Evaluating The Predictive Performance Of LSTM Models On Chilean Pension Funds: A Comparative Study Using The S&P 500 Index

MSc Research Project
Master of Science in Data Analytics Information

Cristian Andrés Gajardo Montecino
Student ID: 24122246

School of Computing
National College of Ireland

Supervisor: Sallar Khan

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Cristian Andrés Gajardo Montecino
Student ID:	24122246
Programme:	Master of Science in Data Analytics Information
Year:	2024
Module:	MSc Research Project
Supervisor:	Sallar Khan
Submission Due Date:	11/08/2025
Project Title:	Evaluating The Predictive Performance Of LSTM Models On Chilean Pension Funds: A Comparative Study Using The S&P 500 Index
Word Count:	7.209
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Cristian Gajardo Montecino
Date:	15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Evaluating The Predictive Performance Of LSTM Models On Chilean Pension Funds: A Comparative Study Using The S&P 500 Index

Cristian Andrés Gajardo Montecino
24122246

Abstract

The expanding need for sustainable pension systems requires data-driven forecasting tools which help both financial planning and institutional decision-making. This research examines how three Long Short-Term Memory (LSTM) architectures—Multivariate LSTM, Attention-Based LSTM, and Multi-Input LSTM, perform in forecasting the daily quota values of "Fondo A" which is managed by AFP Habitat in Chile. The S&P 500 index served as an exogenous input to address traditional model limitations in volatile financial environments. The preprocessed historical time series data underwent standardized normalization followed by sliding window techniques before Python models with TensorFlow/Keras were trained for evaluation through RMSE, MAE and R^2 metrics. Model architecture proved to be a decisive factor in accuracy because the Attention-Based LSTM enhanced temporal feature extraction while the Multi-Input LSTM delivered the highest performance (RMSE: 909.62, R^2: 0.9820) by processing local and global signals independently. The research demonstrates that specific architectural improvements produce superior results than conventional methods which provide a flexible framework to enhance pension fund predictions in emerging market environments. The study faces limitations because it uses data from a single AFP and only one external indicator; future research should include more macroeconomic variables and expand datasets while investigating real-time deployment to maximize operational value.

1 Introduction

Worldwide pension systems face an urgent sustainability challenge because populations are aging and economic markets are becoming more volatile and connected to each other. The Mercer CFA Institute Global Pension Index Knox et al. (2024) ranks Chile among the top-performing pension systems in Latin America regarding overall performance but shows ongoing difficulties in maintaining long-term sustainability. The report demonstrates that combining strong investment plans with predictive risk evaluation methods is crucial for maintaining retirement income systems during demographic and economic challenges. According to Gubalová et al. (2025) the Sustainability Pension Sub-Index reveals that pension systems require transparent governance alongside diversified asset allocation and active management of long-term risks to stay sustainable.

The current worldwide pension fund management practices demonstrate an increasing requirement for improved forecasting methodologies. According to Trends in State and

Local Pension Funds (Giesecke and Rauh; 2023) pension funds have been transitioning their assets to alternative investments while expanding their global market presence and requiring innovative risk management solutions. The changing environment requires pension funds to move beyond market event responses by using predictive modeling systems that analyze various economic and financial signs.

The financial industry has experienced a substantial transformation in forecasting methodologies because of recent advances in machine learning (ML) and deep learning (DL) which provide valuable tools for extended investment planning. The research by Sonkavde et al. (2023) explains how ensemble models including Random Forest and XGBoost and LSTM produce better results than statistical methods when it comes to stock price predictions and market trend assessment and portfolio risk control. The modeling techniques used for short- and medium-term market movements also apply to long-term investment scenario modeling which pension fund management requires. Advanced models that combine sentiment analysis with technical indicators and macroeconomic variables produce forecasts which are stronger and more adaptable. Pension systems need this forecasting capability because accurate scenario modeling helps fund managers make strategic asset allocation decisions which enhances investment sustainability across multiple decades.

Pension fund investments in Chile serve both as a protection mechanism for individual retirement security and a foundation for national economic stability. These funds have such extensive size and influence that small enhancements in forecasting precision would produce major positive effects on both investment returns and risk reduction over time. This research aims to develop scenario modeling for pension fund investments in Chile to provide fund managers with better predictive tools for market trajectory assessment and asset allocation strategy development.

The research investigates whether advanced LSTM architectures that include attention mechanisms along with multi-input designs will boost the predictive accuracy of pension fund performance when analyzing global market indicators such as the S&P 500. Three LSTM-based models will be developed to answer this research question including the Multivariate LSTM and the Attention-Based LSTM and the Multi-Input LSTM.

This study makes three main contributions by using advanced deep learning models in pension fund forecasting for emerging markets and by integrating global financial indicators and assessing different architectural improvements for complex temporal dependencies. The research follows a particular structure which begins with LSTM applications for financial time series prediction then progresses to architectural developments including multi-scale processing and attention mechanisms and hybrid models. The analysis consists of examining the relative performance of LSTM against statistical models and their applications in emerging markets and institutional finance and their combination with other machine learning approaches. The research gap that this thesis addresses is established through the ordered sequence of literature review. The research design has a limited scope because it focuses on Fondo A of AFP Habitat in Chile and uses the S&P 500 index as the benchmark for the global market. However, the methodological framework can be extended to other pension funds and markets.

This report continues as follows: Section 2 discusses relevant research while Section 3 explains the research design. Section 4 explains the model architectures together with their design specifications. Section 5 explains the implementation process. Section 6 reveals the evaluation findings. The final section of this report (Section 7) presents research conclusions alongside recommendations for upcoming investigations.

1.1 Research Hypothesis

The research follows this hypothesis: The research predicts that advanced LSTM architectures with attention mechanisms or multi-input designs will generate more precise forecasts of Chilean pension fund values than a basic multivariate LSTM when using external macroeconomic indicators such as the S&P 500 index. The research evaluates this hypothesis through a comparison of predictive accuracy between three LSTM-based models using historical time series data from AFP Habitat's "Fondo A" while examining the effect of adding a global market variable.

2 Related Work

Financial time series forecasting relies heavily on Long Short-Term Memory (LSTM) networks because they excel at handling temporal dependencies and detecting nonlinear relationships.

Bahador (2018) tested the predictive capabilities of LSTM models on the S&P 500 index through a univariate framework. The research showed that forecasted values differed from actual backtesting results which demonstrated the weaknesses of using single-variable models in complex financial systems. The research established essential theoretical principles which future studies used to develop hybrid models that included external variables before the current study integrated the S&P 500 index into its predictive framework.

