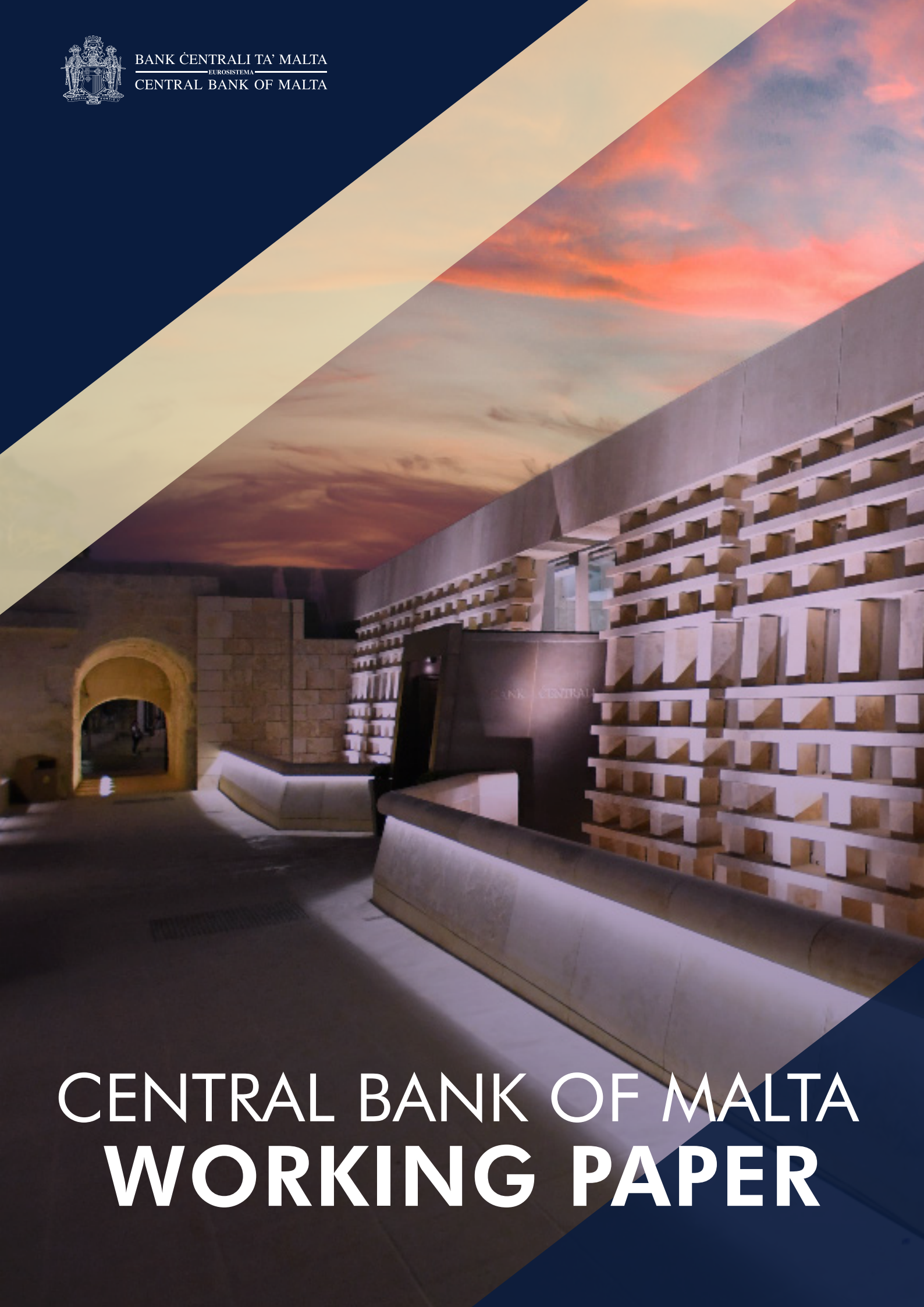


BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

CENTRAL BANK OF MALTA
WORKING PAPER



BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

Forecasting with AI: A Deep Learning Approach to Time Series Analysis of Malta's Foreign Trade[1]

Bilal Kurban[†], Roxanne Spiteri[†]

WP/10/2026

[1] The authors would like to thank the Central Bank of Malta for the opportunity to participate in the International Statistical Institute Regional Statistics Conference 2026. The authors are also grateful to Mr Jesmond Pulé and Ms Karen Caruana for their guidance and helpful feedback, as well as to the members of the Machine Learning and Big Data Network of the Central Bank of Malta for their valuable comments and suggestions. Special thanks are extended to Ms Zlatina Hofmeister, ECB Team Lead Economist Statistician and Chairperson of the ESCB Network of Machine Learning, for reviewing earlier versions of this paper and providing valuable feedback.
The views expressed in this paper are the authors' own and do not necessarily reflect the views of the Central Bank of Malta. Any queries should be sent through the following Contact Us form.

