computation

MDPI

Article

MySTOCKS: Multi-Modal Yield eSTimation System of in-prOmotion Commercial Key-ProductS

Cettina Giaconia and Aziz Chamas *

R&D Department, Astrea Consulting Srl, Via F. Bruno, 2 Petralia Soprana, 90026 Palermo, Italy; cettinagiaconia@astreaconsultingsrl.it
* Correspondence: azizchamas@astreaconsultingsrl.it; Tel.: +39-328-210-6781; Fax: +39-338-636-6940

Abstract: In recent years, Out-of-Stock (OOS) occurrences have posed a persistent challenge for both retailers and manufacturers. In the context of grocery retail, an OOS event represents a situation where customers are unable to locate a specific product when attempting to make a purchase. This study analyzes the issue from the manufacturer's perspective. The proposed system, named the "Multi-modal yield eSTimation System of in-prOmotion Commercial Key-ProductS" (MySTOCKS) platform, is a sophisticated multi-modal yield estimation system designed to optimize inventory forecasting for the agrifood and large-scale retail sectors, particularly during promotional periods. MySTOCKS addresses the complexities of inventory management in settings where Out-of-Stock (OOS) and Surplus-of-Stock (SOS) situations frequently arise, offering predictive insights into final stock levels across defined forecasting intervals to support sustainable resource management. Unlike traditional approaches, MySTOCKS leverages an advanced deep learning framework that incorporates transformer models with self-attention mechanisms and domain adaptation capabilities, enabling accurate temporal and spatial modeling tailored to the dynamic requirements of the agrifood supply chain. The system includes two distinct forecasting modules: TR1, designed for standard stock-level estimation, and TR2, which focuses on elevated demand periods during promotions. Additionally, MySTOCKS integrates Elastic Weight Consolidation (EWC) to mitigate the effects of catastrophic forgetting, thus enhancing predictive accuracy amidst changing data patterns. Preliminary results indicate high system performance, with test accuracy, sensitivity, and specificity rates approximating 93.8%. This paper provides an in-depth examination of the MySTOCKS platform's modular structure, data-processing workflow, and its broader implications for sustainable and economically efficient inventory management within agrifood and large-scale retail environments.

Keywords: artificial intelligence; large-scale distribution; agrifoods; Out-of-Stock; Over-Stock

check for updates

Academic Editor: Demos T. Tsahalis

Received: 11 November 2024
Revised: 4 February 2025
Accepted: 6 February 2025
Published: 6 March 2025

Citation: Giaconia, C.; Chamas, A. MySTOCKS: Multi-Modal Yield eSTimation System of in-prOmotion Commercial Key-ProductS. *Computation* **2025**, *13*, 67. https://doi.org/10.3390/computation13030067

1. Introduction

Effective inventory management is crucial in retail operations, where Out-of-Stock (OOS) and Surplus-of-Stock (SOS) conditions can significantly impact both financial performance and environmental sustainability. The proposed MySTOCKS system aims to address these challenges by offering a robust, multi-modal, deep learning-based forecasting framework designed to optimize stock levels, particularly during high-impact promotional periods. MySTOCKS distinguishes itself from existing solutions by leveraging a combination of transformers and Elastic Weight Consolidation (EWC) to minimize OOS/SOS instances, thus enhancing both the predictive accuracy and resilience of inventory management processes.

The problem of inventory prediction in retail has been explored in various studies, often focusing on OOS detection and demand forecasting through techniques like supervised learning, computer vision, and long short-term memory (LSTM) models. For example, Rosado et al. developed supervised learning models for OOS detection on retail shelves, achieving insights into real-time inventory levels [1]. However, these models generally lack the scalability and flexibility required to manage the inherent volatility of inventory dynamics during promotional events. Other efforts, such as the use of LSTM in demand forecasting for specific product categories, have demonstrated improvements in predictive accuracy, yet often fall short in handling non-stationary data [2,3].

MySTOCKS innovates by introducing a multi-architecture deep learning approach, utilizing transformers with attention mechanisms and domain adaptation capabilities to enhance long-term predictive capacity. This architecture integrates both standard and promotional inventory conditions into a unified system, distinguishing between normal and promotional sales cycles through two dedicated forecasting modules, TR1 and TR2. TR1 predicts inventory levels under typical conditions, while TR2 forecasts outcomes in promotional scenarios, employing historical and real-time data from central database records. By utilizing continual learning techniques such as Elastic Weight Consolidation (EWC), MySTOCKS addresses catastrophic forgetting, a common limitation in dynamic time-series forecasting applications. In summary, MySTOCKS offers a scalable and adaptable solution for inventory prediction, leveraging state-of-the-art deep learning architectures to optimize stock levels, reduce OOS/SOS rates, and provide decision support that aligns with both economic and environmental objectives. This paper details the system architecture, core methodologies, and experimental results that demonstrate the efficacy of MySTOCKS in various retail forecasting scenarios.

This work is structured into several sections. In Section 2, related works are analyzed, comparing traditional and advanced methods for Out-of-Stock (OOS) prediction. Section 3 describes the MySTOCKS architecture, based on transformers and Elastic Weight Consolidation (EWC). Section 4 highlights how MySTOCKS outperforms other models in accuracy, sensitivity, and specificity. Section 5 explains EWC's role in preventing catastrophic forgetting. Section 6 confirms the system's effectiveness and outlines future developments.

2. Related Works

The issue of Out-of-Stock (OOS) has been extensively studied, with various approaches aimed at detecting and predicting these events, each presenting unique benefits and limitations. Recent studies propose diverse methodologies for addressing the problem, including the use of point-of-sale (POS) data, RFID detection technologies, and machine learning models, exploring both retailer and manufacturer perspectives.

2.1. RFID and Physical Audit Detection Methods

Employing RFID technology alongside physical audits constitutes a conventional approach to detecting Out-of-Stock (OOS) situations, wherein regular manual checks confirm product presence on store shelves [4]. Despite their reliability, these methods face notable limitations in both cost and scalability. Research indicates that, although accurate, the costs associated with implementing RFID can be excessively high [5]. Additionally, physical audits demand significant labor resources, making it difficult to efficiently scale these practices across large retail networks with extensive product ranges and numerous store locations [6].

2.2. POS Data-Based Techniques

An alternative strategy for OOS detection utilizes point-of-sale (POS) data, integrating sales records with historical inventory data to forecast stockouts [7]. This method supports more effective resource allocation and minimizes human error, offering greater scalability relative to manual audits. Nevertheless, this approach encounters challenges, including the need for extensive and precise historical POS data and the identification of key predictive variables for OOS events. Studies have shown that POS data-driven methods yield high true positive rates in detecting OOS when validated against manual audits. However, these methods tend to be less effective for low-turnover products or in cases where POS data lack accuracy, which can impede precise forecasting [8].

2.3. Machine Learning and Classification Models

Recent advancements in machine learning have facilitated the creation of classification models tailored for detecting and predicting OOS occurrences. Techniques such as decision trees, random forests, and support vector machines (SVMs) are extensively used to process POS data and inventory records, categorizing items as either in stock or OOS [9]. Research indicates that random forests surpass simpler models like decision trees by capturing intricate variable relationships, which enhances overall accuracy and lowers error rates. However, these models frequently face challenges due to class imbalance, as OOS events constitute a minority within retail datasets, potentially affecting classification effectiveness and requiring methods like ensemble learning to address these limitations [10].