Fu et al. (2022) created an integrated forecasting framework that used ICEEMDAN and VMD decomposition techniques together with multi-factor analysis and an attention-based LSTM (ALSTM) network. The approach improved the LSTM's pattern recognition capabilities for financial data through its decomposition of original time series into more stable components. The proposed decomposition-attention architecture demonstrated superior predictive accuracy compared to Random Forest and GRU and standard LSTM models according to the research findings. The research proves how preprocessing combined with attention mechanisms improves LSTM performance which aligns with the present study's objective to investigate advanced architectural enhancements.

Teng et al. (2022) developed the Multi-Level Context-Aware LSTM (MLCA-LSTM) model to improve LSTM network predictive power through time scale integration with hierarchical attention mechanisms. The model uses local descriptors including Piecewise Aggregate Approximation (PAA), slope, and HOG-1D to detect short-term trends and reduce financial data noise effects. The experimental results showed that MLCA-LSTM outperformed traditional LSTM architectures in trend prediction tasks. The research findings are directly applicable to this study because they demonstrate how attention layers combined with multi-scale feature extraction can enhance forecast stability in unpredictable market conditions.

Zhang et al. (2024) proposed the 1D-CapsNet-LSTM hybrid model which unites the spatial feature extraction abilities of one-dimensional Capsule Networks (CapsNet) with the temporal modeling abilities of LSTM networks. The architecture was developed to overcome the spatial-temporal dependency limitations of traditional LSTMs in financial time series analysis. The model achieved superior performance than baseline approaches in its evaluation of the S&P 500 index and multiple global indices through Multiple-Input Multiple-Output (MIMO) strategies for modeling complex market relationships. The research results validate the effectiveness of hybrid deep learning architectures which match the current investigation into multi-input and attention-based LSTM models.

The research of Assaf et al. (2021) directed LSTM applications toward volatility forecasting instead of price prediction through stacked multivariate LSTM models. The authors used multiple asset prices together with longer lag periods to improve their model's ability to forecast high-frequency intra-day volatility. The proposed LSTM framework achieved better performance than traditional GARCH models in volatility prediction because it successfully detected intricate temporal patterns in volatile financial markets. The research demonstrates how LSTM architectures can handle various financial forecasting needs by showing their ability to model indicators that extend past basic price movements.

Kobiela et al. (2022) conducted a comparative analysis between ARIMA and LSTM models to predict NASDAQ stock prices for short- and medium-term periods. The research showed that ARIMA performed better than LSTM in multiple cases especially when forecasting 30 days ahead because it produced results that were 3.4 times more accurate. The research demonstrated that the choice of forecasting model depends on the dataset features and the length of the prediction period. The research findings support the current investigation by demonstrating the need for model selection based on context and showing instances where deep learning methods do not provide better results.

Research in recent times has concentrated on enhancing LSTM architectures together with the integration of other deep learning paradigms. Geras et al. (2015) developed "model blending" as a method to merge the inductive properties of LSTMs and CNNs. Research revealed CNNs and LSTMs learn distinct patterns and transferring LSTM knowledge to CNNs resulted in better accuracy and computational efficiency than using individual architectures. The research supports financial forecasting methods that utilize hybrid structures and multiple inputs because they require simultaneous detection of temporal patterns and feature-space relationships.

Traditional forecasting methods have received extensive comparison to LSTM models in various studies. Meghana and Shailaja (2024) performed a benchmarking study to evaluate the performance of LSTM networks relative to traditional forecasting models including ARIMA, Support Vector Machines (SVM), and Random Forest. The study evaluated how well the models detected nonlinear relationships in financial time series data. The results showed that LSTM performed better than traditional models when dealing with complex and nonlinear data patterns but the study did not include hybrid enhancements or macroeconomic indicators. The study's findings had restricted applicability to complex financial environments because of this exclusion. The research supports the current study by demonstrating that LSTM performs better than traditional models yet shows how adding external variables and architectural enhancements could produce better results.

Muncharaz (2020) tested the predictive capabilities of LSTM models against Exponential Smoothing (ETS) and ARIMA models using a dataset of 284 S&P 500 stocks. The results demonstrated that LSTM produced lower prediction errors than both traditional models when analyzing complex and noisy financial datasets. The research demonstrated how LSTM models excel at detecting complex time-based relationships which standard statistical approaches struggle to identify. The research supports the current study's implementation of LSTM architectures for financial prediction because market behavior in such contexts involves multiple interacting factors.

Nguyen and Ślepaczuk (2022) analyzed the effect of adding technical indicators to LSTM models for financial prediction tasks. The research analyzed daily returns by adding two technical indicators MACD and RSI as supplementary input features. The research showed that using technical indicators as additional inputs led to better forecast accuracy than using only price data. The research supports the current study by demonstrating how feature engineering improves LSTM performance which justifies the combination of worldwide market indices with domestic financial indicators.

Zhang et al. (2018) developed an LSTM model that combined with a multi-factor selection method to enhance financial time series prediction accuracy. The model achieved 90% predictive accuracy when applied to the CSI 300 index and produced substantial excess returns. The model demonstrated weak performance in predicting neutral returns because it excelled at detecting extreme market movements but struggled during periods of market stability. The research findings from this study match the current investigation because they demonstrate both the strengths and weaknesses of using factor selection with LSTM models which supports further investigation into multi-input structures and attention mechanisms.

The comparative analysis by Kijewski et al. (2024) used two decades of S&P 500 index data to compare the performance of classical forecasting strategies, such as ARIMA, momentum, and contrarian approaches, with recurrent neural networks, including LSTM. The study found that LSTM models outperformed traditional methods during periods of heightened market volatility, such as the 2008 financial crisis and the COVID-19 pandemic. Furthermore, the authors highlighted the benefits of integrating multiple predictive signals to improve accuracy and reduce portfolio risk. This work is relevant to the present study as it reinforces the suitability of LSTM models for volatile market conditions and supports the integration of diverse data sources to enhance predictive robustness.

Santos (n.d.) used recurrent neural networks with Gated Recurrent Unit (GRU) layers to predict age-specific mortality rates in Chile to enhance actuarial predictions for pension fund planning. The results showed that GRU-based models outperformed traditional demographic decomposition methods in terms of accuracy and stability in mortality forecasting. Although the focus was on demographic rather than financial time series, the study is relevant to the present research as it illustrates the value of recurrent architectures in institutional contexts such as pension fund management, where accurate long-term projections are essential for policy and investment decision-making.

