Deep Learning Approach Based on Bi-LSTM for Handling Missing Data in the Stock Market

Ala Alrawajfi, Mohd Tahir Ismail, Sadam Al Wadi, Saleh Atiewi, Abdelrahman Hamad

Abstract— **This study presents a new Attention-Bidirectional Long Short-Term Memory (Attention-BiLSTM) model specifically developed to tackle the problem of imputation for missing data in stock market datasets, emphasizing the Amman Stock Market. The suggested integrates a dynamic attention mechanism that adapts according to the temporal context and the pattern of missing data. This adjustment improves the imputation accuracy by selectively directing attention toward the most pertinent information. This study systematically compares the performance of the Attention-BiLSTM model with conventional imputation techniques such as Random Forest Regression, Support Vector Regression (SVR), LSTM, and BiLSTM. The evaluation metrics, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE), consistently show that the proposed model surpasses these baseline approaches, especially when dealing with larger amounts of missing data. The present work also examines the computational complexity and scalability of the novel approach, together with the assumptions behind the characteristics of the missing data. The findings demonstrate that the Attention-BiLSTM model is a pragmatic and effective approach for real-time stock market applications, resulting in notable enhancements in imputation accuracy and model resilience under various situations.**

Index Terms— **Attention-BiLSTM, Deep Learning, Financial Data Analysis, Machine Learning, Missing Data Imputation, Stock Market Data**

I. Introduction

THE stock exchange, commonly referred to as the stock market, stands out from other markets due to its lack of physical items and commodities in most cases. The stock market operates under a framework of legislative and technical regulations that oversee various aspects of its operations, including the selection of specific investments, the determination of optimal selling points, and the risk of losing financial market data due to failures in data transmission units to the central database [1].

The occurrence of missing data in stock market records is a common phenomenon caused by various factors, including exchange market power outages, disruptions in Internet connectivity, and errors in recording observations. Since stock market time series data encompasses both temporal and geographical characteristics, it is crucial to exercise caution when reconstructing missing periods to preserve the statistical properties of the series without compromising their integrity [2] [3].

One frequently used strategy for addressing missing data is to exclude incomplete observations, hence focusing the analysis solely on complete data [4]. However, this approach can lead to fragmentation, resulting in the loss of valuable insights and potential distortion of findings. To address this challenge, various data estimation approaches have been proposed and extensively discussed in contemporary scholarly discourse. These methods range from traditional statistical methods, such as imputing missing values with the mean, median, or other appropriate values, to more advanced computational techniques [5],[6].

A solution is to use conventional data imputation processes to replace data sets with missing values with calculated estimates. Specifically, data imputation can fill in non-response missing values by providing a data set that accounts for potential error. Conventional imputation approaches include methods such as mean/mode imputation, hot-deck imputation, and regression models[7].

Multiple imputation (MI) is a statistical approach commonly used to handle missing data. This approach provides practical solutions for dealing with incomplete data. MI procedures replace missing values with a set of plausible values that reflect the uncertainty in value substitution, rather than replacing them with a single value. These multiple imputed data sets are then analyzed using standard techniques, and the results of these analyses are combined. The process of synthesizing inferences from multiple data sets remains largely the same, regardless of the complete-data analysis performed [8].

Researchers typically choose one of two primary methods to address the issue of missing data: deletion or imputation, as illustrated in Fig. 1. The deletion technique involves removing cases with missing data, whereas the imputation method estimates or imputes the missing values. The imputation method can be categorized into three primary strategies. The first method, statistical imputation, is beyond the scope of this study. The second method involves intelligent imputation techniques. The third method employs

Manuscript received September 21, 2024; revised March 29, 2025.

Ala Alrawajfi is a postgraduate student at School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Minden, Penang, Malaysia (corresponding author to provide phone: +962-795160000, fax: +962-775160000 email: alaalrawajfi@student.usm.my)

Mohd Tahir is an associate professor of Mathematical Sciences, Universiti Sains Malaysia, 11800, Minden, Penang, Malaysia (email: m.tahir@usm.my)

Sadam Al Wadi is an associate professor of Collage of Business, the University of Jordan, Amman, Jordan (email: sadam_alwadi@yahoo.co.uk)

Saleh Atiewi is an associate professor of Department of Computer Science, Al Hussein Bin Talal University, Ma'an, Jordan (email: Saleh@ahu.du.jo)

Abdelrahman Hamad is an undergraduate student at Department of Computer Science, Al Hussein Bin Talal University, Ma'an, Jordan (email: abdelrahman.hamad2003@gmail.com)

hybrid imputation techniques. This study will focus primarily on intelligent imputation techniques, particularly examining fuzzy learning and machine learning approaches. The primary objective is to evaluate the efficacy of the proposed deep learning (DL)-based approach against other machine learning-based methods [9].

The subsequent sections of this paper are summarized as follows: Section 2 provides an overview of the literature review. Section 3 introduces the research objectives, section 4 presents the dataset, section 5 outlines the research methodology, section 6 discusses missing data types, section 7 introduces missing imputation, section 8 presents the data preprocessing technique, and the discussion and results in section 9, conclusion, and section 10 anticipates future work.

II. Literature Review

Numerous recent studies have examined approaches for identifying missing values in financial data or financial markets by using a range of methodologies that include conventional statistical techniques or intelligent approaches. The primary focus of the conducted study has been on the detection and imputation of missing values. This task is carried out to facilitate the processing of data for subsequent value predictions. However, studies dedicated to the exploration of novel methodologies that aim to compensate for missing values with little error and maximal accuracy are scarce.

Table I displays a portion of the research undertaken to identify missing data in financial markets from 2021 to 2023.