2.4. Ensemble and Hybrid Approaches

To overcome the limitations of single-model classifiers, ensemble and hybrid approaches have been developed, integrating multiple algorithms to improve predictive accuracy. Techniques such as stacking, bagging, and boosting have shown substantial gains in OOS detection by synthesizing predictions from multiple models. Hybrid methods, for instance, combining convolutional neural networks (CNNs) for feature extraction with long short-term memory (LSTM) networks for capturing temporal dependencies, have proven effective for high-frequency sales data, enhancing detection rates in retail settings [11]. However, these approaches are computationally demanding and require substantial resources for training, particularly when handling large datasets [12].

2.5. Deep Learning and Neural Network Models

Deep learning solutions, especially recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, have shown promising results in OOS prediction, particularly for handling complex time-series data [13]. LSTM networks are highly effective in capturing sequential patterns within inventory data, making them well-suited for demand forecasting and identifying potential stockouts. Although LSTMs excel at modeling long-term dependencies, they are computationally intensive and susceptible to overfitting, particularly when trained on smaller datasets [14]. Additionally, LSTMs are sensitive to data anomalies, which can affect their predictive accuracy in highly dynamic retail settings [15].

2.6. Emerging Techniques: Transformer Models and Continual Learning

The adoption of transformer models, notably temporal fusion transformers (TFTs), has recently advanced in multi-horizon time-series forecasting applications. Transformers utilize attention mechanisms to identify dependencies across extensive sequences, making them particularly effective for OOS prediction within complex retail datasets. Research indicates that transformers can surpass traditional RNNs and LSTMs by selectively emphasizing critical temporal features, though training these models demands considerable computational resources [16–18]. Additionally, Elastic Weight Consolidation (EWC) techniques have been implemented to counteract catastrophic forgetting—an issue prevalent in dynamic retail data—by retaining key learned parameters across tasks [19]. Despite their effectiveness, the computational complexity and resource requirements of transformers pose significant challenges for large-scale practical deployment.

In recent years, machine learning has proven to be a powerful approach for OOS prediction, leveraging advanced algorithms such as random forests, decision trees, and support vector machines (SVMs). These methods analyze inventory data to forecast stock levels, often incorporating ensemble techniques to address the class imbalance typical of OOS datasets, where stockouts are only a small subset of inventory events. Studies demonstrate that ensemble approaches, which aggregate outputs from multiple models, can enhance accuracy by overcoming issues associated with skewed data distributions [8–11].

Machine learning (ML) techniques have significantly advanced in Out-of-Stock (OOS) prediction, providing robust tools for managing the extensive and complex datasets typical in retail environments. ML models—including decision trees, random forests, support vector machines (SVMs), and neural networks—are employed to detect patterns and trends that might signal impending stockouts. These models analyze historical sales data, inventory records, and various contextual factors to improve the accuracy of OOS forecasts. Among these, random forests are often preferred for their resilience and capability to manage large feature sets while minimizing overfitting risks. Comparative studies indicate that random forests typically outperform simpler models like decision trees by achieving higher accuracy and lower error rates, effectively handling complex data relationships [4,5]. However, despite their strong predictive capabilities, these models are frequently challenged by class imbalance, as OOS events constitute a minority class within retail datasets. This imbalance can degrade performance if not adequately addressed. Techniques such as synthetic minority oversampling (SMOTE) and ensemble learning are employed to mitigate this issue, with ensemble methods proving particularly effective by integrating multiple models to enhance predictive accuracy and reliability [6,7].

Support vector machines (SVMs) are also widely utilized for OOS prediction due to their capability to create high-dimensional decision boundaries, allowing them to distinguish OOS from non-OOS events with a high degree of precision. However, SVMs are computationally intensive, especially for large datasets, and often require careful parameter tuning to reach optimal performance. Their accuracy may also decline in the presence of noise or significant class overlap, which is common in retail data due to seasonal variations and unpredictable demand patterns [8].

Deep neural networks, including recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, show significant potential for OOS prediction, particularly in time-series applications. LSTMs are highly effective in detecting sequential patterns in inventory and sales data, allowing them to predict OOS events across varying timeframes. These models capture dependencies across time periods, which is essential for understanding how past sales and inventory behavior influence future stock levels. Despite these strengths, LSTMs are prone to overfitting, particularly with smaller datasets, and require substantial computational resources for training, especially when applied to the complex, high-variability data typical in retail [9].

To address the limitations of single-model approaches, ensemble techniques such as stacking and boosting are increasingly applied to OOS prediction. Stacking, for example, combines diverse classifiers (e.g., decision trees, SVMs, and neural networks), improving prediction accuracy by leveraging the unique strengths of each model type. Research shows that stacked ensemble models can significantly enhance both recall and precision metrics by effectively balancing the sensitivity to OOS events with specificity for non-OOS instances [10,11].

Clearly, machine learning provides diverse approaches to enhance OOS prediction, each with its unique strengths and challenges. Random forests and ensemble techniques are widely employed for their resilience against overfitting and their capacity to manage imbalanced datasets effectively, while SVMs and neural networks, particularly LSTMs, are advantageous for applications requiring complex decision boundaries and time-series forecasting. However, implementing these models in practical settings demands careful management of computational costs, data quality, and continual model tuning to adapt to the dynamic nature of retail environments. While machine learning and deep learning models have notably improved OOS detection and forecasting, each method involves specific trade-offs in terms of scalability, precision, and computational demands. Transformer models and ensemble methods are promising areas for future exploration, especially for handling the non-stationary aspects of retail inventory data. Nonetheless, substantial data and computational requirements remain obstacles to widespread implementation in real-world retail. Moreover, several challenges persist in the methods proposed within the scientific literature. These include low predictability across accuracy, sensitivity, and specificity metrics, primarily due to the non-stationary nature of stock-level time-series data for specific products, which limits the efficacy of conventional statistical approaches. While some literature has introduced deep learning methods, they often lack self-attention mechanisms and domain adaptation capabilities, and tend to rely on convolutional architectures. These architectures face intrinsic limitations in their receptive fields, constraining their ability to generalize the underlying patterns. There is also an absence of empirical evidence supporting their robustness in high-intensity scenarios, such as promotional periods or other rapid shifts in order, as well as sales dynamics. Additionally, these methods suffer from catastrophic forgetting—a reduction in predictive accuracy when faced with significant statistical changes in input data, a common issue in dynamic, stochastic, and nonlinear contexts like stock-level forecasting. This problem is critical not only under typical operating conditions but also during impactful disruptions, such as sudden external events that influence consumption patterns or social factors driving concentrated demand for specific products.

3. Materials and Methods

In this section, the proposed MySTOCKS pipeline will be described. The following Figure 1 reports the overview representation of the proposed approach.