[†] Statistics Department, Central Bank of Malta

ISSN 3134-707X (online)

Abstract

This study examines the application of deep learning in forecasting Malta's Foreign Trade monthly statistics using historical time-series data. To do this, we built and tested several models based on recurrent neural networks by aiming to see whether they can increase the accuracy of short-term economic forecasts. Our work compares how each model performs and points out where these methods could fit into the analytical tools used by institutions that produce economic and financial statistics. The project examines both the accuracy and efficiency of different model designs and discusses what they could mean for improving the way economic indicators are produced.

Early findings are encouraging. Ongoing works continues to refine the models, evaluate robustness and explore their operational use in real statistical workflows. Overall, this research adds to the wider conversation about applying artificial intelligence in official statistics and central banking, with the goal of supporting more data-driven decision making.

Keywords: deep learning; artificial intelligence; time series; trade

Contents

Abstract **1**

1 Introduction **3**

2 Methodology **5**

2.1 Recurrent Neural Network Architectures 6

2.2 Experimental Setup and Evaluation Metrics 7

3 Results and Discussion **8**

4 Conclusion **11**

1. Introduction

Forecasting plays a crucial role in economic analysis within central banks, as precise projections guide decision-making and this is typically achieved through time series forecasting in which historical data has been used to predict future economic outcomes. At the Central Bank of Malta (CBM), monthly foreign trade statistics serve as an important measure of economic performance and international competitiveness. However, export series are often characterised by sudden shocks, structural shifts and non-linear patterns, which makes accurate forecasting a challenging exercise.

Forecasting has always been a key part of economic analysis, and trade statistics are no exception. For decades, economic forecasting has been largely driven by traditional methods such as ARIMA and Exponential Smoothing [1, 2]. These techniques continue to be widely used because they are well established, easy to interpret and generally perform well when underlying patterns remain relatively stable. As computing power has increased and larger datasets have become available, machine learning methods have attracted growing interest as an alternative way of modelling these more complex relationships [3].

Within the machine learning literature, recurrent neural networks (RNNs) have emerged as one of the main approaches for analysing sequential data. Unlike traditional forecasting models, RNNs are designed to retain information from previous observations, allowing them to learn patterns that unfold over time. In practice, however, standard RNNs often struggle when important information lies far back in the sequence. This limitation can reduce their ability to model longer-term dependencies. Long Short-Term Memory (LSTM) networks were developed as a solution to this problem. By using a system of memory cells and gates, LSTMs can decide which information should be retained and which can be discarded. This makes them particularly well suited to economic time series, where past events may continue to influence observations many months later. Over the last decade, LSTMs have become one of the most widely used deep learning architectures for forecasting applications.

A number of studies have reported encouraging results when applying LSTMs to trade-related data. Early evidence points to the potential of LSTM models in trade forecasting. The results reported in the literature have generally been promising. Shen et al. [4] reported encouraging results when applying LSTM networks to multinational trade forecasting. This demonstrates that they could achieve higher accuracy than several tra-

ditional approaches. Findings suggest that deep learning models are particularly useful for capturing the complex relationships. Building on this work, Yang et al. [5] found that combining economic indicators with deep learning techniques led to more accurate forecasts, especially in an increasingly interconnected global economy. Hopp [6] applied these models in an economic nowcasting context and demonstrated that they can effectively utilise high-frequency indicators to estimate current economic conditions. This is particularly relevant for central banks and statistical institutions, where timely information is often as important as forecast accuracy itself.

Despite these developments, evidence from small open economies remains relatively limited. Most studies have focused on large economies, global datasets, or broad macroeconomic indicators. Much less attention has been given to official trade statistics produced at the national level, especially in countries where economic activity is heavily influenced by external developments. As a result, there is still a need for empirical evidence on how deep learning models perform in smaller and more specialised economic environments.