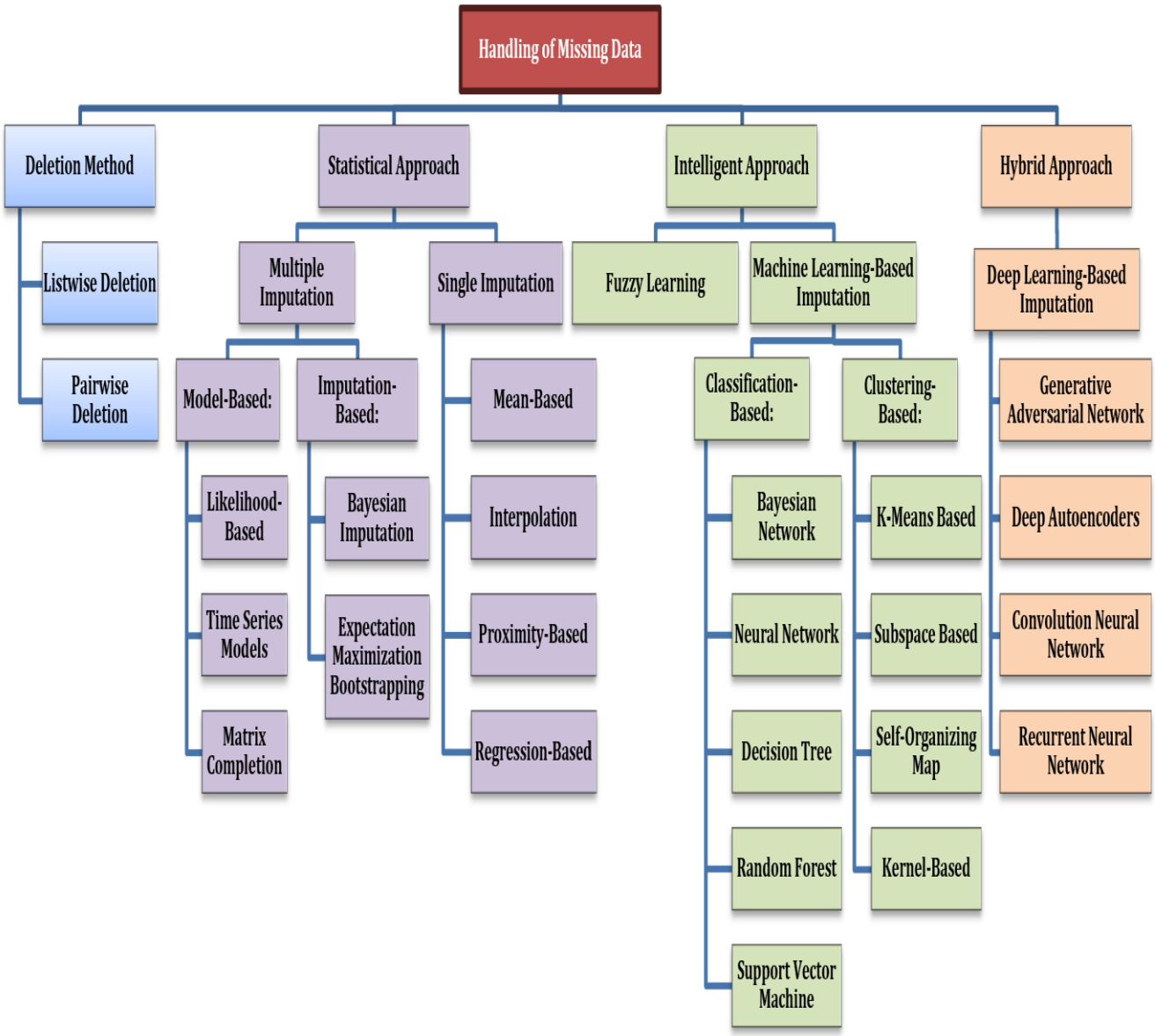

Fig. 1. Methods of Handling Missing Data "redrawing based on [9]"

TABLE I
RECENT RESEARCH IN MISSING DATA IMPUTATION IN FINANCIAL DATA

Year	Study	Objective	Methodology	Findings
2023	[10]	To conduct a comparative analysis and quantitative assessment of several imputation techniques	K-NN, expectation–maximization, classification, and regression tree, and RF	The findings suggest that the expectation-maximization technique exhibits favorable performance relative to other methods.
2021	[11]	To assess the effectiveness of modeling stock market volatility using RF	RF	The random forest (RF) algorithm exhibits the benefit of achieving high performance in scenarios with large feature sizes and effectively managing intricate datasets.
2021	[12]	To build up a prediction model for forecasting stock price.	LSTM, Extreme Gradient Boosting, Linear Regression, Moving Average, and Last Value model	The experimental findings confirm that these models can acquire patterns in time series data. The Long Short-term Memory (LSTM model demonstrated superior performance for the intended objective.
2021	[13]	To evaluate and compare the effects of imputation methods for estimating missing values in a time series.	Mean, LOCF, MICE, K-NN, and NOCB	The results show that the K-NN technique is the most effective in reconstructing missing data and positively contributes to time series forecasting compared with other imputation methods.
2022	[14]	To propose a novel imputation method to obtain a fully observed panel of firm fundamentals that exploits time-series and cross-sectional dependency of data to impute their missing values	Statistical Methods, Backward-Forward-XS, Backward-XS, Forward-XS, Cross-sectional, Time-series, Previous value, and Cross-sectional median	Extensive research has been conducted on the documenting of significant consequences for risk premia calculations, cross-sectional anomalies, and portfolio building.
2021	[15]	New approaches for impute missing time steps are compared to traditional and state-of-the-art methods. The first imputation technique is Decomposition. The second imputation approach is Imputation by Nature.	LSTM, Kalman Filter, Spline Interpolation: Mean, K-NN, Regression, and Decomposition	The Financial Index data set resampled by the weeks' mean yielded the best results when imputed using Kalman, Nature, and Decomposition in sequence. The financial Index data resampled by Friday yielded the best results when imputed using Kalman, Regression, Missing Indicators, and Mean tied for third place. The data sets' findings are imputed.
2021	[16]	To propose a Modified Genetic Algorithm (MGA) in which the local optima problem of Genetic Algorithm (GA) is addressed	K-NN (KNN), mean, median, or zero/constant	The performance of MGA is compared with those of GA and PSO. For evaluation, accuracy, precision, recall, and F-score are used. The MGA is better, according to the findings.
2022	[17]	To provide the unique Rolling Mice Forest (RMF) approach for missing values in financial panel data. RMF imputes missing data using RFs, reducing future observation leakage	list-wise, deletion, cross-sectional mean/median imputation, and K-NN imputation, Rolling Mice Forest (RMF)	Gradient Boosted Regression Trees GBRTs outperformed OLS models in stock return prediction. In the U.S. and Norway, the median and RMF imputation improve the GBRT stock return predictions best. By contrast, list-wise missing data deletion lowers the GBRT stock return prediction performance.
2023	[18]	To propose a novel framework that leverages generative adversarial networks (GANs) and an iterative approach using the gradient of the complimentary	Generative adversarial networks	The experimental findings demonstrate that imputeGAN exhibits superior performance compared with the conventional complementation techniques in terms of complementation accuracy.

Table I illustrates that the use of the intelligent methodology for identifying missing data was limited, and no hybrid strategy was used across all intelligent techniques. However, the emphasis was placed on using either a sophisticated approach or conventional statistical techniques, with no studies that investigated the attention LSTM method. A review of prior studies showed that the majority (66.7%) of the studies examined a single type of missing data, namely, Missing at Random (MAR). The remaining third of the studies investigated two categories of missing data, namely, MAR and Missing Not at Random (MNAR). Consequently, this research was motivated to investigate the MAR and MNAR types of missing data. The quantity of research pertaining to Data Imputation in financial markets is somewhat scarce compared with other sectors, such as the medical profession.

III. Research Objectives

The objective of this study is to overcome the limitations identified in prior research by introducing a methodology that utilizes DL methods for imputing missing values in the Amman stock market dataset. This study aims to achieve two objectives:

A. To develop and implement a novel Attention-Bi-Long Short-Term Memory (Bi-LSTM) model for imputing missing data in the Amman Stock Market. This objective would focus on the modification of the traditional LSTM approach by incorporating an attention mechanism to enhance the accuracy of imputation.