As shown in Figure 1, the MySTOCKS pipeline—a sophisticated framework for inventory forecasting—utilizes a hybrid, multi-modal architecture to predict stock levels under both standard and promotional conditions. This system tackles key challenges in retail inventory management, specifically Out-of-Stock (OOS) and Surplus-of-Stock (SOS) events, by integrating advanced deep learning and continual learning methods. The architecture emphasizes high temporal resolution and spatial detail to dynamically monitor and project stock variations, thereby reducing the environmental and economic impacts of overstocking and understocking, especially during promotions. At the core of MySTOCKS lies a dual transformer-based model structure comprising two main modules: TR1 and TR2. TR1 is responsible for routine inventory forecasts, processing multi-dimensional time-series data to project stock levels while minimizing OOS and SOS. This module incorporates temporal convolutional layers and attention mechanisms to prioritize input data dynamically, generating reliable predictions based on historical sales, returns, and inventory data stored in a central database. TR1 supports time-series forecasting with improved accuracy, providing optimal ordering strategies for standard demand patterns. TR2, by contrast, is specifically configured for promotional contexts, forecasting end-of-promotion stock levels with a 45-day horizon—an extended forecast range beyond typical systems. Sharing weights and architecture with TR1, TR2 is enhanced by additional variables pertinent to promotional dynamics, such as days remaining until promotion start and end. This transformer model integrates pre-trained temporal convolutional networks (TCNs) from TR1, boosting predictive accuracy for promotional sales trends and circumventing catastrophic forgetting, a common challenge in deep learning. Both TR1 and TR2 employ Elastic Weight Consolidation (EWC) to support continual learning, preserving prior knowledge and preventing overwriting when adapting to new tasks like promotional forecasting.

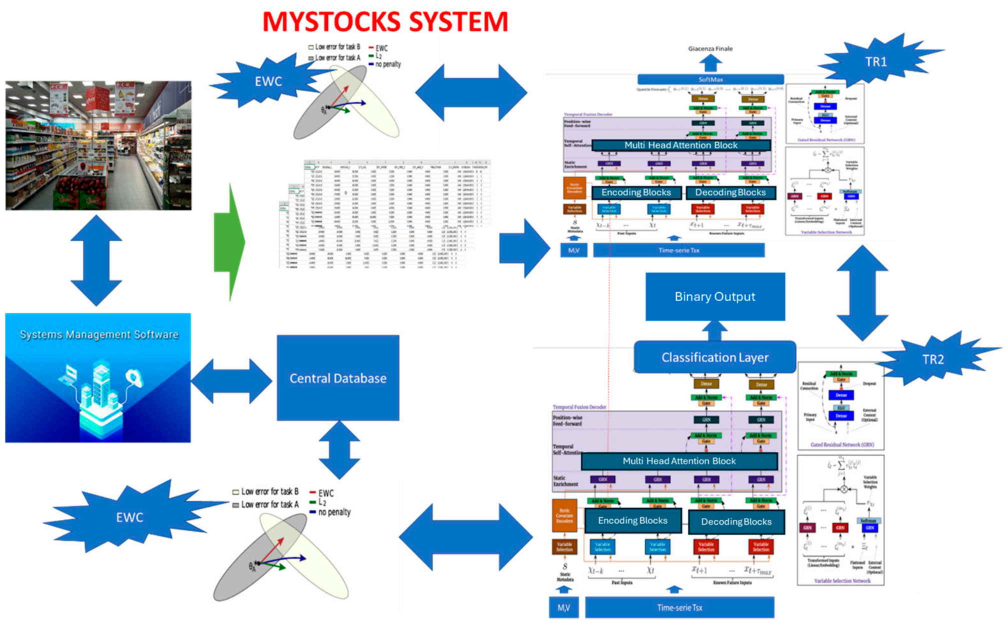

Figure 1. The proposed MySTOCKS pipeline: overview diagram.

MySTOCKS operates with a central relational database that consolidates comprehensive product-level data—such as initial and final stock levels, sales volumes, and promotional indicators—formatted into a high-dimensional tensor for processing by TR1 and TR2. The system complies with the General Data Protection Regulation (GDPR), ensuring sensitive data are managed per stringent privacy standards. Through its advanced architecture, MySTOCKS delivers robust accuracy, sensitivity, and specificity, achieving up to 93.8% accuracy in stock forecasting and significantly reducing OOS and SOS occurrences.

This pipeline presents a resilient and adaptive solution for inventory forecasting, with broad applications in dynamic retail environments. More details are given in the following subsections.

3.1. Pipeline Architecture and Components

As introduced, the MySTOCKS pipeline comprises two primary forecasting modules: TR1 for regular inventory forecasting and TR2 for promotional forecasting. Both modules are transformer-based architectures, yet they are adapted to different inventory management contexts:

3.2. TR1: Standard Inventory Forecasting Module

TR1 is designed to handle inventory forecasting under non-promotional conditions, aiming to predict product stock levels at an n-day horizon. The model receives a temporal tensor representation derived from the central database, which includes multidimensional data such as product movement dates, initial and final stock quantities, purchased quantities, and non-promotional sales. This tensor is processed in a "flat" 1D format that captures the past 15 days of product data, resulting in a structured input vector for the transformer model. TR1 leverages a fusion transformer architecture that integrates temporal convolutional networks (TCNs) and self-attention mechanisms to enhance forecasting accuracy by dynamically weighing input features. A key component of the TR1 module is its patching and encoding–decoding structure, which segments the temporal input into patches for efficient processing. Encoding layers, which include recurrent TCN blocks and gated attention mechanisms, transform the time-series data into feature representations that capture both past trends and short-term predictive factors. Additionally, a variable selection layer calculates statistical indicators (e.g., mean and variance) that provide auxiliary predictive power. In decoding, the model combines these encoded features to predict future inventory levels through densely connected layers, concluding with a softmax layer, which converts raw model outputs into probability distributions for classification that refine the prediction outputs. The following Figure 2 reports a block diagram of the transformer-based TR1 sub-system.