Silvestre-Ponce et al. (n.d.) used a multi-output LSTM model to predict Chile's bond market yield curve. The method sought to reduce model complexity while enhancing the directional accuracy of forecasts for various maturities. The model generated better directional predictions than traditional benchmarks although it did not produce the highest returns which makes it suitable for strategic planning in developing markets. The research findings are relevant to this study because they show how LSTM architectures work in Chilean institutional finance and validate the use of multi-output and multi-input designs for complex financial forecasting problems.

Sebastian and Tantia (2025) used an attention-based multivariate LSTM model to predict Indian stock prices. The Diebold-Mariano statistical test was used to validate that the attention-based LSTM outperformed traditional forecasting models in terms of predictive accuracy. The attention layer allowed the model to filter out noise and focus on the most relevant time steps, resulting in more reliable predictions under volatile market conditions. This study is relevant to the present research as it highlights the benefits of integrating attention mechanisms into LSTM architectures, a core aspect of one of the models evaluated in this thesis.

Wekwete et al. (2023) investigated the application of Deep Reinforcement Learning (DRL) for Asset Liability Management (ALM) in institutional finance. Their framework dynamically optimized portfolio strategies by effectively detecting and responding to interest rate movements, outperforming traditional ALM models in both accuracy and adaptability. Although not based on LSTM, the study is relevant to the present research as it demonstrates how advanced machine learning techniques can enhance decision-making in institutional contexts, such as pension fund management, by adapting to changing market conditions in real time.

Jahan et al. (2025) used a Multi-Head Attention LSTM model to predict stock prices in the Bangladeshi market which is known for its volatility and limited data availability. The architecture used multiple attention heads to capture different aspects of temporal dependencies which improved the model's ability to extract relevant patterns from noisy datasets. Experimental results showed that the Multi-Head Attention LSTM outperformed GRU, Bi-LSTM, and standard LSTM models, delivering more accurate and robust forecasts. This study is relevant to the present research as it underscores the effectiveness of advanced attention mechanisms in enhancing predictive performance, particularly in emerging market contexts similar to Chile's pension fund system.

The financial prediction tasks benefit significantly from the implementation of attention mechanisms that enhance LSTM performance. Pardeshi et al. (2023) proposed the LSTM-SSAM model, which integrates a sequential self-attention mechanism into the LSTM architecture to enhance its ability to focus on the most relevant time steps in a sequence. Tested across various financial datasets, the LSTM-SSAM outperformed Bi-LSTM, CNN, ARIMA, and Prophet models in terms of predictive accuracy. The incorporation of self-attention allowed the model to dynamically adjust its focus, improving performance in environments with high volatility and complex temporal relationships. This study is relevant to the present research as it demonstrates how attention-based enhancements can significantly improve LSTM forecasting, aligning with the objectives of one of the models evaluated in this thesis.

Qiu et al. (2020) created an attention-based LSTM framework that utilized wavelet transforms to predict stock prices from major indices such as the S&P 500, DJIA and HSI. The wavelet transform processed historical stock data to remove noise while extracting vital features before the LSTM model received the data. The preprocessing technique improved the model's pattern recognition abilities which produced a coefficient of determination greater than 0.94 and a mean squared error less than 0.05. The research remains relevant to this study because it shows how combining signal processing techniques with attention mechanisms leads to better predictive results in unstable financial markets.

The Chinese stock market benefited from the attention-based Multi-Input LSTM model of Li et al. (2018) proposed an attention-based Multi-Input LSTM model to handle noisy auxiliary inputs in financial forecasting. The architecture used separate LSTM layers to process the primary and auxiliary input sequences while using an attention mechanism to select the most relevant features from each source. The model achieved better forecasting accuracy and robustness than standard LSTM and other baseline methods when applied to the Chinese stock market. The research is relevant to the present study because it shows how multi-input structures can be combined with attention mechanisms which is a design principle used in one of the models in this thesis.

Li et al. (2019) introduced the Evolutionary Attention-based LSTM (EA-LSTM) which combines an attention mechanism with competitive random search optimization inspired by evolutionary algorithms. The optimization strategy selects better time steps by escaping local minima and speeding up convergence than standard gradient-based training of attention layers. The EA-LSTM achieved better predictive accuracy on multivariate time series datasets through the integration of evolutionary computation methods with deep learning. The research supports the current investigation because it demonstrates that optimizing attention mechanisms leads to improved LSTM performance which matches the thesis objective to explore architectural enhancements.

The 2022 Entropy publication conducted by Nabipour et al. (2020) introduced a Multi-Input LSTM framework which processed various financial time series simultaneously before uniting them for prediction purposes. The LSTM branches handled individual input streams to maintain their temporal characteristics before the feature fusion process. The experimental results demonstrated that this architecture delivered better forecasting accuracy and robustness than single-input LSTM models when processing data streams with different statistical characteristics. The research findings from this study validate the implementation of multi-input architectures for uniting local pension fund data with global market indicators in one of the models developed in this thesis.

Hybrid approaches that combine LSTM with other learning paradigms further expand the modeling frontier. Zhou (2024) created a hybrid attention-based forecasting framework which combines LSTM networks with XGBoost and Support Vector Machines (SVM) to enhance stock price prediction. The architecture utilized LSTM for temporal dependency modeling and XGBoost for structured data handling and SVM for classification boundary robustness. The combined model achieved superior performance than all individual baselines especially during periods of high market volatility. The research findings support the current investigation because they demonstrate how combining deep learning models with other machine learning techniques produces effective financial market forecasting systems.

The research by Chou et al. (2020) presented an interval forecasting framework for financial time series through an Accelerated Particle Swarm-Optimized Multi-Output Machine Learning System. The method focused on generating forecast ranges instead of single point predictions because it aimed to handle financial market uncertainties. The system used simultaneous modeling of multiple outputs to exploit financial indicator correlations which resulted in better predictive accuracy. The accelerated particle swarm optimization algorithm outperformed traditional optimization methods by delivering faster convergence rates and superior solution quality. The research findings support the current investigation because they show how multi-output forecasting and metaheuristic optimization work together to enhance predictive accuracy through advanced modeling techniques.

Roszyk and Ślepaczuk (2024) developed a hybrid volatility forecasting model for the S&P 500 by merging GARCH with VIX and LSTM which produced better results than individual models and showed that combining statistical and sentiment with deep learning approaches delivers robust predictions. Research into financial forecasting now uses attention optimization techniques to blend hybrid modeling approaches with architectural blending methods which demonstrates a growing research trend for leveraging multiple algorithmic strengths.

