization of historical data to make future predictions. This transformation allows us to incorporate temporal dependencies and patterns into our predictive models. Furthermore, we addressed the issue of consistency within the dataset by dropping rows where the item/store values differed from the shifted columns.

This step ensures that the dataset remains coherent and aligned, facilitating accurate modeling and prediction. Finally, the dataset was split into training and testing sets, a fundamental procedure in machine learning model development. This division allows for the evaluation of model performance on unseen data, enabling us to assess its generalization capability effectively.

Throughout these steps, the dataset's structure, consisting of columns such as date, ID item, and demand, remained integral. These attributes serve as the foundation for analyzing demand dynamics and developing predictive models to forecast future demand accurately.

4.3. LSTM implementation

In this study, the LSTM model comprises four hidden layers, with the parameters of neurons influencing the total trainable parameters. Additionally, the output layer consists of a neuron responsible for predicting demands numbers. The training dataset is split into 95% for training and 5% for testing.

However, during training, the LSTM model may encounter overfitting issues, where it becomes too specialized to the training data and performs poorly on unseen data. To address this, the study utilizes the mean absolute error (MAE) as the loss function to measure in-sample error. The model optimization employs the Adam optimization algorithm with a learning rate set to 0.02.

The training process spans 100 epochs, where each epoch represents training the model on the entire training dataset 100 times. The batch size, which refers to the number of samples used in each training iteration, is set to 1. Additionally, a rolling training method is employed, where the initial training values are discarded after each prediction to prevent bias in subsequent predictions.

4.4. Discussion

In our study, the data used specifically originates from a company operating within the Gulf countries. This data is collected from this company, which operates in the Gulf region, and they reflect the demand patterns observed in its business operations. We acknowledge the importance of clarifying this information to ensure an accurate understanding of the data's origin and its geographical relevance. This clarification will be integrated into our revised text to provide appropriate contextualization of the data used in our analysis.

In **Figure 7**, the forecasting of demands is depicted, with actual demands presented in **Figure 7a**, predicted demands using the RNN model in **Figure 7b**, and predicted demands using the VAE model in **Figure 7c**. Upon thorough analysis, it becomes evident that the RNN model outperforms the VAE model in predicting demand values. The RNN model exhibits higher accuracy and demonstrates a closer alignment with the actual demand values, as showcased in the comparison of actual demand, RNN demand, and VAE demand in **Figure 7d**. These findings underscore the superior performance of the RNN model in capturing the underlying patterns and trends in the data, particularly in scenarios where capturing sequential dependencies and temporal dynamics is crucial for accurate predictions. These results provide an assessment of the performance of the RNN and VAE models in predicting demand compared to the actual values. For the RNN model, the RMSE (Root Mean Square Error) is 11.03 and the MAE (Mean Absolute Error) is 7.81. This means that, on average, the predictions of the RNN model have an error of about 11.03 units for RMSE and an absolute error of about 7.81 units for MAE relative to the actual demand values. In contrast, for the VAE model, the RMSE is 28.14 and the MAE is 21.84. This indicates that the predictions of the VAE model have a higher error, with higher RMSE and MAE compared to the RNN model.