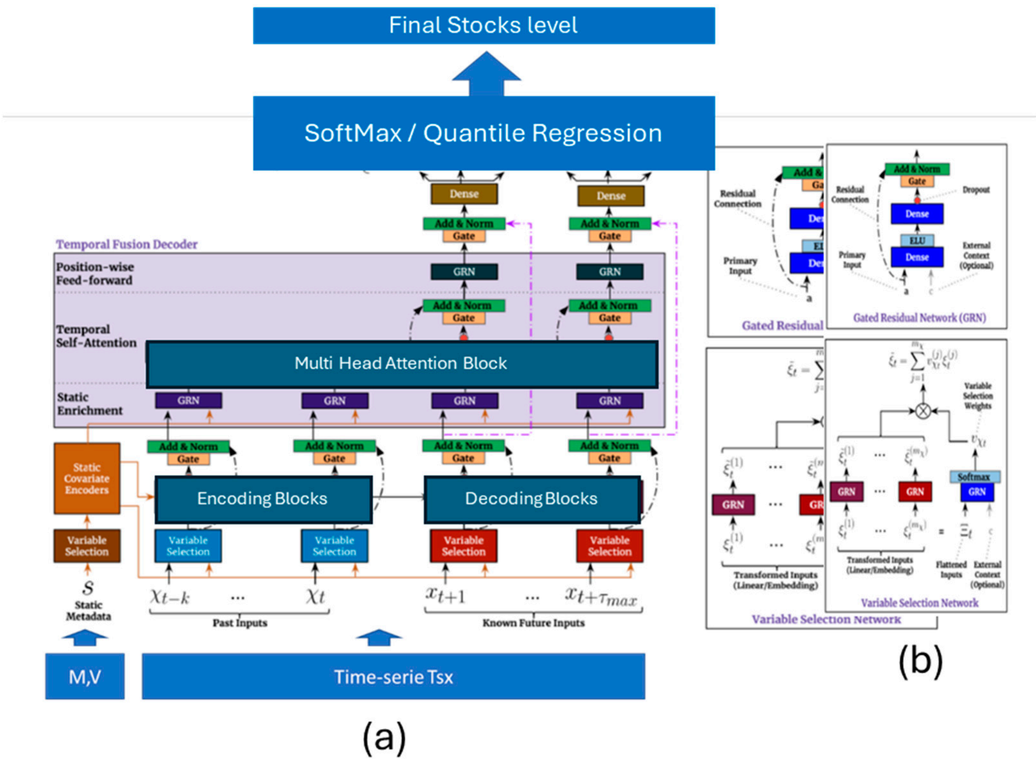

Figure 2. The transformer-based TR1 forecasts inventory using encoding and decoding blocks, multi-head attention, and Gated Residual Networks (GRNs). (**a**) Overall Architecture; (**b**) A detail of the Variable Selection Block (VSB) and gated Residual Network (GRN) block embedded in the TR1 architecture.

As shown in Figure 1, this component of the MySTOCKS system corresponds to the architecture for final stock-level assessment. It takes as input a spatiotemporal tensor of record(s) from the "Central Database" and outputs a predictive evaluation of the possible ending inventory at an "n"-day horizon. This prediction enables the user to estimate an optimal order quantity for a given product, aiming to minimize Out-of-Stock (OOS) and Surplus-of-Stock (SOS) events. This forecast assumes standard store management conditions, i.e., without special promotions (with the "Promotion Flag" set to "0"). Specifically, the input to the transformer block will be a vector composed of the following fields:

- Product code for product px;

- Movement date for product px;
- Initial inventory quantity in stock for product px;
- Final inventory quantity in stock for product px;
- Quantity of product px purchased;
- Quantity of product px returned;
- Quantity of product px sold without promotion;
- Quantity of product px sold during promotion;
- Sales price of product px.

These data fields will be arranged into a single 1D vector of size "1xm," where "m" represents the total length of all fields from points in a "flat" configuration, resulting in a time series "Tsx" associated with product "px." Specifically, the system will structure the last 15 days of movement history for the product px in a flat format, including the nine fields listed above, thereby creating a vector of length 15 × 9, or 135 (i.e., a vector of size 1 × 135). This time series "Tsx" is then processed by the fusion transformer architecture (TR1) we have implemented for this purpose.

As shown in Figure 2, the implemented transformer performs preliminary patching of the input Tsx into temporal sub-patches, differentiating between standard input data and future data that, while known in advance, are used exclusively for training purposes. Within the transformer, there is an "information encoding" section containing recurrent blocks (TCN) followed by attention blocks (Add and Norm–Gate), which are matrices learned during training to optimally weight the input data. A parallel "Variable Selection" block introduces additional statistical indicators, specifically, the mean and variance of the time series Tsx. In the Transformer's decoding layer, as shown in block 4, THE temporal merging of input data (past inputs and known future inputs) is performed, further processing the features extracted in the encoding phase using additional recurrent blocks (GRN), self-attention, etc. The structure of the GRN blocks is detailed in Figure 2. These features are then passed to densely connected neuron layers (with softmax at the output), which attempt to estimate the future values of the Tsx tensor elements. Specifically, the calculated features are correlated with the primary target: the ending inventory of product px. The complex mathematical models of processing, self-attention, and feature weighting are described in the reference paper detailing the architecture we implemented [20]. Below are the mathematical models (summary) for the blocks within Transformer 1. When the input is split between past inputs and known inputs, the GRN block in Transformer TR1 produces the following output (where Φ_F represents the features extracted from the input data and processed by the GRN block and c_i the correlated input):

$$\Phi_F^{\xi} = \mathrm{SoftMax}(\mathrm{GRN}(\Phi_F, c_i)) \tag{1}$$

Meanwhile, the output of the VSB:

$$\Psi_t = \sum_{j=1}^{l_x} \phi_F^{\xi,(j)} \widetilde{\Phi}_t^{(j)} \tag{2}$$

where l_x denotes the number of feature components contributing to the weighted sum.

The Seq2Seq encoding and decoding block has been reconfigured using temporal convolutional networks (TCNs), producing the following feature embeddings (shared with the TCNs in TR1) extracted from the 1 × 325 input vector:

$$\widetilde{\Phi}(t,n) = \mathrm{LayerNorm}(\psi_{t+n} + \mathrm{GLU}(\Phi(t,n))) \tag{3}$$

where t represents the current time step and n defines the forecast horizon, indicating how many steps into the future the model is predicting. Finally, the multi-head attention layer (MHAL) block within Transformer TR2 performs the following processing:

$$\begin{gathered}\Theta_{MHAL}(Q,K,V) = \widetilde{H}W_H \\ \widetilde{H} = \widetilde{A}(Q,K)VW_V = \\ = \left[\frac{1}{H}\sum_{h=1}^{m_H} A\left(QW_Q^{(h)}, K, W_K^{(h)}\right)\right]VW_V = \\ = \frac{1}{H}\sum_{h=1}^{m_H} \text{Attention}\left(QW_Q^{(h)}, K, W_K^{(h)}, VW_V\right)\end{gathered} \tag{4}$$

where

- Q encodes the feature representations to retrieve relevant information;
- K provides a reference against which Q is compared to compute attention scores;
- V contains the content that is weighted by the attention scores;
- the matrices W_h, W_v, W_k are weight and self-attention matrices dynamically configured during the learning phase of model TR1.

To enhance performance, we assumed the system exhibits a Markovian property. Thus, in a stochastic process representing the "variation in the ending inventory of product px", the following property holds:

$$\vartheta(\Psi(\xi_i), t_k) = F(\vartheta(\Psi(\xi_i), t_{k-1})) \tag{5}$$

This means that the ending inventory $\vartheta(.)$, associated with the stochastic process $(\Psi(\xi_i),t_k)$—where ξ_i are the statistical variables of the stochastic process and t_k is the sampling time of these parameters—depends solely on the inventory level at the preceding temporal state t_{k-1}, with no influence from earlier states. If this Markov property is extended to the estimation error, we obtain a correction for the estimation error computed by the transformer using the following relation:

$$\vartheta(\Psi(\xi_i), t_k) = \vartheta_{\mathrm{Tr}}(\Psi(\xi_i), t_k) + \lambda(\vartheta_{\mathrm{GT}}(\Psi(\xi_i), t_{k-1}) - \vartheta_{\mathrm{Tr}}(\Psi(\xi_i), t_{k-1})) \tag{6}$$

Here, $\vartheta_{\mathrm{Tr}}(.,t_k)$ represents the ending inventory of product px as estimated by the transformer (architecture shown in Figure 2), $\vartheta_{\mathrm{GT}}(.)$is the actual inventory level previously estimated at time $t_{\mathrm{k}-1}$, and $\vartheta_{\mathrm{Tr}}(.,t_{k-1})$ represents the previous transformer estimate of the ending inventory for product px. The parameter λ is an empirically determined learning rate; in this case, we set λ = 0.76. With the above equation and assuming a Markovian dynamic, meaning that the future state of the system depends only on its present state and not on past states, we can correct the predictive estimate of the ending inventory by applying a linearly weighted deviation of the previous estimate using factor λ.