B. To compare the performance of the proposed Attention Bi-LSTM model with other established imputation techniques, such as Random Forest, Regression, Support Vector Regression (SVR), and Bidirectional LSTM. This objective aims to validate the effectiveness of the proposed model against other well-known methods in the field.

C. To analyze the relationship between the proportion of missing data and the accuracy of imputation using various machine learning techniques. This objective would help in understanding how different levels of missing data affect the performance of the imputation models.

The suggested approach introduces a dynamic adjustment to the attention weights, considering both the temporal context and the missingness pattern. The attention mechanism is now context-aware, meaning it can focus on the most reliable data points for imputing missing values. The final output is a combination of short-term and long-term context vectors, weighted according to their relevance.

IV. Dataset

The Amman Stock Market (ASM) was established in March 1999 as a private corporation that does not seek to make a profit and has complete administrative and financial independence. ASM is authorized to function as a stock exchange, which means that trading in securities may take place there. The exchange is overseen by a board of directors consisting of seven individuals. The day-to-day operations are overseen and reported on by the chief executive officer, who is accountable to the board of directors. ASM is comprised of Jordan's 68 different brokerage firms as members. Moreover, ASM officially became a state-owned business on February 20, 2017 and has been referred to as "The ASM Company" since then [19],[20].

The ASM adheres to the ideals of equity, openness, efficiency, and liquidity. The exchange strives to create a robust and safe environment for its listed securities while also protecting and ensuring the rights of its investors. Furthermore, the ASM has established globally accepted guidelines on market divisions and listing requirements to create this transparent and efficient market.

The ASM and the Jordan Securities Commission (JSC) work closely together on a number of different fronts to address issues relating to surveillance. ASM maintains strong relationships with other exchanges, associations, and international organizations to keep its membership up to date and guarantee that it conforms to international standards and best practices. This initiate helps ASM ensure that it complies with international standards. The World Federation of Exchanges and the Federation of Euro-Asian Stock Exchanges recognize this exchange as a full member in their respective organizations (FEAS). In addition, the exchange is a contributing member of the Arab Federation of Exchanges [20],[21],[22].

ASM is responsible for proving companies with a means of raising capital by showing their equities on the Exchange and encouraging an active market in listed securities based on the effective determination of fair prices. Additionally, the ASM is entrusted with the duty of encouraging an active market in listed securities. The ASM is responsible for transparent trading, the provision of modern facilities and equipment for trading, the recording of trades and the publication of prices, the monitoring and regulation of market trading, and the coordination with the JSC, as required, to ensure legal compliance, in addition to a number of other responsibilities [21].

This study analysed the data obtained from ASM for the period of 2010 to 2021 for 11 different industries. These industries include educational services, insurance, financial services, real estate, financials, health care services, banks, technology, hotels and tourism, transportation, and communications. Table II displays the key characteristics of the ASM dataset.

TABLE II
ASM DATASET CHARACTERISTICS

Dataset Characteristics	Value
Dataset Name	ASM
Data Source	Amman Stock Market
Time Period	2010-2021
Missing Ratio	10%-50%
No. of Records	2940
No. of Observations	32340
No. of Variables	11

V. Research Methodology

This section may be broken down even further into two primary subsections. In the first part, the various methods for making educated guesses about missing data are addressed. Meanwhile, the second part provides a description of how the performance of the techniques used is assessed. This research evaluated the efficacy of various infilling strategies by using a cross-validation approach on data spanning the years 2010 to 2021, as can be seen in

Table III. Because complete data was readily available for this time period, it was decided to use it as the baseline. After being subjected to a random simulation, the whole set of time series data was combed through in order to extract the daily streamflow data that was missing. The process of preparing the dataset and inserting missing data into the whole time series is shown in Fig. 2

TABLE III
SAMPLE DATA FROM ASM

Num	Date	Banks	Insurance	Financial Services	Real Estate	….	….	Technology and Communications
1	3/1/2010	3668.370331	2834.380308	3711.604593	3191.064065	….	….	1818.240146
2	4/1/2010	3644.745756	2851.338919	3786.265458	3252.617986	….	….	1848.458801
3	5/1/2010	3657.163882	2849.223076	3807.878181	3264.147445	….	….	1839.393204
4	6/1/2010	3644.747857	2826.439994	3750.459862	3230.736836	….	….	1847.645109
5	10/1/2010	3673.577171	2824.853649	3778.417771	3295.199304	….	….	1879.491146
6	11/1/2010	3700.293942	2810.926028	3805.1322	3296.317397	….	….	1869.611858
7	12/1/2010	3692.968741	2819.674784	3844.672627	3331.162393	….	….	1862.754436
8	13/1/2010	3691.443754	2794.487538	3823.099369	3332.005821	….	….	1865.776301
9	14/1/2010	3688.066699	2788.332212	3811.341849	3363.918204	….	….	1852.875148
….	….	….	….	….	….	….	….	….
….	….	….	….	….	….	….	….	….
….	….	….	….	….	….	….	….	….
2940	30/12/2021	3894.838925	1929.036136	1363.214108	1679.803958	….	….	603.7909358

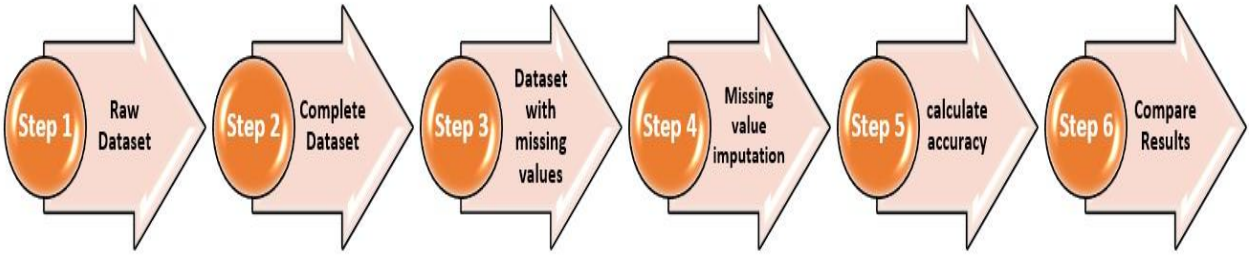

Fig. 2. Process of dataset preparation and incorporating missing data into the entire time series

Fig 3 illustrates the simulation process used for dataset imputation. This process encompasses four primary steps. First, data are collected from the Amman stock market over a span of 12 years. Subsequently, the missing values are randomly generated within the collected dataset. The second step involves using five distinct methods, along with one proposed method, for data imputation. Thereafter, the accuracy of all five methods is calculated in relation to the original dataset. The accuracy is evaluated based on five criteria: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE). The fourth step entails analyzing the outcomes of the simulation. The aforementioned three procedures were executed with a total of 1000 iterations each to maximize the amount of accuracy that could be attained [23], [24], [25], [26].