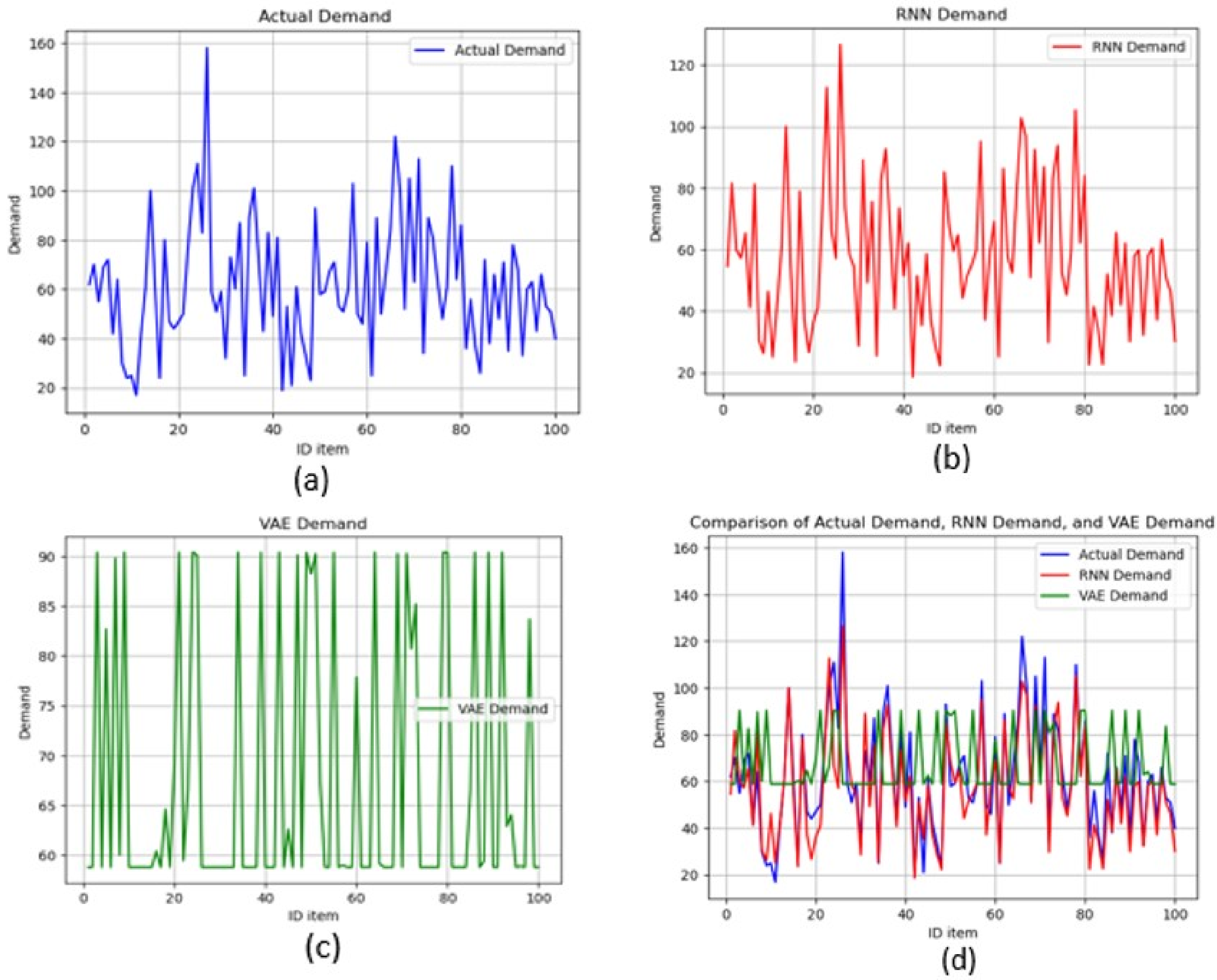

Figure 7. VAE and RNN demand forecasting. **(a)** represents actual demands; **(b)** predicted demands using RNN model **(c)** and **(d)** predicted demands using VAE model; **(d)** comparison of actual demand, RNN demand, and VAE demand.

Interpreting these results, it can be concluded that the RNN model outperforms the VAE model in demand prediction in this scenario. The RNN model produces predictions that are closer to the actual demand values, with lower average errors compared to the VAE model. This suggests that the RNN model better captures underlying patterns and trends in the demand data, making it a more reliable choice for demand prediction.

Table 1 presents a comparison of actual demand values with the predicted demand values generated by both the Recurrent Neural Network (RNN) model and the Variational Autoencoder (VAE) model. Upon analysis, it's evident that the RNN model generally performs better than the VAE model in predicting demand across most items. For instance, for Item 1, the RNN model predicted a demand of 55, closer to the actual demand of 62, whereas the VAE model predicted 59. Similarly, for Item 5, the RNN model predicted a demand of 65, while the VAE model predicted 83, further from the actual demand of 72.

Table 1. Comparison of results.