No study has examined advanced LSTM variants including multi-input and attention mechanisms and evolutionary optimization in the Chilean pension fund system or integrated S&P 500 economic indicators to model local fund performance. This study evaluates and compares three LSTM-based architectures which use Multivariate LSTM and Attention-Based LSTM and Multi-Input LSTM to train on AFP Habitat's Fondo A data with S&P 500 index values to find the optimal design for improving predictive accuracy in institutional finance.

3 Methodology

The research employed an experimental approach to assess how three Long Short-Term Memory (LSTM) architectures namely Multivariate LSTM, Attention-Based LSTM and Multi-Input LSTM performed in predicting daily quota values of AFP Habitat's Fondo A in Chile. The methodology follows a structured pipeline, illustrated in Figure 1, which integrates data acquisition, preprocessing, model training, evaluation and output.

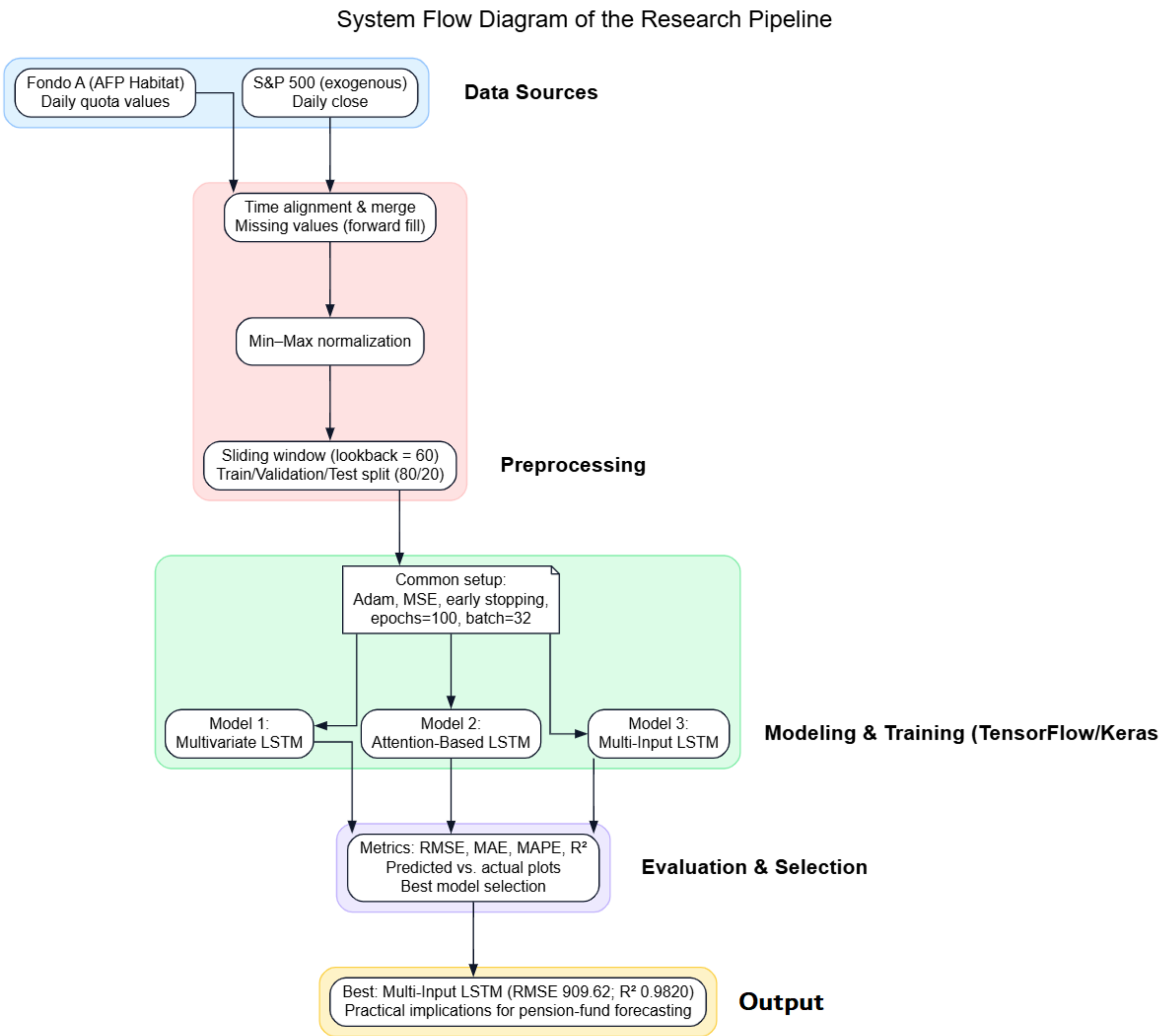


Figure 1: System Flow Diagram of the Research Pipeline

3.1 Data Collection and Preprocessing

The main dataset consisted of daily quota values of different Chilean pension fund administrators (AFPs) together with their corresponding values from the S&P 500 index. The consolidated data was stored in a CSV file named Cuota_value_&_SP500.csv. The experiments utilized quota values of AFP Habitat's "Fondo A" as well as S&P 500 index values which served as external inputs. The preprocessing pipeline included: The approach for dealing with missing values consisted of forward filling. Min-Max scaling normalization was used to prepare data for neural network training purposes. The preprocessing involved converting date indexes and organizing time series data for joint multi-input processing. Lag-based features were developed to incorporate temporal dependencies according to LSTM input specifications.

3.2 Model Architectures

The development of three different LSTM variants happened in Python using TensorFlow and Keras platforms. The Multivariate LSTM served as the baseline model since it took both the Chilean pension fund value and the S&P 500 index as multiple input variables. The model contained one hidden LSTM layer and a dense output layer which worked together to forecast the fund value for the following day. The attention mechanism enabled the multivariate model to become more accurate and interpretable through its implementation in the model. The attention mechanism assigned weights to time steps which improved both the interpretability and predictive power of the model. The Multi-Input LSTM architecture consisted of two separate LSTM branches that processed fund values and S&P 500 series before joining them into a single output. The model structure followed Li, Shen, and Zhu (2018) by permitting independent learning of exogenous information and gate control over external inputs.