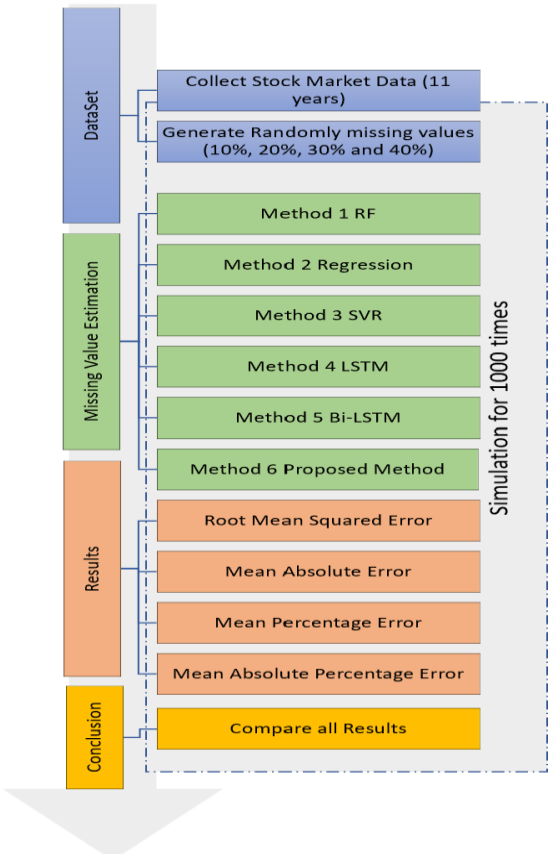

Fig. 3. Simulation Process for Dataset Imputation

VI. MISSING DATA TYPES AND MECHANISMS

Three categories may be used to classify the different forms of missing data. These categories are determined by the link that exists between the missing data mechanism, the seen values, and the missing values. Understanding these categories is essential because the challenges presented by missing values and the approaches used to address those challenges are distinct for each of these three groups [27].

Missing Completely at Random (MCAR): In MCAR, the

missing values for an attribute are not reliant on either the observable data or the unobserved data. The missing value are independent of both types of data. In this type of randomness, adding any sort of bias whatsoever would be impossible, regardless of the treatment used for the missing data.

MAR: In Fig , an occurrence in MAR that has a missing value may be inferred to be reliant on any of the seen values, but it cannot be inferred to be dependent on the unobserved data.

Fig 4. MAR Data Pattern

MNAR: In Fig 5, an instance in MNAR with a missing value depends on the data that have not been seen. Few different approaches may be used to address with this type of unpredictability. One of these techniques is called instance substitution, and it may also be used with mean or mode substitution. This naïve technique can induce bias. Thus, this technique must be treated with caution.

The MCAR and MAR mechanisms are commonly regarded as ignorable, whereas non-ignorable missing value mechanisms are abbreviated as MNAR.

In this research, five different approaches of imputation were compared with the proposed imputation method to identify which approach was the most suitable for filling in missing values in the time series datasets. These approach include RF Imputation method, Regression Imputation (R) method, Support Vector Regression (SVR) imputation method, LSTM, and Bidirectional Long Short-term Memory (BiLSTM) method. In the beginning, a dataset from the time series dataset was arbitrarily removed from the primary dataset. The amounts of missing data ranged from 10% to 50%, with 20%, 30%, 40%, and 50% being the most common. Several imputation approaches were used to fill in all of the missing variables. The accuracy of each approach was determined by comparing the entire datasets that were produced by each method with the dataset that was initially used. The four accuracy parameters used to compare the five imputation methods are MAE, RMSE, MPE, and MAPE.

In the ASM dataset, the first experiment included the generation of missing data in a random manner, with rates ranging from 10% to 50%. In the second experiment, the missing data were intentionally generated in a non-random way, with varying rates ranging from 10% to 50%.

Fig 5. MNAR Data Pattern

VII. Missing Data Imputation

This section explore the Different imputation method that has been implemented for the ASM datasets

RF Imputation

RF imputation is an effective method for addressing missing data, and it leverages the ensemble learning algorithm called RF. The methodology involves constructing a collection of decision trees based on the available data, which are then used to estimate the absent values within the dataset. The use of RF imputation leverages the inherent capability of RF to effectively handle datasets with a large number of dimensions and accurately capture intricate associations among variables [28].

The RF imputation procedure involves the partitioning of the dataset into two distinct components: the observed variables, which possess full data, and the target variable, which has missing values. The predictor variables included in the analysis are utilized to train the RF model. After the completion of the training process, the model is then used to make predictions about the undisclosed values in the target variable, taking into account the observed variables. The process of imputation involves the computation of the mean predictions derived from several decision trees inside the RF. This approach serves to alleviate any possible bias that may arise from a single imputation [28].

Recent research has brought attention to the efficacy of RF imputation across several fields. In a recent study conducted by [29], various imputation techniques were examined for their efficacy in addressing missing data within electronic health records. The findings revealed that RF imputation exhibited superior performance compared with the alternative methods, such as mean imputation or regression imputation, in terms of imputation accuracy and predictive capabilities. A recent investigation conducted by Zhang and [30] examined the issue of missing data imputation within environmental monitoring datasets. The research asserted that the utilization of RF imputation yielded notable outcomes in terms of retaining the fundamental structure of the data and effectively capturing its temporal trends.

Regression Imputation

Simple and multiple linear regression imputation is a technique used to handle missing data by estimating the missing values based on the relationships observed in the data. Simple linear regression imputation involves using a single independent variable to predict the missing values, while multiple linear regression imputation incorporates multiple independent variables. These methods leverage the linear relationships between variables to impute missing values and are commonly used when the missingness is assumed to be related to the available data [31].

In a simple linear regression imputation, the missing values are imputed by regressing the variable of interest on a single independent variable. The regression equation is estimated using the observed data, and the missing values are predicted based on this relationship. The imputed values are obtained by substituting the missing variable values into the regression equation [31].

Multiple linear regression imputation expands on simple linear regression by incorporating multiple independent variables in the regression equation. This mechanism allows for a more comprehensive modeling of the relationships between variables. The coefficients of the regression equation are estimated using available data, and the missing values are predicted based on the relationship between the dependent and the independent variables [31].

Recent studies have explored the application of simple and multiple linear regression imputation in various domains. Reference [32] investigated the use of simple linear regression imputation for handling missing data in a healthcare dataset. The results showed that simple linear regression imputation performed well in imputing missing values and improved the accuracy of subsequent analyses. Another study by [32] applied multiple linear regression imputation to handle the missing values in a dataset related to water quality monitoring. The findings demonstrated the effectiveness of multiple linear regression imputation in capturing the relationships among variables and producing accurate imputations.

SVR Imputation

SVR imputation is a powerful technique for handling missing data by leveraging the principles of SVR. SVR extends Support Vector Machines (SVMs) to the regression setting and aims to find a hyperplane that best fits the relationship between the observed and the missing variables. SVR imputation works by training an SVR model on the observed data and using it to predict the missing values based on the patterns learned from the available data [33].