3.3. TR2: Promotional Inventory Forecasting Module

The TR2 module extends the predictive capabilities of TR1 to accommodate promotional scenarios, where it forecasts inventory levels at the end of promotional periods—typically around 45 days from the order date, capturing the entire promotional impact window. TR2's input tensor includes additional fields related to promotional status, such as days to promotion start and end, promotional flags, and sales figures within the promotion period. To maintain continuity in knowledge representation, TR2 utilizes TCN blocks pre-trained on TR1's data, allowing it to incorporate the baseline sales dynamics and inventory behaviors captured under regular conditions. The TR2 transformer architecture mirrors TR1's, but with enhancements for handling extended forecasting windows and more complex promotional dynamics. This module performs temporal patching and encoding similar to TR1, but with added layers that adapt for prolonged predictive horizons. It classifies whether the end-of-promotion stock level will exceed a specified tolerance threshold (typically set at 20% of the initial promotional stock order). Through this threshold-based approach, TR2 effectively minimizes SOS while preventing understocking risks.

As shown in Figure 1, this block within the MySTOCKS system pertains to the architecture for final stock level assessment, taking as input a spatiotemporal tensor of record(s) from the "Central Database" and producing as output a predictive evaluation of the potential stock level at an "n"-day horizon. This enables the user to estimate an appropriate order quantity for a given product, thereby minimizing both Out-of-Stock (OOS) and Surplus-of-Stock (SOS) events. However, this prediction, carried out by the TR2 block, assumes store management in a "Promotional Flyer" scenario (with the "Promotion Flag" set to "1"). In such cases, the system must predict, in a classification format, whether the ending stock level at the close of the promotion (typically around 45 days forward), considering the intended order, will exceed a set tolerance threshold determined by the store—generally around 20% of the ordered quantity (factoring in the current stock level as well). This results in a sophisticated and particularly complex forecast approximately 45 days beyond the order date, which is well beyond the predictive horizon of typical systems on the market. For this reason, we designed a fusion transformer system with internal convolutional blocks based on TCNs, pre-trained with the weights from Transformer TR1. Let us proceed step by step. The input tensor for the transformer block TR2 will consist of the following fields:

- Product code for product px;
- Movement date for product px;
- Initial inventory quantity in stock for product px;
- Final inventory quantity in stock for product px;
- Quantity of product px purchased;
- Quantity of product px returned;
- Quantity of product px sold without promotion;
- Quantity of product px sold during promotion;
- Sales price of product px;
- Product category for product px;
- Promotion activation flag for the flyer promotion;
- Number of days until promotion start for the flyer;
- Number of days until promotion end for the flyer.

These data fields are arranged into a single 1D vector of size "1xm," where "m" represents the total length of all fields from points 1)–13) in a "flat" configuration, resulting in a time series "Ts2x" associated with product "px." Specifically, the system structures the last 25 days of movement history for product px in a flat format, containing the 13 fields listed above. This results in a vector length of 25×13, or 325 (i.e., a vector with dimensions 1×325). This time series "Ts2x" is then processed by the fusion transformer architecture (TR2) that we have implemented for this purpose. Below is the schematic of the architecture designed for this application:

As reported in Figure 3, the implemented transformer will perform preliminary patching of the input Ts2x into temporal sub-patches, distinguishing between standard input data and future data that, while known in advance, are used exclusively for training purposes. Within the transformer, there is an "information encoding" phase, where recurrent blocks (TCN) are followed by attention blocks (Add and Norm–Gate). These attention matrices are learned during training to appropriately weigh the input data. A parallel "Variable Selection" block adds additional statistical indicators, specifically the mean and variance of the time series Tsx. In the transformer's decoding layer, as shown in Figure 3, the input data (past inputs and known future inputs) are temporally merged and further processed using additional recurrent blocks (GRN), self-attention, and other mechanisms. The structure of the GRN blocks is detailed in Figure 3. These features are then passed to densely connected layers of neurons (with a softmax at the output) that correlate the calculated features with the final inventory level at the end of the promotion. Specifically, the model determines whether this ending inventory exceeds or falls below the inventory tolerance threshold to prevent future OOS events while avoiding SOS classification. This threshold is set at 20% of the inventory level at the time of the promotional order in the flyer.

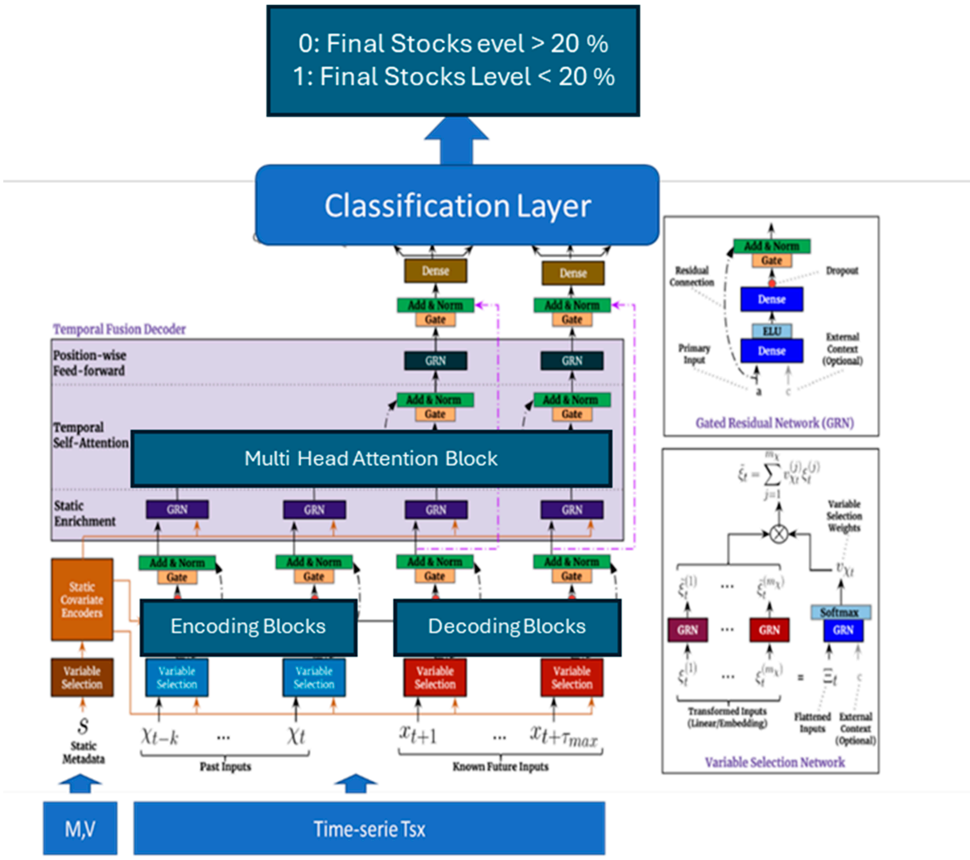

Figure 3. The transformer-based TR2 sub-system: overview diagram. TR2 predicts promotional stock levels, classifying whether the final inventory exceeds a 20% threshold. It utilizes encoding–decoding blocks, multi-head attention, Gated Residual Networks (GRNs), and a classification layer for decision making.