ID	Actual Demand	Predicted demand (RNN)	Predicted demand (VAE)	Best Predicted model
1	62	55	59	RNN
2	70	82	59	VAE
3	55	60	90	-
4	69	57	59	RNN
5	72	65	83	RNN
6	42	41	59	RNN
7	64	81	90	-
8	30	30	60	RNN
9	24	26	90	-
10	24	25	90	-

However, there are exceptions where the VAE model outperforms the RNN model. For example, for Item 2, the VAE model predicted a demand of 59, which is closer to the actual demand of 70 compared to the RNN model's prediction of 82. Additionally, for Item 3, the VAE model predicted a demand of 90, aligning more closely with the actual demand of 55 compared to the RNN model's prediction of 60.

Overall, the RNN model appears to be the better-performing model, as indicated by its consistently closer alignment with the actual demand values across most items. However, it's essential to consider each model's strengths and weaknesses and select the most suitable model based on specific forecasting requirements and constraints.

5. Conclusion

This paper gives a detailed understanding for the utilization of differential autoencoders (VAEs) and repetitive brain organizations (RNNs) popular determining in store network the board. By examining the gauging execution and contrasting and genuine limit values, we have clarified the unpretentious elements of these high-level brain network calculations in estimating situations.

In the case study, we employed a comprehensive approach to forecasting demand using recurrent neural networks (RNNs). Our dataset consisted of 913,001 entries, with columns for date, store, item ID, and demand. Initially, data collection involved gathering historical demand data from the company's records. Subsequently, data preparation included steps such as data cleaning, handling missing values, and normalization to ensure the quality and consistency of the dataset. For predictive modeling, we implemented an RNN architecture, specifically utilizing Long Short-Term Memory (LSTM) networks due to their suitability for time-series data analysis. The model was trained on a portion of the dataset while being validated on another segment to assess its generalization performance. Evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used to quantify the accuracy of the forecasts generated by the model. Our findings indicated that the RNN model achieved an RMSE of 11.03 and an MAE of 7.81, significantly outperforming the Variational Autoencoder (VAE) model, which had an RMSE of 28.14 and an MAE of 21.84. Through this methodology, we aimed to develop a robust demand forecasting system that could effectively anticipate future demand patterns and aid decision-making processes within the retail industry.

Our study reveals a microscopic situation where the RNN model tends to exhibit more accurate predictions compared to the VAE model. For a range of products, the RNN model shows consistency with actual demand values, which means it can capture the time-dependent complexity and nonlinear relationships of demand data.

The rationale for adopting these sophisticated machine learning techniques lies in the ability to describe complex patterns in terms of demand data, thus facilitating accurate forecasting Using VAE and RNN capabilities, suppliers can gain deeper insights into demand dynamics, enabling them to make more proactive decisions. The advantages of our proposed method extend beyond mere prediction accuracy.

Integrating VAEs and RNNs into supply chain management practices empowers organizations to optimize inventory levels, streamline production processes, and ultimately, enhance customer satisfaction. By harnessing the power of data-driven forecasting methodologies, businesses can mitigate risks, reduce costs, and gain a competitive edge in dynamic market environments.

Looking ahead, further refinement of model architectures, exploration of ensemble forecasting techniques, and integration of additional data sources could contribute to improving predictive performance. Additionally, the evaluation of models in real-world Supply Chain settings and the investigation of uncertainty quantification methodologies present exciting opportunities for advancing the field of demand forecasting.

In conclusion, while the RNN model emerges as the preferred choice for demand forecasting in our study, the journey towards effective supply chain management is ongoing and multifaceted. By embracing innovation, fostering collaboration, and continuously refining forecasting methodologies, organizations can navigate complexities, seize opportunities, and drive sustainable growth in today’s dynamic business landscape.