Furthermore, SVR imputation involves finding an optimal hyperplane that maximizes the margin around the observed data points while minimizing the deviation of the predicted values from the actual values. SVR uses a kernel function to map the data into a higher-dimensional feature space, allowing for nonlinear relationships to be captured. During the training phase, SVR identifies support vectors that are crucial for determining the hyperplane and uses them to make predictions for the missing values. The predicted values are then imputed as the missing values [33].

Recent research has explored the applications and advancements of SVR imputation in various domains. A study by [29] applied SVR imputation to handle missing data in electronic health records. The results demonstrated the effectiveness of SVR in imputing missing values and improving the performance of downstream analysis tasks, such as disease prediction. Another study by [30] investigated the use of SVR imputation for missing data in environmental monitoring datasets. The research showcased the ability of SVR in capturing complex relationships and producing accurate imputations in environmental data.

LSTM Imputation

LSTM imputation is an advanced technique for handling missing data that leverages the power of LSTM neural networks, a variant of Recurrent Neural Networks (RNNs), to model and impute missing values in sequential data. LSTM networks are specifically designed to overcome the vanishing gradient problem of traditional RNNs and effectively capture long-term dependencies in sequential data. LSTM imputation works by training an LSTM model on the observed data and utilizing the learned patterns to fill in the missing values [13].

Furthermore, LSTM imputation (Fig. 4) involves encoding the sequential nature of the data and utilizing the internal memory cells of the LSTM network. The network consists of memory cells that store information over time and three gates (input, forget, and output) that control the flow of information. During the training phase, the LSTM model learns to predict the next value based on the previous values. Once the model is trained, it can be used to impute missing values by providing the observed values as inputs and generating predictions for the missing values [13].

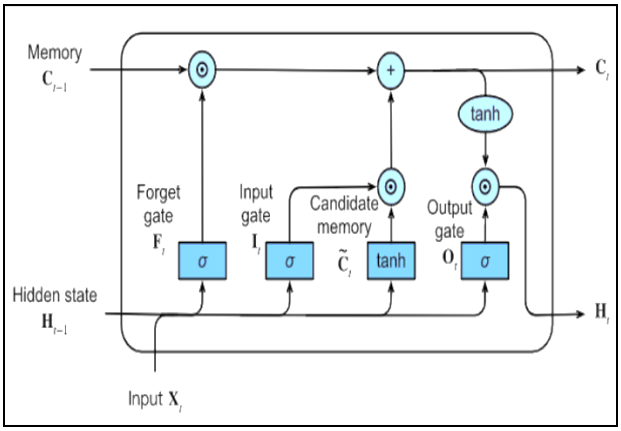

Fig. 4. Structure of LSTM

Recent studies have demonstrated the effectiveness of LSTM imputation in various domains. A study by [34] applied LSTM imputation for handling missing values in smart grid data and achieved improved imputation accuracy compared with the traditional imputation methods. Another study by [35] explored the use of LSTM imputation for missing data in electronic health records and showed that it outperformed other imputation techniques in terms of preserving the temporal dependencies and improving the performance of downstream predictive models.

BiLSTM Imputation

BiLSTM imputation is a technique used to fill in missing values in sequential data, such as time series data or natural

language sentences, using bidirectional LSTM networks. LSTM is a type of RNN that is well-suited for handling sequences of data due to its ability to capture long-term dependencies.

The BiLSTM imputation procedure may be described in the following sequential manner:

Input Data Preparation: The first step is to prepare the input data. This step involves structuring the sequential data into a format that can be fed into the BiLSTM model. The data should be represented as a sequence of data points, with missing values replaced by placeholders (e.g., NaN or zero).

BiLSTM Architecture: The BiLSTM model consists of two LSTMs (Fig. 5), one processing the sequence in the forward direction and the other in the backward direction. The forward LSTM reads the sequential data from the beginning to the end, while the backward LSTM processes the data in the reverse order. This bidirectional architecture enables the model to capture information from past and future contexts, which helps in effectively imputing missing values [36].

- Training: The BiLSTM model is trained on the data with missing values. During the training, the model learns to predict the missing values based on the available context from the forward and backward directions. The loss function used during training measures the discrepancy between the predicted values and the actual observed values at the non-missing positions in the data.

- Masking: A technique called masking is employed to handle the missing values during training. A binary mask is created that marks the positions of missing values as zero and the positions of observed values as one. The mask is multiplied element-wise with the loss function. Thus, the model does not consider the missing values when calculating the loss, focusing only on the observed values.

- Imputation: After the BiLSTM model is trained, it can be used to impute the missing values in the unseen data. During the imputation phase, the model takes the entire sequence as input, including the positions with missing values. The model generates predictions for the missing values based on the available context in the forward and backward directions.

- Post-processing: The imputed values are obtained as the outputs of the BiLSTM model. The post-processing steps may be applied to the imputed values, depending on the application. For example, you might clip imputed values to certain bounds, round them to integers, or use a threshold to convert them into binary values.

- Performance Evaluation: The performance of the BiLSTM imputation can be assessed using various metrics, such as Mean Squared Error (MSE) or MAE, by comparing the imputed values with the true observed values in the validation or test dataset.

BiLSTM imputation is especially useful when dealing with long sequences of data with complex temporal dependencies. BiLSTM models can make more accurate predictions for missing values and provide robust imputation results in various applications, including time series forecasting, natural language processing, and medical data analysis, by capturing context from past and future observations [37].

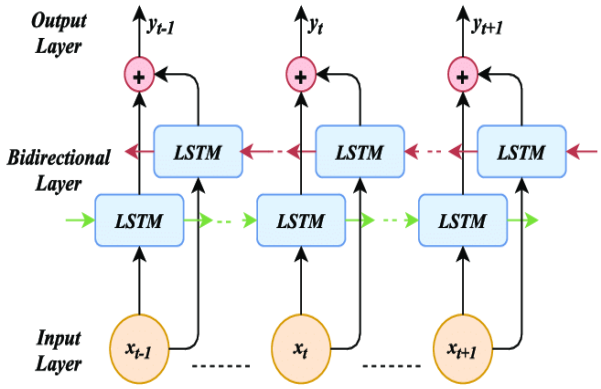

Fig. 5. Structure of a Bi-LSTM algorithm "redrawing based on [37]"

Attention Bi-LSTM

In attention Bi-LSTM the attention mechanism is designed to allow the model to focus on specific parts of the input sequence when making predictions. Instead of treating all elements of the sequence equally, the attention mechanism assigns different weights (importance) to different parts of the sequence. This is particularly useful in scenarios where certain parts of the input sequence are more relevant to the task at hand, such as in machine translation or speech recognition.

A. Attention Bi-LSTM mechanism

Fig. 6 summarizes the main idea of the attention Bi-LSTM imputation method. The attention Bi-LSTM model involves three main steps as follows:

1- Processing with Bi-LSTM

The input sequence is first fed into the Bidirectional LSTM layer, where the sequence is processed in both forward and backward directions. This results in two sets of hidden states: one from the forward LSTM and one from the backward LSTM. These hidden states are then concatenated to form a comprehensive representation of the sequence at each time step.