3.3 Experimental Setup

All models were trained using an 80/20 train-test split. The data was structured into sequences by applying the sliding window technique which employed 60 previous time steps for predicting the next value. The training process employed Mean Squared Error (MSE) as the loss function together with the Adam optimizer. To ensure robustness and comparability: All models received identical hyperparameter settings which included epochs set to 100 and batch size set to 32. The model implemented early stopping to stop training when validation loss reached a certain threshold for preventing overfitting. Performance was evaluated using RMSE, MAE, and MAPE metrics. The analysis of predicted vs actual values employed matplotlib plots for visual inspection.

3.4 Tools and Environment

All experiments were implemented in Python 3.10 using: Pandas and NumPy for data manipulation. TensorFlow and Keras for model development. Matplotlib and Seaborn for visualization.

4 Design Specification

This section describes the architectures used for the LSTM forecasting models that were used in the assessment, along with the key techniques and frameworks used during the development of the models. The architectures were designed to handle time series data representing daily quota values of Fondo A Chilean pension fund and the macroeconomic signals from the S&P 500 index. It was also meant to test if the level of complexity and innovation in models improves the predictive performance when the data and preprocessing methods are kept the same.

4.1 Common Framework and Tools

All models were developed using Python programming language and Keras API on TensorFlow as the backend. The models were designed to take sequences of 60 historical observations and forecast the next-day quota value. Some of the key software libraries used include: TensorFlow/Keras for deep learning model construction. Pandas and NumPy for data handling and sequence transformation. Matplotlib for visual diagnostics of the outputs.

4.2 Model 1: Multivariate LSTM

The forecasting model used in this study employs a multivariate Long Short-Term Memory (LSTM) architecture which predicts the upcoming quota value of Fondo A (AFP Habitat) through two synchronized time series: (i) the daily quota value of Fondo A and (ii) the daily closing value of the S&P 500 index.
The architecture, illustrated in Figure 2, consists of the following layers:
First LSTM Layer: 64 memory units, configured with return_sequences=True to preserve temporal information across all time steps and facilitate deeper sequential modeling.
First Dropout Layer: Dropout rate of 0.2 to mitigate overfitting by randomly deactivating a fraction of the LSTM units during training.
Second LSTM Layer: 32 memory units for higher-level temporal feature extraction and refinement.
Second Dropout Layer: Dropout rate of 0.2, providing additional regularization.
Dense Output Layer: A single neuron with linear activation (default) to produce a continuous-valued prediction corresponding to the next-day quota value.
This configuration exploits the temporal dependencies between the two input features. The internal gating mechanisms of the LSTM layers enable the model to capture both short-term fluctuations and longer-term dependencies in the data. The use of multiple LSTM layers, combined with dropout regularization, aims to improve generalization while maintaining the ability to model complex, non-linear relationships between financial time series.

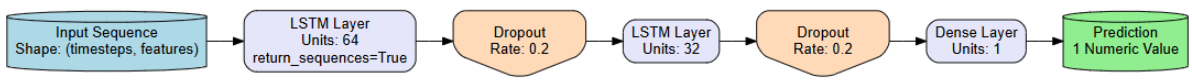


Figure 2: Multivariate LSTM Architecture

4.3 Model 2: Attention-Based LSTM

The proposed architecture builds upon Fu et al. (2022) by adding a custom soft attention mechanism to a baseline LSTM to focus on the most important time steps in the input sequence. The model starts with an LSTM layer that uses return_sequences=True to maintain the complete temporal information of the processed sequence. A custom attention layer applies a trainable linear transformation (Dense(1)) to each time step before computing softmax over the temporal dimension to obtain attention weights. The attention weights serve to calculate a weighted sum of LSTM outputs which produces a context vector that contains the most important historical patterns. The context vector moves directly to a fully connected (Dense) output layer to generate the forecast value. The attention mechanism automatically reduces the temporal dimension so an explicit flattening operation becomes unnecessary. The architectural design improves interpretability because it enables the model to focus more on crucial historical observations instead of treating all time lags with equal importance. The architecture diagram appears in Figure 3.

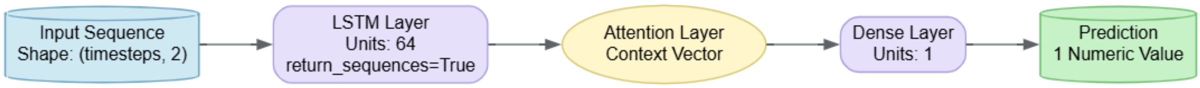


Figure 3: Attention-Based LSTM Architecture

4.4 Model 3: Multi-Input LSTM

The model employs two parallel LSTM branches to analyze fund quota values and S&P 500 index values independently. The model includes an input layer with (timesteps, 1) dimensions before an LSTM layer with 32 units that extracts temporal features from its respective series. The model combines the outputs of both branches before passing them to a Dense layer that generates a single prediction output. The model structure enables it to learn specific patterns from each variable type before uniting them to detect fund behavior and index effects. The model architecture lacks Dropout layers which simplifies its structure yet potentially increases its vulnerability to overfitting. The described architecture is illustrated in Figure 4.

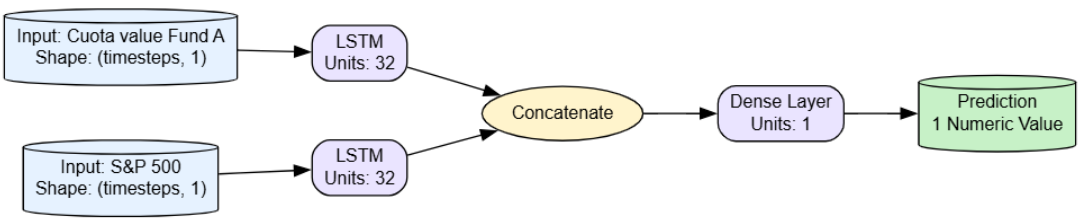


Figure 4: Multi-Input LSTM Architecture

4.5 Design Rationale

The move from a simple multivariate LSTM to attention-based and multi-input architectures was meant to check how the architectural improvements affect the forecast performance. Each model uses the same preprocessing pipeline and forecasting task, and thus any observed performance differences are due to the model design itself, and not the data handling.

5 Implementation

The final phase of implementation included the creation of three LSTM-based deep learning models for predicting daily quota values of the Chilean pension fund "Fondo A" managed by AFP Habitat. The entire work was carried out in Python using Jupyter Notebooks with TensorFlow and Keras as the primary frameworks for neural network development.

5.1 Development Environment and Tools

The implementation was done on a Windows 11 development machine with Intel i7 processor and 16GB RAM. The following software libraries and tools were used: Python 3.10: Main programming language. Jupyter Notebooks: Development and experimentation interface. TensorFlow/Keras: Deep learning framework for model construction and training. Pandas and NumPy: Data manipulation and time series transformation. Matplotlib and Seaborn: Visualization of predictions and model performance.