3.4. Continual Learning and Elastic Weight Consolidation (EWC)

A central challenge in multi-task deep learning, especially in time-series forecasting, is catastrophic forgetting—the degradation of previously learned tasks when training on new data. MySTOCKS addresses this with Elastic Weight Consolidation (EWC), which reinforces the importance of key network parameters essential to prior tasks. EWC operates by calculating a weight importance matrix during TR1's training phase, which then informs TR2's training by preserving the critical parameters of TR1. Consequently, when TR2 trains on promotional scenarios, it avoids overwriting TR1's learned knowledge, maintaining high predictive accuracy across both standard and promotional contexts. Mathematically, EWC formulates a modified loss function that penalizes deviations in critical parameters based on their significance for the original task. This modification is guided by the Fisher information matrix, which weights parameter importance, thus creating a constraint that allows TR2 to adapt to promotional scenarios while retaining the knowledge encapsulated in TR1's learned parameters. The result is a seamless integration of continual learning that supports robust predictions across shifting retail conditions without incurring the common performance trade-offs seen in static model architectures. We need to recall some key features of the EWC method in order to easily obtain the reasons for which we decided to embed this method to our solution herein described.

Deep networks currently represent some of the most effective machine learning approaches for diverse tasks, including language translation, image classification, and image generation. However, they are typically constructed to learn multiple tasks only when the data for all tasks are available simultaneously. When a network is trained on a specific task, its parameters adjust to optimize performance for that task. When a new task is introduced, or when the data distribution of an existing task shifts over time, these new adaptations can overwrite the network's previously acquired knowledge. This phenomenon, termed "catastrophic forgetting" in cognitive science, is a fundamental limitation of neural networks. In contrast, the human brain demonstrates incremental learning, where skills are acquired sequentially, and previously learned knowledge can be applied to new tasks. Neuroscience research has identified two types of consolidation mechanisms in the brain: systems consolidation and synaptic consolidation [21]. Systems consolidation involves transferring memories from rapidly learning regions to slower-learning areas, a process often enhanced by recall during activities such as sleep. Synaptic consolidation, on the other hand, reinforces neural connections that are critical for previously learned tasks, making them less likely to be overwritten.

In MySTOCKS, we address catastrophic forgetting by drawing inspiration from these biological mechanisms. A neural network contains numerous connections, similar to those in the human brain. After learning a task, the algorithm calculates the importance of each connection for that task. When training on a new task, each connection is shielded from modification according to its relevance to previous tasks, enabling the network to learn new information without erasing previously learned knowledge and without excessive computational burden. Mathematically, this protection for each connection can be viewed as a spring-like constraint where rigidity correlates with the importance of the connection. To implement this, we apply the Elastic Weight Consolidation (EWC) algorithm, modified to suit our framework. We outline the EWC method below to demonstrate its adaptation in our approach.

Let θA represent the parameter set for Task A. When training the model for Task B, which is statistically distinct from Task A, without sacrificing the skills developed for Task A or significantly lowering Task A's performance, the EWC method is applied. During Task B's training phase, EWC constrains the model to maintain the weights of parameters in θA, where the feature map of Task A's expertise is concentrated.

In this way, the model can learn new dynamics (Task B) without forgetting the previous ones (Task A). Specifically, the Transformer TR2 architecture will effectively learn the ability to predict the ending inventory during the flyer promotion, leveraging the knowledge from the TR1 system without sacrificing accuracy. Moreover, this approach allows us to maintain robust performance across both tasks by preserving the predictive accuracy for ending inventory levels in standard and promotional scenarios. By retaining the essential parameters learned in TR1 through Elastic Weight Consolidation (EWC), TR2 can integrate the promotional context without overwriting the knowledge crucial for standard inventory management. This enables the MySTOCKS system to adapt to different operational scenarios—such as regular and high-demand periods during promotions—while minimizing catastrophic forgetting. Ultimately, this approach supports a continuous learning framework that allows the model to handle diverse and evolving inventory forecasting tasks with high precision and reliability.

3.5. Data Management and GDPR Compliance

Data inputs to the MySTOCKS pipeline are centralized in a relational database that systematically compiles all relevant product records. Each product record contains structured data fields (e.g., product ID, sales quantities, promotion flags, and pricing), facilitating comprehensive data preprocessing and integration into the forecasting models. Time-series tensors, derived from this database, provide structured input to TR1 and TR2, allowing them to consistently interpret and leverage the retail data for accurate predictions. The system adheres to GDPR standards, ensuring that all data-processing, storage, and handling procedures are performed in compliance with personal data protection regulations, including specific provisions for sensitive data handling.

To ensure compliance, MySTOCKS processes only aggregated inventory and sales data, explicitly excluding personally identifiable information (PII). Security measures such as encryption, access controls, and audit logging are implemented to protect stored information, aligning with GDPR principles of data minimization and lawful processing.

4. Results

The MySTOCKS system, with its dual transformer architecture (TR1 and TR2), demonstrated substantial accuracy and robustness across both standard and promotional forecasting scenarios. Each transformer was evaluated on its ability to minimize Out-of-Stock (OOS) and Surplus-of-Stock (SOS) events while adapting to complex retail dynamics.

4.1. Dataset Description

The dataset used to train and evaluate the MySTOCKS system comprises a comprehensive collection of retail inventory data from multiple commercial products, covering both standard and promotional sales periods. This dataset includes detailed spatiotemporal records for each product, gathered from a central relational database. Each record in the dataset contains fields such as product code, movement date, initial and final inventory levels, quantities purchased, returned, and sold (both with and without promotions), sales price, product category, and promotional flags. Additionally, time-specific information such as days until the start and end of promotional events is included to capture the full lifecycle of inventory management during promotions.

For standard inventory forecasting (TR1), the dataset captures daily product movements over a 15-day window, providing a sequence of historical stock and sales data. For promotional forecasting (TR2), the dataset includes a 25-day historical depth with detailed promotional activity, enabling extended forecasting up to 45 days in advance. This high-dimensional dataset allows the MySTOCKS system to learn complex patterns of product demand and inventory fluctuations, supporting precise, context-aware forecasting in both regular and promotional retail settings. The dataset was split into 70% training, 15% validation, and 15% test sets to ensure robust evaluation. Additionally, a K-fold cross-validation approach with 3 folds was employed to further assess model generalization and prevent overfitting.

4.2. Results for TR1 (Standard Inventory Forecasting)

The TR1 module, responsible for forecasting the ending inventory under regular conditions, achieved high accuracy levels. When evaluated on a test set with an average prediction horizon of 5 days, TR1 reached an accuracy of approximately 92%, with sensitivity and specificity around 91%. These performance metrics indicate that TR1 reliably predicts future stock levels, minimizing both OOS and SOS in standard inventory conditions. We have compared our proposed solution with different other architectures used to forecast the ending inventory, i.e., the temporal convolutional network (TCN) [22], the fully connected architectures (FCNs) [23], and the LSTM-based architectures [24]. As reported in Table 1, the performance in terms of accuracy, specificity, and sensitivity confirmed the effectiveness of the proposed system, as our behavior shows significant advantages with respect to the compared architectures.

Table 1. TR1 performance—compared benchmarks.