2- Applying Attention Mechanism:

The concatenated hidden states from the Bi-LSTM are then passed through the attention mechanism. The attention mechanism calculates a set of attention weights, which determine the importance of each hidden state in making predictions.

These weights are used to create a weighted sum of the hidden states, which emphasizes the most relevant parts of the sequence.

3- Final Prediction:

The output of the attention mechanism (a context vector) is then used as the input to the final layer(s) of the model, which could be a dense layer or another LSTM layer, depending on the specific application.

The final output is generated based on this context vector,

which now contains a focused and enriched representation of the sequence.

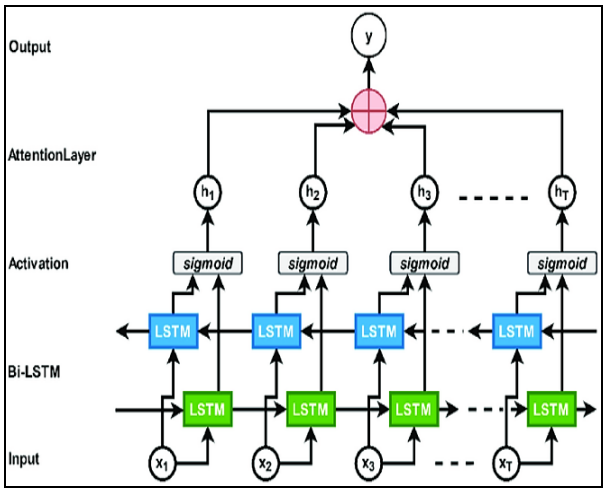

Fig. 6. The Architecture of the Attention Bi-LSTM

B. *Mathematical Model for Traditional Attention Bi-LSTM*

The LSTM cell maintains a memory state that is updated at each timestep. The operations within an LSTM cell are as in (1) to (6).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{(Forget Gate)} \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{(Input Gate)} \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \text{(Candidate Memory Cell)} \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \text{(Memory Cell Update)} \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{(Output Gate)} \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad \text{(Hidden State Update)} \quad (6)$$

C. *Bidirectional LSTM*

In a Bidirectional LSTM, the input sequence is processed in both forward and backward directions. Let $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the hidden states of the forward and backward LSTM layers, respectively. The final hidden state at time t is given by (7).

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \quad (7)$$

D. *Attention Mechanism*

The attention mechanism computes a context vector that allows the model to focus on specific parts of the input sequence as in (8) to (10).

$$e_{t,s} = \text{score}(h_t, \bar{h}_s) \quad \text{(Alignment Score)} \quad (8)$$

$$\alpha_{t,s} = \frac{\exp(e_{t,s})}{\sum_{s'} \exp(e_{t,s'})} \quad \text{(Attention Weights)} \quad (9)$$

$$c_t = \sum_s \alpha_{t,s} \cdot \bar{h}_s \quad \text{(Context Vector)} \quad (10)$$

The context vector c_t is combined with the hidden state h_t to produce the final output at each timestep as in (11).

$$y_t = f(c_t, h_t) \quad (11)$$

E. *Proposed Attention Bi-LSTM Imputation*

The modified Attention Bi-LSTM model introduces dynamic adjustments to the attention weights based on the temporal context and the missingness pattern in the data. This modification allows the model to pay more attention to time steps that provide the most reliable context for imputing missing values.

Each timestep updates the memory state maintained by the LSTM cell. The processes within an LSTM cell are identical to those of a conventional LSTM. In addition, the input sequence is processed in both forward and backward directions, similar to Bi-LSTM. Furthermore, the attention component has been developed to incorporate a dynamic context-aware attention mechanism that adapts attention weights according to both the temporal context and the missingness pattern.

A description of the functions of each model component for the modified attention Bi-LSTM is provided in the following steps and components.

Input Data and LSTM: The input data provided in this study consists of a sequence of items, namely a sequence of closing prices. This series is inputted into the LSTM network, which has the ability to acquire knowledge of extended relationships and capture sequential patterns.

Encoder–Decoder Architecture (Optional): Attention LSTMs are commonly used in encoder–decoder architectures, wherein the input sequence is first processed by an encoder LSTM, and the decoder LSTM generates the output sequence based on the encoded information. This architecture is commonly used in tasks, such as machine translation, where the input sequence (source language) is encoded, and the output sequence (target language) is generated.

Dynamic Attention Mechanism: The attention mechanism is introduced between the encoder and decoder LSTMs. This mechanism allows the decoder LSTM to selectively focus on certain parts of the encoded input sequence while generating each element of the output sequence. The attention weights are dynamically adjusted based on the temporal context and the missingness pattern as in (12).

$$\alpha_{t,s}^{(\text{dynamic})} = \frac{\exp(e_{t,s} + \delta_{t,s})}{\sum_{s'} \exp(e_{t,s'} + \delta_{t,s'})} \quad (12)$$

where $\delta_{t,s}$ is an adjustment factor that depends on the missingness pattern at time s and the temporal context.

Attention Weights: The attention mechanism computes attention weights for each element in the input sequence. These attention weights indicate how relevant each element is to the generation of the current output element. The attention weights are typically computed based on a similarity measure between the current state of the decoder LSTM and the hidden states of the encoder LSTM.

Context Vector: The attention weights are used to compute a context vector, which is a weighted sum of the encoder LSTM's hidden states. The context vector represents the parts of the input sequence that the decoder LSTM should pay attention to when generating the current output element. The model generates multiple context vectors by focusing on different parts of the sequence, such as short-term and long-term dependencies as in (13) and (14).

$$c_t^{(\text{short})} = \sum_{s \in \text{short}} \alpha_{t,s}^{(\text{short})} \cdot \bar{h}_s \quad \text{(Short-term Context Vector)} \quad (13)$$

$$c_t^{(\text{long})} = \sum_{s \in \text{long}} \alpha_{t,s}^{(\text{long})} \cdot \bar{h}_s \quad \text{(Long-term Context Vector)} \quad (14)$$

Final Context Vector: The final context vector is a weighted combination of the different context vectors as in (15).

$$c_t^{(\text{final})} = \beta^{(\text{short})} \cdot c_t^{(\text{short})} + \beta^{(\text{long})} \cdot c_t^{(\text{long})} \quad (15)$$

where $\beta^{(\text{short})}$ and $\beta^{(\text{long})}$ are dynamically determined weights that prioritize the most relevant context based on the missingness pattern and temporal dependencies

Context-Aware LSTM: The context vector is concatenated with the input to the decoder LSTM at each time step. This approach makes the decoder LSTM "context-aware", which means that it now has access to information about the relevant parts of the input sequence, as determined by the attention mechanism.

Output Generation: With the attention mechanism guiding the decoder LSTM's focus, it generates the output sequence, one element at a time. The context-aware nature of the decoder LSTM helps it make more informed decisions based on the relevant information from the input sequence. The final output at each timestep is computed using the dynamically weighted context vector as in (16).

$$y_t^{(\text{modified})} = f\left(c_t^{(\text{final})}, h_t\right) \quad (16)$$

Training: During training, the attention LSTM learns to adjust its attention weights and context vector calculation to optimize the task-specific objective function, such as minimizing the cross-entropy loss in machine translation or text generation.