5.2 Data Preparation and Output Generation

The dataset used for all models was a merged time series of daily quota values for multiple Chilean AFPs and the S&P 500 index. The data was filtered to extract the values relevant to AFP Habitat's Fondo A and aligned with the corresponding S&P 500 values. Key outputs from this stage included: Cleaned and normalized dataset ready for LSTM input. Sliding window sequences for time series forecasting, with 60-day lookback periods. Input arrays reshaped for LSTM model requirements (samples $\times$ timesteps $\times$ features).

5.3 Model Implementation and Training

Three distinct models were implemented: Multivariate LSTM: A baseline model using both the fund value and the S&P 500 as inputs in a single sequence. Attention-Based LSTM: An advanced architecture with an attention layer applied over LSTM outputs to enhance temporal focus. Multi-Input LSTM: A dual-branch model where the fund data and S&P 500 series were processed separately before being merged. Each model was trained under consistent conditions using: 80/20 train-test split. 100 epochs with early stopping. Adam optimizer and Mean Squared Error loss function.

5.4 Results Produced

The final outputs of the implementation stage were: Trained models stored in memory for evaluation. Forecasted quota values compared against actual values in the test set. Performance metrics including RMSE, MAE, and MAPE for each model. Visualization plots showing predicted vs. actual values, error trends, and attention weight distributions (for the attention model). This implementation provided a complete pipeline from raw financial data to interpretable forecasts, which were then used in the evaluation stage to compare the predictive power and practical applicability of each model.

6 Evaluation

The evaluation section contains experimental evaluation outcomes from the three LSTM-based forecasting models developed in this research. The models received quantitative assessments and visual representation to check their predictive precision and reliability. The analysis compared the models through equal data portions and uniform preprocessing methods for maintaining experimental validity.

6.1 Experiment 1: Multivariate LSTM

The Multivariate LSTM served as the baseline model in this study. It utilized both the AFP Habitat fund value and the S&P 500 index as combined input features. The model achieved the following performance metrics: MAE = 753.05, RMSE = 945.02, and R^2 = 0.9805. It effectively captured the general patterns of fund value fluctuations, but it exhibited notable errors during periods of heightened market volatility. As illustrated in Figure 5, the model's predicted series aligns closely with the observed values under stable market conditions, but deviations become more pronounced during volatile intervals.

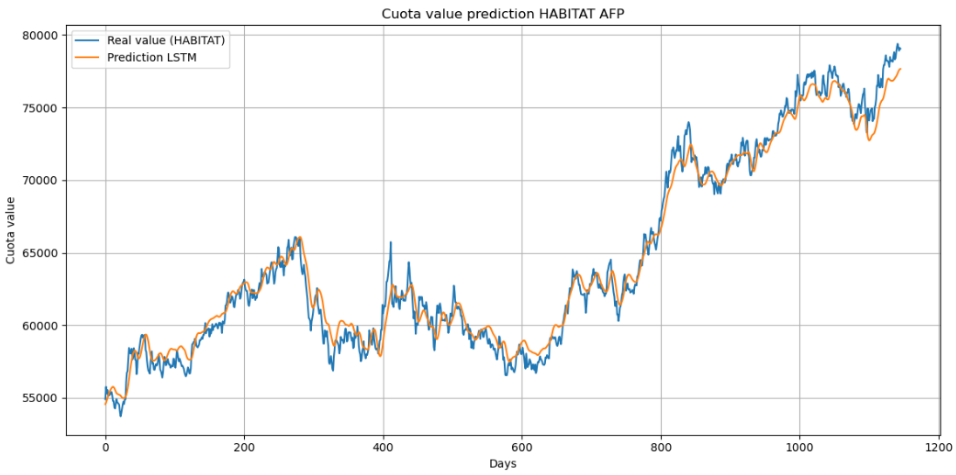


Figure 5: Multivariate LSTM Architecture

6.2 Experiment 2: Attention-Based LSTM

The Attention-Based LSTM model received an attention mechanism in its LSTM layer which enabled the model to focus on the most important time steps. The model achieved the following performance metrics: MAE = 1056.44, RMSE = 1409.25, and $R^2 = 0.9567$. The model outperformed the baseline model by reducing both error magnitude and variability. The attention mechanism proved effective in capturing essential temporal dependencies because the predicted series matches actual values throughout the entire time period as shown in Figure 6.

6.3 Experiment 3: Multi-Input LSTM

The Multi-Input LSTM processed the AFP fund value and the S&P 500 index as independent sequence data streams, which were subsequently merged after each LSTM layer's

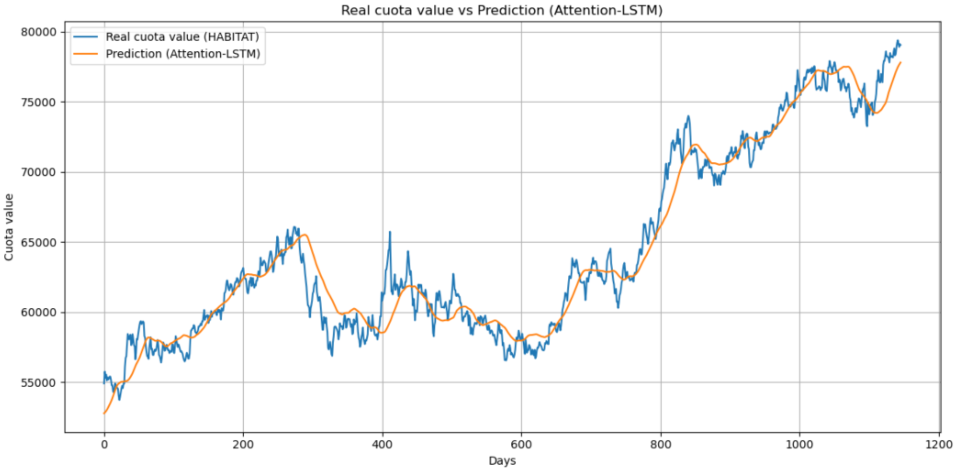


Figure 6: Attention-Based LSTM Architecture

processing stage. The model achieved the following performance metrics: MAE = 684.22, RMSE = 909.62, and R^2 = 0.9820. This architecture delivered the best overall results, showing its strongest performance in minimizing RMSE. Its design effectively separated time-dependent patterns from each input stream, thereby reducing signal interference. As illustrated in Figure 7, the predicted series closely follows the actual values, confirming the model's superior capacity to capture and integrate both local and global temporal dynamics.