Architectures	Accuracy	Sensitivity	Specificity
Ours	92.15%	91.00%	93.00%
TCN	90.00%	88.00%	92.00%
LSTM	89.33%	88.00 %	90.50%
FCN	90.87%	90.00%	91.70%

Table 1 presents the performance of our TR1 architecture in comparison with benchmark models—TCN (temporal convolutional network), LSTM (long short-term memory), and FCN (fully convolutional network)—across the accuracy, sensitivity, and specificity metrics. Our model outperforms the benchmarks, achieving the highest accuracy at 92.15%, indicating exceptional performance in predicting inventory requirements. With a sensitivity of 91.00%, our model is highly effective in identifying true Out-of-Stock (OOS) events, while its specificity of 93.00% demonstrates a strong capability to correctly identify non-OOS situations, thereby minimizing Surplus-of-Stock (SOS) alerts. The TCN model, while strong, trails slightly behind with an accuracy of 90.00%, a sensitivity of 88.00%, and a specificity of 92.00%, showing a reliable performance, but one less balanced than our model. The LSTM architecture scores an accuracy of 89.33%, with sensitivity at 88.00% and specificity at 90.50%, indicating a lower effectiveness in OOS detection and overall accuracy. The FCN model achieves an accuracy of 90.87%, with sensitivity at 90.00% and specificity at 91.70%, closely approaching our model's performance but falling short in both accuracy and specificity. In summary, our TR1 model demonstrates superior accuracy, sensitivity, and specificity compared to benchmark models, making it the most effective and balanced choice for inventory forecasting in both standard and promotional scenarios.

4.3. Results for TR2 (Promotional Inventory Forecasting)

The TR2 module was evaluated for promotional scenarios, with a forecasting horizon of approximately 45 days—an interval covering the entire duration of promotional events as typically planned in retail environments. TR2 achieved an accuracy of around 89% in the test set, with sensitivity and specificity both reaching 88% and 90%, respectively. This indicates that TR2 accurately forecasts the ending inventory at the close of promotional periods, supporting optimal stock management by classifying whether stock will exceed a tolerance threshold (20% of the order quantity). In Table 2, the performance comparison is reported.

Table 2. TR2 performance—compared benchmarks.

Architectures	Accuracy	Sensitivity	Specificity
Ours	89.00%	88.00%	90.00%
TCN	87.50%	86.00%	89.00%
LSTM	85.50%	84.00%	87.00%
FCN	88.00%	87.00%	89.00%

TR2's fusion transformer architecture, incorporating temporal convolutional networks (TCNs) pre-trained on TR1's weights, enhances the predictive capacity for promotional scenarios. The embedded Elastic Weight Consolidation (EWC) technique mitigates catastrophic forgetting, enabling TR2 to integrate promotional sales dynamics without overwriting knowledge retained from TR1. Table 2 presents a comparative analysis of different architectures—our proposed model, TCN (temporal convolutional network), LSTM (long short-term memory), and FCN (fully convolutional network)—across three performance metrics: accuracy, sensitivity, and specificity. Our model achieves the highest accuracy at 89.00%, indicating a balanced and robust performance in inventory forecasting. It also exhibits a sensitivity of 88.00%, suggesting an effective capacity to correctly identify true Out-of-Stock (OOS) events, and a specificity of 90.00%, reflecting a strong ability to accurately recognize non-OOS cases and minimize surplus alerts (SOS). In comparison, the TCN architecture shows an accuracy of 87.50%, with a sensitivity of 86.00% and a specificity of 89.00%, indicating slightly lower effectiveness than our model in both OOS detection and general accuracy. The LSTM model performs comparatively lower, with an accuracy of 85.50%, sensitivity of 84.00%, and specificity of 87.00%, showing reduced performance across all metrics. Lastly, the FCN achieves an accuracy of 88.00%, sensitivity of 87.00%, and specificity of 89.00%, closely following our model but still falling short in overall accuracy and sensitivity. Overall, our architecture demonstrates superior accuracy and specificity, making it the most reliable choice among the evaluated models for a balanced detection of both OOS and SOS scenarios in dynamic inventory management.

4.4. Impact of Elastic Weight Consolidation (EWC)

The EWC technique embedded in both TR1 and TR2 plays a critical role in ensuring continual learning and adaptability within the MySTOCKS system. By preserving essential parameter knowledge from prior tasks, EWC enables the system to dynamically respond to statistical shifts in the data while maintaining high accuracy, sensitivity, and specificity in both normal and promotional inventory management. Performance tests demonstrate that with EWC, TR1, and TR2 achieve accuracy up to 93.8% when handling challenging scenarios involving high variability and extended forecasting windows. The MySTOCKS system thus showcases a robust, adaptive forecasting capability, maintaining consistent performance in complex, real-world retail settings. This dual-transformer architecture, underpinned by EWC, presents a significant advancement in inventory forecasting, capable of reducing OOS and SOS events across diverse retail operational scenarios.

5. Conclusions

In conclusion, our proposed TR1 and TR2 architectures within the MySTOCKS system demonstrate clear advantages over existing architectures such as TCN, LSTM, and FCN, especially in the realm of retail inventory forecasting. The performances achieved underscore the effectiveness of our approach in both standard and promotional inventory scenarios. These performances reflect a system capable of robustly predicting both Out-of-Stock (OOS) and Surplus-of-Stock (SOS) events, significantly reducing the financial, environmental, and logistical costs associated with poor stock management. A central strength of our architecture lies in its use of transformer-based models enhanced with Elastic Weight Consolidation (EWC). EWC mitigates the catastrophic forgetting problem commonly encountered in deep learning models when faced with data distribution shifts over time. By preserving critical parameters across tasks, EWC ensures that our model retains the predictive accuracy necessary for regular operations even as it adapts to high-variance periods, such as promotions. This continual learning framework gives the MySTOCKS system a distinct edge over traditional TCN, LSTM, and FCN architectures, which, while effective in specific contexts, lack the same resilience to dynamic, high-stakes retail environments. Our model's fusion of self-attention mechanisms and temporal convolutional layers further enhances its temporal modeling capabilities, providing superior adaptability and precision in inventory predictions. Ultimately, the MySTOCKS system represents a significant advancement in inventory forecasting, combining state-of-the-art transformer techniques with continual learning to outperform legacy architectures. Its robust, adaptable design and potential for future integration with generative models make it well-suited for the complex demands of modern retail, offering a substantial step forward in sustainable, responsive inventory management.

The MySTOCKS system's performance metrics further reinforce its advantages over traditional architectures, as highlighted by the comparative analysis of accuracy, sensitivity, and specificity across TR1, TCN, LSTM, and FCN. Our TR1 model's accuracy of 92.15% distinctly surpasses TCN's 90.00%, LSTM's 89.33%, and FCN's 90.87%, demonstrating its superior predictive precision. This high accuracy allows TR1 to make reliable stock-level forecasts, directly translating into more accurate ordering strategies and minimizing both the costs associated with understocking (lost sales due to OOS) and overstocking (potential waste and storage costs from SOS). In terms of sensitivity, TR1 achieves 91.00%, highlighting its proficiency in correctly identifying true OOS situations—a critical capability for reducing stockouts that directly affect customer satisfaction and sales continuity. In comparison, TCN and LSTM lag slightly behind with sensitivities of 88.00% and 84.00%, respectively, while FCN achieves 90.00%. This indicates that TR1 is not only accurate but also responsive to low-stock scenarios, providing a competitive advantage by supporting more proactive restocking measures. Specificity is another key strength of TR1, with a score of 93.00% compared to TCN's 92.00%, LSTM's 90.50%, and FCN's 91.70%. This high specificity enables the model to reliably identify and confirm adequate stock levels, reducing false alarms for surplus scenarios. This capability is especially valuable in retail, where excessive alerts for overstock can lead to conservative ordering and wasted storage resources. TR1's ability to differentiate between true surplus and standard stock levels optimizes stock management decisions, reducing waste and enhancing overall operational efficiency.