2. Methodology

The data set used in this study was obtained from CBM Statistics [7] and comprises 272 monthly observations that span the period from January 2003 to August 2025. Five variables are included in the dataset, these being: Year, Month, Exports, Imports and Balance of Trade. The forecasting task focuses on monthly exports, which serve as a high-frequency economic indicator within official statistics.

An exploratory analysis was conducted to examine the statistical properties of the dataset. The mean monthly export value over the study period was €286.64 million. The export series also exhibited clear seasonality together with periods of increased volatility and a pronounced spike around 2012 as shown in Figure 2.1.

To facilitate time-series modelling, the Year and Month variables were combined to construct a chronological DateTime index, thereby preserving the temporal ordering of the observations throughout the forecasting process. With the new DateTime index, the entire export series could then be split in the training and test set. For this split, a chronological split of 80/20 was utilised. Moreover, since neural networks are sensitive to difference in the input feature scale, the `MinMaxScaler` was applied exclusively to the training set of the export series to normalise this to the range [0,1] and crucially, to prevent data leakage.

The univariate time series was subsequently transformed into a supervised learning problem using a sliding window approach. Rather than fixing the look-back period to a single sequence length, three sequence lengths of 6, 12, and 24 months were evaluated. These were selected to examine how the amount of historical information provided to the recurrent networks influences forecasting performance, while capturing temporal dependencies over short-, medium-, and longer-term horizons. For each sequence length, the preceding 6, 12, or 24 monthly export observations were used as input to predict the subsequent month's exports, as illustrated in Figure 2.2. Consequently, the effective number of observations available after sequence construction varied according to the selected look-back window, decreasing from the original 272 observations to 266, 260, and 248 observations for the 6-, 12-, and 24-month sequence lengths, respectively.



Figure 2.1: Plot of Exports, Imports and Balance of Trade

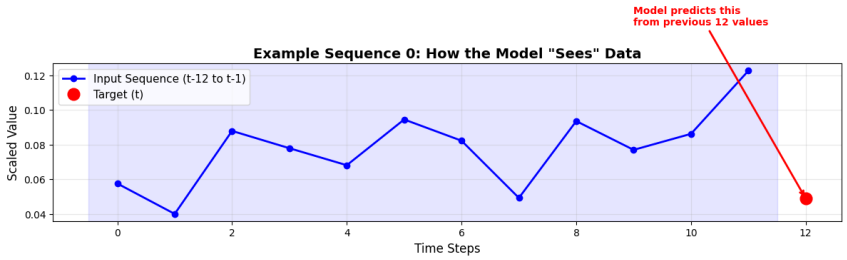


Figure 2.2: Illustration of the sliding window approach using a 12-month sequence length

2.1 Recurrent Neural Network Architectures

Three recurrent deep learning architectures were evaluated to assess their suitability for forecasting monthly exports: a Simple Recurrent Neural Network (RNN), a standard Long Short-Term Memory (LSTM) network, and an Advanced LSTM architecture.

The Simple RNN served as the baseline recurrent model. By maintaining a hidden state across successive time steps, the network is capable of modelling temporal dependencies within sequential data. However, conventional RNNs are susceptible to the vanishing gradient problem, limiting their ability to capture long-term relationships in extended sequences. To address this limitation, a standard LSTM architecture was implemented. LSTM networks incorporate memory cells together with forget, input, and output gates that regulate the flow of information through time, enabling the network to retain rel-

evant long-term information while discarding less useful historical observations. This gating mechanism has made LSTMs one of the most widely adopted architectures for sequential forecasting problems.

The final architecture consisted of an LSTM network comprising three LSTM layers with 50 hidden units each. Dropout layers with a rate of 0.2 were inserted between successive LSTM layers to reduce overfitting, while a fully connected dense layer produced the final export forecast. Increasing model depth allows higher-level temporal representations to be learned, although this is accompanied by increased computational complexity. The Simple RNN, standard LSTM, and Advanced LSTM models contained approximately 8,951, 31,901, and 54,651 trainable parameters, respectively.

2.2 Experimental Setup and Evaluation Metrics

To ensure reproducibility, all experiments were conducted using a fixed random seed. Model parameters were optimised using the Adam optimiser with a learning rate of 0.001, while the Mean Squared Error (MSE) loss function was employed during training. Each of the three recurrent architectures was independently trained and evaluated using sequence lengths of 6, 12, and 24 months, resulting in nine model configurations. All remaining experimental settings were kept consistent across configurations to facilitate a direct comparison of forecasting performance across both model architectures and sequence lengths.