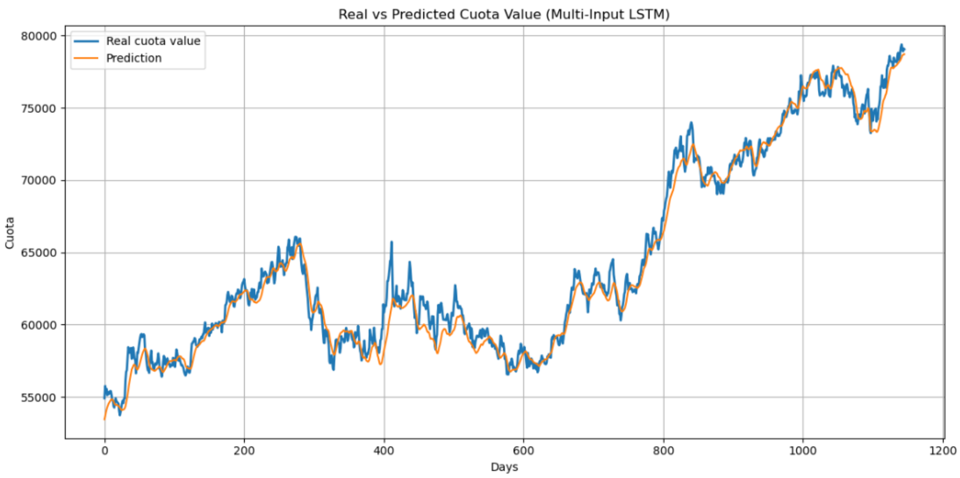


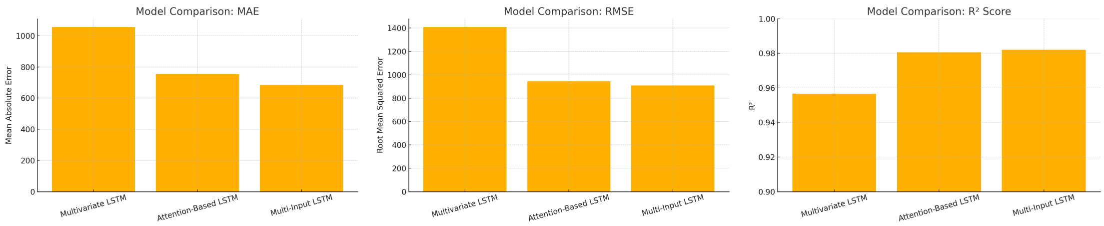
Figure 7: Multi-Input LSTM Architecture

6.4 Visual Results

Each model displayed line graphs that presented predicted quota values against actual quota values during the test duration. The Attention and Multi-Input LSTM models displayed better trend alignment with smaller error margins than the baseline model. The attention model displayed interpretable attention weights that indicated which time steps were essential for prediction purposes.

Table 1: Performance comparison of LSTM-based models

Model	**RMSE**	**MAE**	R^2
Multivariate LSTM	945.02	753.05	0.9805
Attention-Based LSTM	1409.25	1056.44	0.9567
Multi-Input LSTM	**909.62**	**684.22**	**0.9820**



(a) MAE (b) RMSE (c) R-square score

Figure 8: Results

6.5 Discussion

The evaluation shows that LSTM networks achieve improved forecasting performance through architectural enhancements that incorporate attention mechanisms and separate inputs. The study supports the findings of Pardeshi et al. (2023) and Li et al. (2018) which demonstrate how enhanced LSTM variants produce better results for financial time series prediction. The Multivariate model provided acceptable results yet failed to detect different patterns that emerged from separate input sources. The Attention-Based LSTM model improved temporal performance and the Multi-Input model optimized the processing of individual input streams. Potential improvements include: Experimenting with longer lookback windows. The model should incorporate external variables such as interest rates and inflation data. The combination of different models through ensemble methods to leverage their individual strengths. The research supports the idea that complex LSTM architectures boost pension fund prediction precision especially when using S&P 500 data as a global economic indicator.

The research findings support previous studies about deep learning systems used for financial prediction tasks. However, the research confirms that advanced recurrent neural architectures work well for pension fund forecasting despite previous studies mainly focusing on stock market and volatility prediction.

The Multi-Input LSTM outperformed other models and the S&P 500 index integration proved beneficial because Chilean pension funds react to international market conditions according to macroeconomic research on emerging economy capital flows.

The results demonstrate that diverse features together with complex architectures produce strong predictive models.

7 Conclusion and Future Work

7.1 Conclusion

This research investigated whether advanced LSTM-based architectures could enhance Chilean pension fund value predictions through the inclusion of external macroeconomic indicators such as the S&P 500 index. The research evaluated three architectures—Multivariate LSTM, Attention-Based LSTM, and Multi-Input LSTM—using daily quota values from AFP Habitat's "Fondo A."

The research successfully achieved its objectives. The research results showed that architectural enhancements produce quantifiable improvements in forecasting precision. The baseline Multivariate LSTM achieved acceptable results but its performance fell short of the more complex designs. The Attention-Based LSTM model improved its accuracy through temporal attention mechanisms which proved most effective during periods of market volatility. The Multi-Input LSTM achieved the highest performance (RMSE: 909.62; R^2: 0.9820) by processing local and global indicators independently, capturing the unique patterns of each data stream. The inclusion of the S&P 500 index resulted in better predictive outcomes which indicates Chilean pension funds show some level of response to international market trends.

The research expands deep learning applications into pension fund forecasting by demonstrating that attention and multi-input strategies outperform basic neural architectures according to empirical evidence. The research findings hold special value for emerging markets because they share similar data limitations.

7.2 Future Work

Future research directions show potential in several areas. The input space should be expanded by adding local economic indicators which include interest rates and inflation as well as GDP. The use of ensemble learning techniques which combine XGBoost with LSTM models should be investigated for enhanced generalization performance. The implementation of Explainable AI (XAI) tools will enhance the ability to explain how input variables affect the system. Multiple AFPs need to be tested for evaluating how performance varies between institutions. The implementation of these models in production environments through dashboard or API systems would enable real-time pension fund monitoring capabilities. Deep learning models prove to be effective tools for financial time series forecasting when properly designed according to research findings. This forecasting method enables pension fund managers together with policymakers to make better data-driven decisions in markets that are becoming more unpredictable.