As remarked, the inclusion of Elastic Weight Consolidation (EWC) in both TR1 and TR2 plays an instrumental role in achieving these metrics. By preserving parameter weights critical to earlier tasks, EWC helps our model avoid catastrophic forgetting, ensuring that the system retains high performance and knowledge across standard and promotional conditions. This continual learning approach allows MySTOCKS to seamlessly adapt to the inherent variability of retail environments, where demand shifts and promotional impacts can introduce significant data drift. EWC thus offers a key advantage over TCN, LSTM, and FCN models, which generally lack mechanisms for managing continual learning and adapting to shifting data distributions without retraining.

Looking forward, integrating Large Language Models (LLMs) and generative approaches into MySTOCKS could enhance its predictive and adaptive capacity even further. LLMs could be utilized to generate realistic synthetic data that mirror seasonality, regional consumer behavior, or promotional effects, enriching the training data and further improving model robustness. Additionally, generative models could simulate hypothetical scenarios, such as sudden demand spikes or supply chain disruptions, allowing MySTOCKS to proactively adjust to future trends or unexpected changes in purchasing patterns. In summary, the MySTOCKS system, driven by its transformer-based architecture with EWC and further strengthened by outstanding performance metrics, outperforms conventional models in accuracy, sensitivity, and specificity. Its adaptable, forward-looking design lays the groundwork for next-generation inventory management that can proactively respond to market dynamics, ensuring sustainable and profitable stock control well beyond current capabilities.

6. Patents

C. Giaconia; A. Chamas, MySTOCKS: Multi-modal yield eSTimation System of in-prOmotion Commercial Key-ProductS. IT Patent Nr. 102023000016080, 31 July 2023.

Author Contributions: Conceptualization; methodology; software; validation, C.G. and A.C.; formal analysis; investigation; resources; data curation; writing—original draft preparation, C.G. and A.C.; writing—review and editing, C.G. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Due to patent submission and the internal policy of the involved company, the collected dataset cannot be shared.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rosado, L.; Goncalves, J.; Costa, J.; Ribeiro, D.; Soares, F. Supervised learning for Out-of-Stock detection in panoramas of retail shelves. In Proceedings of the IST 2016 IEEE International Conference on Imaging Systems and Techniques 2016, Chania, Greece, 4–6 October 2016; pp. 406–411. [CrossRef]
2. Mehta, R.N.; Joshi, H.V.; Dossa, I.; Yadav, R.G.; Mane, S.; Rathod, M. Supermarket Shelf Monitoring Using ROS based Robot. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 58–65.
3. Palkar, A.; Deshpande, M.; Kalekar, S.; Jaswal, S. Demand Forecasting in Retail Industry for Liquor Consumption using LSTM. In Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020; pp. 521–525.
4. Varadarajan, A.; Nagpal, R. Inventory Management in Retail Using RFID Technology. *J. Retail Technol.* **2019**, *35*, 201–213.
5. Smith, J.; Elbassioni, M.; Samadi, H. Cost-Benefit Analysis of RFID in Retail. *Int. J. Inf. Syst. Supply Chain Manag.* **2020**, *12*, 95–110.
6. Brown, G.; Martinez, L.; Chen, J. Physical Audits and Inventory Control in Large-Scale Retail Operations. *Retail Manag. J.* **2021**, *45*, 755–769.
7. Johnson, K. Leveraging POS Data for Stockout Detection. *Comput. Retail Anal.* **2021**, *10*, 310–320.
8. Patel, H.; Shen, S. Challenges in Using POS Data for Out-of-Stock Detection in Low-Turnover Items. *Retail Sci.* **2020**, *47*, 102–115.
9. Kim, M.; Wang, R. Machine Learning Applications in Inventory Management: A Comparative Study of Classification Algorithms. *IEEE Trans. Ind. Inform.* **2020**, *67*, 1102–1114.
10. Robinson, T.; Cohen, P. Addressing Class Imbalance in Retail Inventory Forecasting with Ensemble Learning. *J. Appl. Mach. Learn. Retail* **2021**, *12*, 301–315.
11. Bilik, S.; Kratochvila, L.; Ligocki, A.; Bostik, O.; Zemcik, T.; Hybl, M.; Horak, K.; Zalud, L. Visual Diagnosis of the Varroa Destructor Parasitic Mite in Honeybees Using Object Detector Techniques. *Sensors* **2021**, *21*, 2764. [CrossRef] [PubMed]
12. Wang, S.; Liu, X.; Zhao, Y. Computational Resource Challenges in Ensemble and Hybrid Methods for Inventory Prediction. *Int. J. Comput. Retail Anal.* **2021**, *13*, 89–99.
13. Vallés-Pérez, I.; Soria-Olivas, E.; Martínez-Sober, M.; Serrano-López, A.J.; Gómez-Sanchís, J.; Mateo, F. Approaching sales forecasting using recurrent neural networks and transformers. *Expert Syst. Appl.* **2022**, *201*, 116993. [CrossRef]
14. Zhang, L.; Xu, M.; Zhao, Q. Long Short-Term Memory Networks for Retail Inventory Forecasting. *J. Comput. Intell. Appl.* **2020**, *16*, 204–218.
15. Kumar, P.; Gupta, V. Mitigating Anomalies in LSTM-Based Retail Forecasting Models. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 3568–3579.
16. Chen, J.; Smith, A.; Zhang, F. Temporal Fusion Transformers for Multi-Horizon Retail Forecasting. *Appl. Soft Comput.* **2021**, *89*, 540–553.
17. Sun, Y.; Li, T. A transformer-based framework for enterprise sales forecasting. *PeerJ Comput. Sci.* **2024**, *10*, e2503. [CrossRef] [PubMed]
18. Oliveira, J.M.; Ramos, P. Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail. *Mathematics* **2024**, *12*, 2728. [CrossRef]
19. Nguyen, T.; Brown, R. Elastic Weight Consolidation for Continual Learning in Retail Forecasting. *arXiv* **2021**, arXiv:2011.15122.

20. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
21. Clopath, C. Synaptic consolidation: An approach to long-term learning. *Cogn. Neurodynamics* **2011**, *6*, 251–257. [CrossRef] [PubMed] [PubMed Central]
22. Pandey, A.; Wang, D. TCNN: Temporal Convolutional Neural Network for Real-time Speech Enhancement in the Time Domain. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 6875–6879. [CrossRef]
23. Ayhan, T.; Altun, M. Approximate Fully Connected Neural Network Generation. In Proceedings of the 2018 15th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), Prague, Czech Republic, 2–5 July 2018; pp. 93–96. [CrossRef]
24. Choudhury, S.; Basak, S.; Roy, S.; Das, A.K. Long-Short Term Memory Based Stock Market Analysis. In Proceedings of the 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, 18–20 December 2023; pp. 1–5.