Overall, the attention mechanism in an Bi-LSTM allows the model to dynamically focus on different parts of the input sequence, improving its ability to effectively process long sequences and generate more accurate and contextually relevant output sequences.

The sequential structure of an LSTM layer includes multiple memory cells with inputs x_t of closing price in stock market and two other states: the previous cell state C_{t-1} and the hidden state h_{t-1}.

The LSTM cell state is described in (4) [37].

Table IV presents a comprehensive overview of the parameters used in the suggested methodology. The utilization of the Model Checkpoint callback is employed in combination with the model training process. The *fit()* function is used to save a model or weights in a checkpoint file at regular intervals. This mechanism allows for the subsequent loading of the model or weights, enabling the continuation of training from the previously stored state. This callback has many possibilities.

The decision on whether to retain just the model that has attained the highest level of performance so far or to preserve the model at the conclusion of each epoch irrespective of its performance is a matter of consideration. The question at hand pertains to the decision of retaining just the model that has shown the highest level of performance so far or preserving the model at the conclusion of each epoch irrespective of its performance.

TABLE IV
MODEL CONFIGURATION

Parameter	Value
Epoch	100
Learning rate	0.001
Batch size	32
Optimizer	Adam
Hidden layer	2
Input layer	1,2
Units	50
Callback	Model check points

VIII. DATASET PREPROCESSING

To guarantee the efficacy and dependability of the suggested Attention Bi-LSTM model for missing data imputation, it is essential to perform a number of data pretreatment evaluations. The Stationarity Test, Correlation Analysis, and Seasonality Analysis are crucial for comprehending the fundamental structure, relationships, and trends within the dataset. The Stationarity Test determines if time series data have consistent statistical features across time, which is essential for numerous time series models. Correlation analysis elucidates the interdependencies among variables, essential for understanding multivariate interactions during imputation. Finally, Seasonality Analysis reveals cyclical patterns and variations, guaranteeing that the temporal context is properly represented. The preprocessing methods establish a solid basis for the implementation of the Bi-LSTM model, improving its capacity to reliably impute absent values in intricate multivariate time series datasets [38].

A. *Stationarity Test*

The Augmented Dickey-Fuller (ADF) test was employed to evaluate the stationarity of the dataset across all numeric columns. The findings suggested that certain variables, including "Banks," had non-stationarity, as evidenced by p-values over 0.05. This indicates that these time series exhibit trends or other types of non-stationarity that may necessitate

differencing or alternative transformations prior to additional analysis or modeling. Conversely, the columns "Insurance" and "Financial Services" exhibited stationarity, as their p-values fell below the 0.05 threshold, indicating consistent statistical features over time.

TABLE V
RESULTS OF THE AUGMENTED DICKEY-FULLER TEST FOR STATIONARITY OF NUMERIC VARIABLES

Sector	ADF Statistic	p-value	Stationary
Banks	-1.6672	0.4481	No
Insurance	-3.9964	0.0014	Yes
Financial Services	-2.9182	0.0433	Yes
Real Estate	-3.2165	0.0190	Yes
Financials	-2.0191	0.2783	No
Health Care Services	-2.4557	0.1266	No
Educational Services	-0.6670	0.8551	No
Hotels and Tourism	-2.3156	0.1670	No
Transportation	-4.2753	0.0005	Yes
Technology and Communications	-1.4100	0.5775	No

Table V displays the outcomes of the ADF test conducted on all numeric variables within the dataset. Variables such as "Insurance" and "Financial Services" were determined to be stationary, as evidenced by p-values below the 0.05 level, indicating steady statistical features. In contrast, non-stationary variables, such as "Banks," exhibited elevated p-values, indicating the existence of trends or dynamic patterns that may necessitate differencing or other transformations for analysis.

B. *Correlation analysis*

Correlation analysis was conducted to examine the links among numeric variables in the dataset. The correlation matrix revealed both robust and feeble connections among variables. "Financial Services" and "Real Estate" demonstrated a strong positive association, suggesting that both variables may display interdependent patterns. Conversely, minor correlations were noted between the variables "Banks" and "Insurance," indicating negligible interaction or influence between these factors. These insights offer essential direction for feature selection and modeling methodologies.

Table VI presents a detailed overview of the correlations between numeric variables in the dataset. Robust positive correlations, exemplified by the relationship between "Financial Services" and "Real Estate," signify considerable dependency, which can guide imputation and predictive modeling methodologies. Weak or negative correlations, such as those between "Banks" and "Insurance," indicate restricted interactions, implying low mutual influence. This approach facilitates the prioritization of characteristics for sophisticated modeling and imputation tasks.

C. *Seasonality Analysis*

The seasonal decomposition of chosen time series, namely "Banks," "Insurance," "Financial Services," and "Real Estate," was performed to analyze their fundamental patterns. Each series was analyzed into observable, trend, seasonal, and residual components:

1. Observed: The original time series data offers an extensive perspective of all integrated patterns.
2. Trend: The trend component indicates the long-term trajectory of the data, emphasizing sustained gains or reductions over time.
3. Seasonality: The seasonal component identifies recurring trends within a defined timeframe, such as annual or monthly cycles.
4. Residuals: Residual components signify random noise or anomalies that are not accounted for by the trend or seasonality.

In the "Banks" time series, a distinct ascending trend and annual seasonal swings were noted, signifying predictable periodic oscillations. Comparable patterns were observed in "Insurance" and "Financial Services," where seasonal trends align with cyclical habits in the financial industry. These decompositions offer critical insights into the structure and dynamics of the time series data, facilitating model selection and parameter optimization.

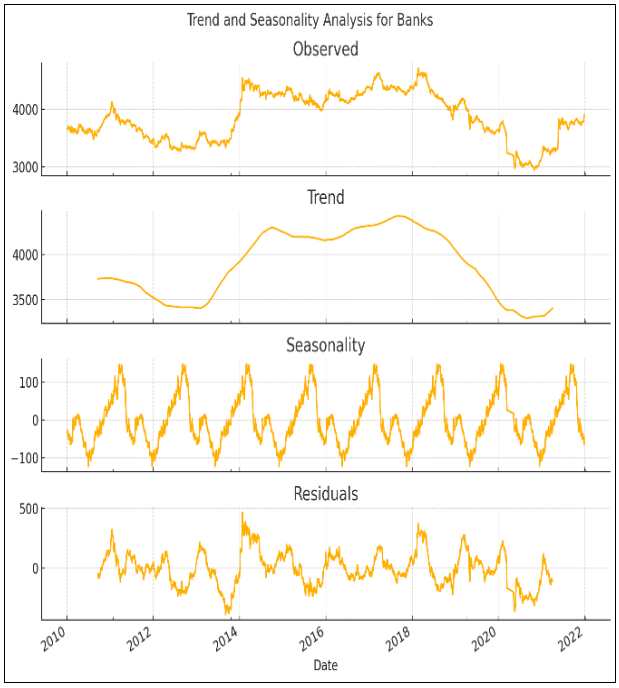

Fig 7. Seasonal Decomposition of the "Banks" Time Series.