Training efficiency and model generalisation were improved through the use of two callback mechanisms. As a regularisation technique, early stopping monitored validation loss and terminated training after fifteen consecutive epochs without improvement, thereby reducing the likelihood of overfitting. In addition, the ReduceLROnPlateau scheduler automatically decreased the learning rate whenever validation performance stagnated, allowing the optimiser to converge more effectively towards local minima.

Model performance was evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), two of the most widely used metrics for regression-based forecasting problems. Predictions were inverse-transformed prior to evaluation so that both metrics were reported in the original units of monthly exports, thereby facilitating meaningful interpretation and comparison of forecasting performance across the different model configurations.

3. Results and Discussion

The forecasting performance of the three recurrent neural network architectures was evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) on both the training and test datasets. To additionally examine the influence of the look-back window on forecasting performance, each architecture was evaluated using sequence lengths of 6, 12, and 24 months. Table 3.1 summarises the results obtained across the resulting nine model configurations.

The results demonstrate that forecasting performance was influenced by both the recurrent architecture and the selected sequence length. No single sequence length consistently produced the lowest test error across all three architectures. For the Simple RNN, the 6-month sequence length achieved the lowest test RMSE of 72.70, while increasing the look-back period to 12 and 24 months resulted in higher test errors. In contrast, both LSTM-based architectures performed best with a 12-month sequence length. The standard LSTM achieved its lowest test RMSE of 71.52 using the 12-month window, compared with 80.83 and 73.81 for the 6- and 24-month sequences, respectively. The Advanced LSTM exhibited a similar pattern, with the 12-month configuration producing a test RMSE of 67.15, compared with 77.87 for the 6-month window and 81.32 for the 24-month window.

Across all nine configurations, the Advanced LSTM with a 12-month sequence length achieved the lowest overall test RMSE of 67.15, while also recording a test MAE of 50.16. The stronger performance of both LSTM architectures at 12 months may reflect their ability to retain and selectively utilise relevant historical information, allowing them to benefit from temporal patterns occurring across an annual horizon. However, extending the look-back window to 24 months did not improve their forecasting performance, suggesting that additional historical information does not necessarily translate into greater predictive accuracy.

Overall, all three recurrent architectures were able to capture the general trend of Malta's monthly export series across the sequence lengths considered, although their out-of-sample performance varied. Based on the quantitative results presented in Table 3.1, the Advanced LSTM with a 12-month sequence length was identified as the best-performing configuration and was therefore selected for further analysis.

Model	Sequence Length	Training RMSE	Test RMSE	Training MAE	Test MAE
RNN	6	53.1699	72.7029	36.7405	53.3354
	12	54.8219	85.4476	36.9690	65.7535
	24	56.1616	81.2486	38.2445	62.1821
LSTM	6	61.1003	80.8325	44.2635	60.5035
	12	61.1450	71.5160	46.5617	52.0276
	24	60.1478	73.8135	43.4045	54.5625
Advanced LSTM	6	76.6490	77.8687	64.1575	57.2828
	12	57.5060	**67.1546**	39.5678	**50.1635**
	24	61.2567	81.3174	42.8019	61.0883

Table 3.1: Comparative performance of the three architectures considered across varying sequence lengths

Figure 3.1 presents the actual and predicted monthly exports generated by the selected Advanced LSTM configuration for both the training and test datasets. The model broadly followed the observed export series and captured its underlying temporal behaviour, although some of the more pronounced short-term fluctuations were not fully reproduced. Nevertheless, its ability to track the general movements of the series, provides further visual support for its comparatively strong out-of-sample performance.

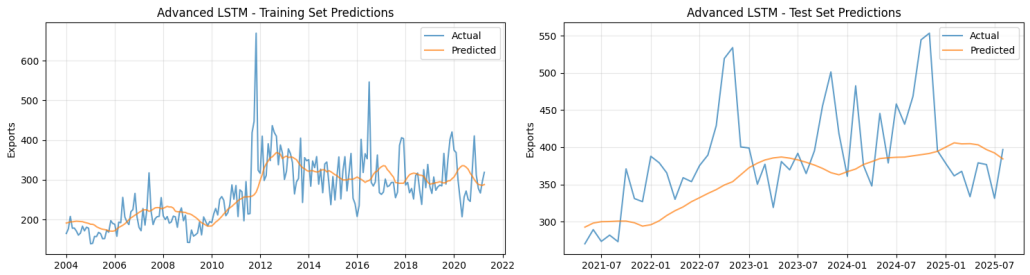


Figure 3.1: Actual and predicted monthly export values for the Advanced LSTM with a 12-month sequence length on the training and test datasets

The Advanced LSTM contained substantially more trainable parameters than the standard LSTM and Simple RNN, reflecting the greater computational complexity associated with its deeper architecture. Nevertheless, this additional complexity remained manageable for the dataset considered and was accompanied by stronger performance across the test set.