Author contributions: Conceptualization, KK; methodology, WF; software, WF; validation, KK; formal analysis, IJ; resources, ZB; data curation, WF; writing—original draft preparation, IJ; writing—review and editing, ZB; supervision, KK. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. Engineering Applications of Artificial Intelligence, 131, 107881. https://doi.org/10.1016/j.engappai.2024.107881

Almaleh, A., Almushabb, R., & Ogran, R. (2023). Malware API Calls Detection Using Hybrid Logistic Regression and RNN Model. Applied Sciences, 13(9), 5439. https://doi.org/10.3390/app13095439

Baskin, I. I. (2020). The power of deep learning to ligand-based novel drug discovery. Expert Opinion on Drug Discovery, 15(7), 755–764. https://doi.org/10.1080/17460441.2020.1745183

Bassiouni, M. M., Chakrabortty, R. K., Hussain, O. K., et al. (2023). Advanced deep learning approaches to predict supply chain risks under COVID-19 restrictions. Expert Systems with Applications, 211, 118604. https://doi.org/10.1016/j.eswa.2022.118604

Benmamoun, Z., Fethallah, W., Ahlaqqach, M., et al. (2023). Butterfly Algorithm for Sustainable Lot Size Optimization. Sustainability, 15(15), 11761. https://doi.org/10.3390/su151511761

Benmamoun, Z., Khlie, K., Dehghani, M., & Gherabi, Y. (2024). WOA: Wombat Optimization Algorithm for Solving Supply Chain Optimization Problems. Mathematics, 12(7), 1059.

Chen, Z., Ma, M., Li, T., et al. (2023). Long sequence time-series forecasting with deep learning: A survey. Information Fusion, 97, 101819. https://doi.org/10.1016/j.inffus.2023.101819

Cheng, F., He, Q. P., & Zhao, J. (2019). A novel process monitoring approach based on variational recurrent autoencoder. Computers & Chemical Engineering, 129, 106515. https://doi.org/10.1016/j.compchemeng.2019.106515

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. Available online: http://arxiv.org/abs/1409.1259 (accessed on 4 May 2024).

de Castro Moraes, T., Qin, J., Yuan, X. M., et al. (2023). Evolving Hybrid Deep Neural Network Models for End-to-End Inventory Ordering Decisions. Logistics, 7(4), 79. https://doi.org/10.3390/logistics7040079

de Bruijne, M., Cattin, P. C., Cotin, S., et al. (2021). Medical Image Computing and Computer Assisted Intervention—MICCAI. In: Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-87199-4

El Filali, A., Lahmer, E. H. B., El Filali, S., et al. (2022). Machine Learning Applications in Supply Chain Management: A Deep Learning Model Using an Optimized LSTM Network for Demand Forecasting. Available online: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2185310X&AN=155479850&h=QNfWAvzGl6HNvFi%2FLjb2GUtZRfjGDzA15IGsntYvOkKEF7OvyNJJGmtU%2BgdYRrOTviGX%2B2cFyWSwjN%2BHkPboMQ%3D%3D&crl=c (accessed on 5 May 2024).

Germain, M., Gregor, K., Murray, I., & Larochelle, H. (2015). Made: Masked autoencoder for distribution estimation. Available online: http://proceedings.mlr.press/v37/germain15.html (accessed on 4 May 2024).

Giannakopoulou, K. M., Roussaki, I., & Demestichas, K. (2022). Internet of Things Technologies and Machine Learning Methods for Parkinson's Disease Diagnosis, Monitoring and Management: A Systematic Review. Sensors, 22(5), 1799. https://doi.org/10.3390/s22051799

Hangl, J., Behrens, V. J., & Krause, S. (2022). Barriers, Drivers, and Social Considerations for AI Adoption in Supply Chain Management: A Tertiary Study. Logistics, 6(3), 63. https://doi.org/10.3390/logistics6030063

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jackson, I., Ivanov, D., Dolgui, A., et al. (2024). Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. International Journal of Production Research, 1–26. https://doi.org/10.1080/00207543.2024.2309309

Jebbor, I., Benmamoun, Z., & Hachimi, H. (2023). Optimizing Manufacturing Cycles to Improve Production: Application in the Traditional Shipyard Industry. Processes, 11(11), 3136. https://doi.org/10.3390/pr11113136

Khlie, K. (2024). Optimizing Inventory Control in Supply Chain Management. In: Proceedings of the 2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). pp. 53–56.

Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. Available online: http://arxiv.org/abs/1312.6114 (accessed on 4 May 2024).