The research provides practical solutions to the Chilean pension fund industry by enhancing financial forecasting accuracy for data-based decision-making.

The research demonstrates how pension fund managers can use advanced LSTM-based architectures to integrate local and global indicators for market prediction and investment optimization.

The research methodology shows potential for application in other emerging markets with comparable data limitations to enhance institutional risk management and policy development.

7.3 Limitations

Several limitations should be acknowledged. The results from this study cannot be generalized to other institutions because the dataset contained only daily quota values from Habitat AFP. The study used only the S&P 500 index as its external macroeconomic indicator which might have excluded other important economic factors such as interest rates and inflation and GDP. The evaluation period was restricted to historical data and no validation with future out-of-sample data was conducted which may affect the robustness of conclusions. The attention mechanisms provided better interpretability than standard LSTMs but the models still lack full explainability of their decision-making processes.

References

Assaf, O., Di Fatta, G. and Nicosia, G. (2021). Multivariate lstm for stock market volatility prediction, *International Conference on Machine Learning, Optimization, and Data Science*, Springer, pp. 531–544.

Bahador, M. (2018). The accuracy of the lstm model for predicting the s&p 500 index and the difference between prediction and backtesting.

Chou, J.-S., Truong, D.-N. and Le, T.-L. (2020). Interval forecasting of financial time series by accelerated particle swarm-optimized multi-output machine learning system, *Ieee Access* **8**: 14798–14808.

Fu, E., Zhang, Y., Yang, F. and Wang, S. (2022). Temporal self-attention-based conv-lstm network for multivariate time series prediction, *Neurocomputing* **501**: 162–173.

Geras, K. J., Mohamed, A.-r., Caruana, R., Urban, G., Wang, S., Aslan, O., Philipose, M., Richardson, M. and Sutton, C. (2015). Blending lstms into cnns, *arXiv preprint arXiv:1511.06433* .

Giesecke, O. and Rauh, J. (2023). Trends in state and local pension funds, *Annual Review of Financial Economics* **15**(1): 221–238.

Gubalová, J., Medveďová, P. and Špirková, J. (2025). The sustainability pension subindex in the context of the indicators weights., *Statistika: Statistics & Economy Journal* **105**(1).

Jahan, M. K., Chowdhury, M. E., Shuba, S. J., Khan, F. S., Rahman, M. M., Faruk, M. O., Chowdhury, M. A. A., Islam, S. J. and Arman, M. (2025). Multi-head attention based lstm model for forecasting bangladeshi stock market prices, *2025 IEEE Conference on Computer Applications (ICCA)*, IEEE, pp. 1–7.

Kijewski, M., Ślepaczuk, R. and Wysocki, M. (2024). Predicting prices of s&p 500 index using classical methods and recurrent neural networks.

Knox, D., Mercer, Institute, C. and for Financial Studies, M. C. (2024). Mercer cfa institute global pension index 2024: Helping dc members get the best retirement outcome. 16th edition. Lead author: Dr. David Knox AM, Senior Partner, Mercer.
URL: *https://www.mercer.com/what-we-do/wealth-and-investments/global-pension-index.html*

Kobiela, D., Krefta, D., Król, W. and Weichbroth, P. (2022). Arima vs lstm on nasdaq stock exchange data, *Procedia Computer Science* **207**: 3836–3845.

Li, H., Shen, Y. and Zhu, Y. (2018). Stock price prediction using attention-based multi-input lstm, *Asian conference on machine learning*, PMLR, pp. 454–469.

Li, Y., Zhu, Z., Kong, D., Han, H. and Zhao, Y. (2019). Ea-lstm: Evolutionary attention-based lstm for time series prediction, *Knowledge-Based Systems* **181**: 104785.

Meghana, R. and Shailaja, K. (2024). Predicting stock prices using lstm networks: A comprehensive approach, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* **12**(VII): 1046–1057.
URL: *https://doi.org/10.22214/ijraset.2024.63696*

Muncharaz, J. O. (2020). Comparing classic time series models and the lstm recurrent neural network: An application to s&p 500 stocks, *Finance, Markets and Valuation* **6**(2): 137–148.

Nabipour, M., Nayyeri, P., Jabani, H. and Mosavi, A. (2020). Deep learning for stock market prediction.

Nguyen, T. T. G. and Ślepaczuk, R. (2022). *The efficiency of various types of input layers of LSTM model in investment strategies on S&P500 index*, Vol. 29, University of Warsaw, Faculty of Economic Sciences.

Pardeshi, K., Gill, S. S. and Abdelmoniem, A. M. (2023). Stock market price prediction: A hybrid lstm and sequential self-attention-based approach, *Applications of AI for Interdisciplinary Research*, CRC Press, pp. 122–140.

Qiu, J., Wang, B. and Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism, *PloS one* **15**(1): e0227222.

Roszyk, N. and Ślepaczuk, R. (2024). The hybrid forecast of s&p 500 volatility ensembled from vix, garch and lstm models, *arXiv preprint arXiv:2407.16780* .

Santos, J. M. B. V. (n.d.). Backtesting recurrent neural networks with gated recurrent unit: Probing with chilean mortality data.

Sebastian, A. and Tantia, V. (2025). Multi-variate lstm with attention mechanism for the indian stock market, *International Journal of Information Management Data Insights* **5**(2): 100350.

Silvestre-Ponce, C., Diaz, D. and Astudillo, C. (n.d.). Modeling and predicting the yield curve of the chilean bond market through multi-output lstm, *Available at SSRN 4643746* .

Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D. and Bhat, S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications, *International Journal of Financial Studies* **11**(3): 94.

Teng, X., Zhang, X. and Luo, Z. (2022). Multi-scale local cues and hierarchical attention-based lstm for stock price trend prediction, *Neurocomputing* **505**: 92–100.

Wekwete, T. A., Kufakunesu, R. and van Zyl, G. (2023). Application of deep reinforcement learning in asset liability management, *Intelligent Systems with Applications* **20**: 200286.

Zhang, C., Sjarif, N. N. A. and Ibrahim, R. (2024). 1d-capsnet-lstm: A deep learning-based model for multi-step stock index forecasting, *Journal of King Saud University-Computer and Information Sciences* **36**(2): 101959.

Zhang, R., Huang, C., Zhang, W. and Chen, S. (2018). Multi factor stock selection model based on lstm, *International Journal of Economics and Finance* **10**(8): 1–36.

Zhou, J. (2024). Predicting stock price by using attention-based hybrid lstm model, *Asian Journal of Basic Science & Research* **6**(2): 145–158.