Based on the quantitative and visual evaluation, the Advanced LSTM with a 12-month sequence length was selected for the subsequent forecasting exercise. To further assess

the practical applicability of the proposed framework, the Advanced LSTM was used to generate forecasts beyond the original sample period. Figure 3.2 presents the resulting forecast trajectory together with the corresponding 95% confidence interval.

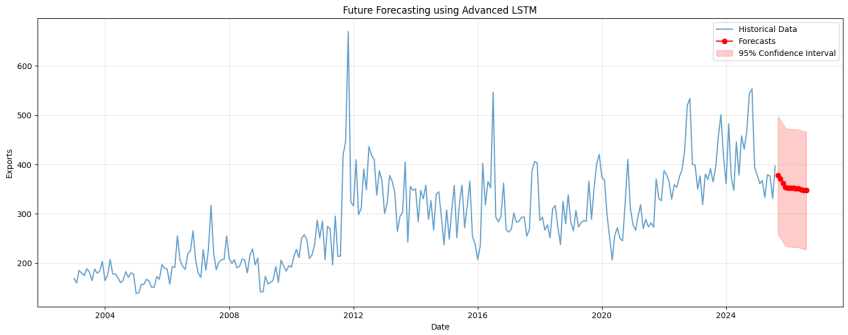


Figure 3.2: Out-of-sample export forecasts generated using the Advanced LSTM with a 12-month sliding window approach

Subsequent observations have since become available for part of the forecast horizon, allowing the predictive performance of the Advanced LSTM to be further assessed. All realised export values from August 2025 to April 2026 fall within the corresponding 95% confidence intervals of the forecasts, despite the considerable month-to-month volatility of the series. This result provides additional support for the robustness of the Advanced LSTM, as the forecast intervals adequately capture the uncertainty surrounding future export movements. While individual point forecasts may not reproduce every short-term fluctuation, the ability to encompass all subsequently observed values reinforces the model's practical applicability for short-term forecasting within official statistics.

4. Conclusion

This study investigated the use of recurrent deep learning models for forecasting Malta's monthly exports and considered their suitability for official statistics. The results demonstrate that forecasting performance was influenced by both model architecture and sequence length, with no single look-back window performing best across all three architectures. Among the nine configurations evaluated, the Advanced LSTM with a 12-month sequence length achieved the strongest overall performance, producing the lowest test RMSE. The results further indicate that increasing the amount of historical information supplied to the models does not necessarily improve forecasting accuracy, highlighting the importance of evaluating sequence length alongside model architecture.

The results indicate that recurrent deep learning can be a practical addition to the forecasting methods already used in official statistics rather than a replacement for them. Integrating these models into automated forecasting processes, together with reproducible training and regular model validation, could support more consistent and timely economic forecasts in central banks and other official statistical institutions.

References

[1] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control* (Wiley Series in Probability and Statistics). Wiley, 2016, ISBN: 9781118674925.

[2] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2021, https://otexts.com/fpp3/, ISBN: 9780987507136.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[4] M.-L. Shen, C.-F. Lee, H.-H. Liu, P.-Y. Chang, and C.-H. Yang, "Effective multinational trade forecasting using lstm recurrent neural network," *Expert Systems with Applications*, vol. 182, 2021, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2021.115199 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421006333

[5] C.-H. Yang, C.-F. Lee, and P.-Y. Chang, "Export- and import-based economic models for predicting global trade using deep learning," *Expert Systems with Applications*, vol. 218, p. 119 590, 2023, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2023.119590 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741742300091X

[6] D. Hopp, "Economic nowcasting with long short-term memory artificial neural networks (lstm)," *Journal of Official Statistics*, vol. 38, pp. 847–873, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232360845

[7] *Central bank of malta*. [Online]. Available: https://www.centralbankmalta.org/external-statistics