Liao, W., Ge, L., Bak-Jensen, B., et al. (2022). Scenario prediction for power loads using a pixel convolutional neural network

and an optimization strategy. Energy Reports, 8, 6659–6671. https://doi.org/10.1016/j.egyr.2022.05.028
Lin, H., Liu, G., Wu, J., et al. (2024). Deterring the Gray Market: Product Diversion Detection via Learning Disentangled Representations of Multivariate Time Series. INFORMS Journal on Computing, 36(2), 571–586. https://doi.org/10.1287/ijoc.2022.0155
Luleci, F., & Catbas, F. N. (2023). A brief introductory review to deep generative models for civil structural health monitoring. AI in Civil Engineering, 2(1). https://doi.org/10.1007/s43503-023-00017-z
Ma, S. (2024). Retail store-SKU level replenishment planning with attribute-space graph recurrent neural networks. Expert Systems with Applications, 249, 123727. https://doi.org/10.1016/j.eswa.2024.123727
Molęda, M., Małysiak-Mrozek, B., Ding, W., et al. (2023). From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry. Sensors, 23(13), 5970. https://doi.org/10.3390/s23135970
Nguyen, N. T., Chbeir, R., Exposito, E., et al. (2019). Computational Collective Intelligence. In: Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-28377-3
Pan, Z., Wang, Y., Cao, Y., Gui, W. (2023). VAE-based interpretable latent variable model for process monitoring. Available online: https://ieeexplore.ieee.org/abstract/document/10149177/ (accessed on 5 May 2024).
Raja Santhi, A., & Muthuswamy, P. (2022). Influence of Blockchain Technology in Manufacturing Supply Chain and Logistics. Logistics, 6(1), 15. https://doi.org/10.3390/logistics6010015
Rimal, Y., Sharma, N., & Alsadoon, A. (2024). The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-18426-2
Salehinejad, H., Sankar, S., Barfett, J., et al. (2018). Recent Advances in Recurrent Neural Networks. Available online: http://arxiv.org/abs/1801.01078 (accessed on 4 May 2024).
Schroeder, M., & Lodemann, S. (2021). A Systematic Investigation of the Integration of Machine Learning into Supply Chain Risk Management. Logistics, 5(3), 62. https://doi.org/10.3390/logistics5030062
Sefati, S. S., Arasteh, B., Halunga, S., et al. (2023). Meet User's Service Requirements in Smart Cities Using Recurrent Neural Networks and Optimization Algorithm. Available online: https://ieeexplore.ieee.org/abstract/document/10210672/ (accessed on 5 May 2024).
Serrou, D., Khlie, K., & Abouabdellah, A. (2016). Improvement of the lean-maintenance by hospital logistics. In: Proceedings of the 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt). pp. 19–24.
Simian, D., & Husac, F. (2022). Challenges and Opportunities in Deep Learning Driven Fashion Design and Textiles Patterns Development. Available: https://link.springer.com/chapter/10.1007/978-3-031-27034-5_12 (accessed on 5 May 2024).
Singh, A., & Ogunfunmi, T. (2021). An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. Entropy, 24(1), 55. https://doi.org/10.3390/e24010055
Talasila, V., Madhubabu, K., Mahadasyam, M., et al. (2020). The Prediction of Diseases using Rough Set Theory with Recurrent Neural Network in Big Data Analytics. International Journal of Intelligent Engineering and Systems, 13(5), 10–18. https://doi.org/10.22266/ijies2020.1031.02
Trivedi, S., & Negi, S. (2023). The metaverse in supply chain management: application and benefits. International Journal of Advanced Virtual Reality, 1(1), 36–43.
Wang, T., & Hong, L. J. (2023). Large-Scale Inventory Optimization: A Recurrent Neural Networks–Inspired Simulation Approach. INFORMS Journal on Computing, 35(1), 196–215. https://doi.org/10.1287/ijoc.2022.1253
Yilmaz, B., & Korn, R. (2022). Synthetic demand data generation for individual electricity consumers : Generative Adversarial Networks (GANs). Energy and AI, 9, 100161. https://doi.org/10.1016/j.egyai.2022.100161
Zietlow, D., Rolinek, M., & Martius, G. (2021). Demystifying inductive biases for (beta-) VAE based architectures. Available online: http://proceedings.mlr.press/v139/zietlow21a.html (accessed on 5 May 2024).