As seen from *Fig 7* the analysis of the "Banks" time series indicates a distinct increasing trend and significant annual seasonal variations. The trend indicates a sustained positive trajectory, whereas the seasonal component reflects regular periodic fluctuations. The residual component signifies noise or anomalies that are negligible compared to the prevailing patterns.

As illustrated in Fig 8 The breakdown of the "Insurance" time series reveals consistent seasonal trends along with a little rising trend. The detected seasonality corresponds with cyclical patterns characteristic of the banking industry. The

residuals exhibit negligible variation, indicating that the model accounts for the majority of the variability in the data.

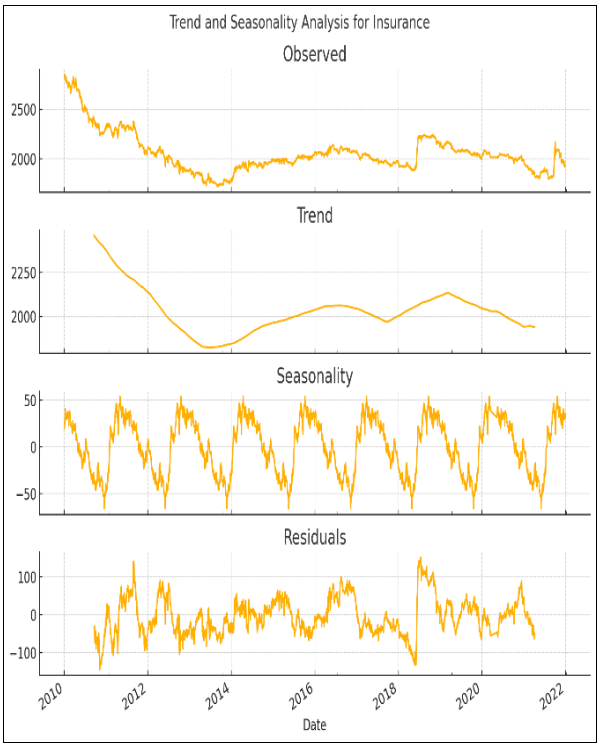

Fig 8. Seasonal Decomposition of the "Insurance" Time Series.

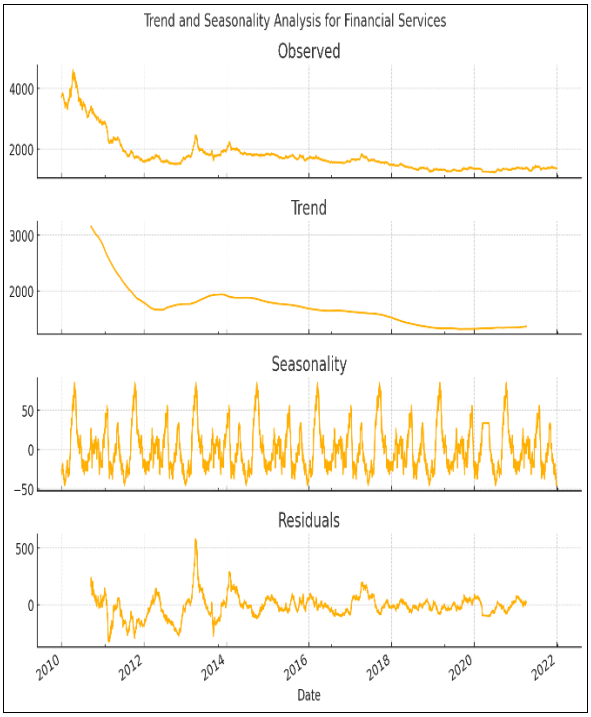

Fig 9. Seasonal Decomposition of the "Financial Services" Time Series.

As illustrated in Fig 9, the "Financial Services" time series exhibits a pronounced increasing trajectory accompanied by notable seasonal variations. These cyclical patterns indicate consistent economic or market cycles. The residuals exhibit negligible noise, hence reinforcing the precision of the decomposition.

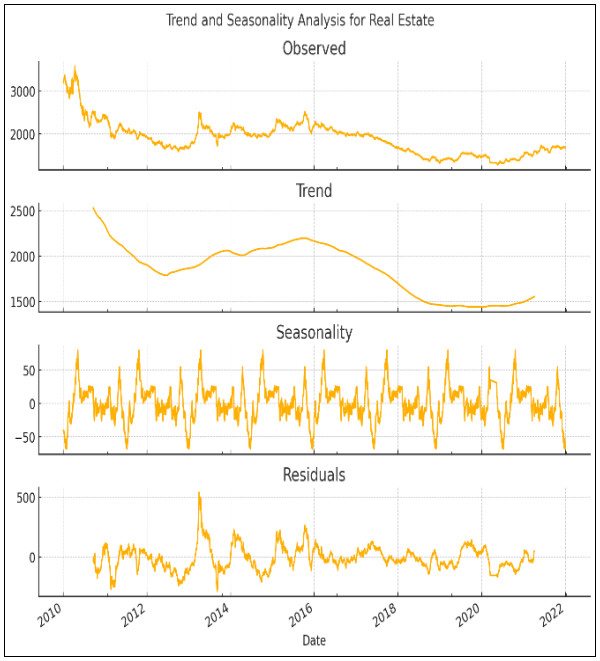

Fig 10. Seasonal Decomposition of the "Real Estate" Time Series.

as depicted in Fig 10 The "Real Estate" time series decomposition reveals a consistent rising trend accompanied by mild seasonal fluctuations. These patterns indicate persistent behaviors in the sector, while the residuals imply minimal irregularity, underscoring the reliability of observed trends.

IX. Discussion and Results

In this section, the 2940 records of Amman stock market data is considered to investigate the various imputation methods considered in this study. The experiments were conducted using RStudio software with R packages to compare the proposed algorithm with the existing algorithms. Imputation methods are compared based on five measures of performance, including MAE, RMSE, computed average of percentage errors (MAPE), MPE, and MAPE, which are defined as in (17) to (20).

$$MAE(H) = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{17}$$

$$RMSE(H) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}}, \tag{18}$$

$$MPE(H) = \frac{100\%}{n} \sum_{i=1}^{n} \frac{y_i - x_i}{y_i}, \tag{19}$$

$$MAPE(H) = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - x_i}{y_i} \right|. \tag{20}$$

The results are presented inTable VII, and a graphical representation is utilized in Fig 13–18.

TABLE VI
PAIRWISE CORRELATION MATRIX HIGHLIGHTING RELATIONSHIPS AMONG VARIABLES.

	Banks	Insurance	Financial Services	Real Estate	Financials	Health Care Services	Educational Services	Hotels and Tourism	Transportation	Technology and Communications
Banks	1.0000	-0.0364	-0.0455	0.1720	0.8671	0.2375	0.7275	-0.0571	-0.0821	-0.2780
Insurance	-0.0364	1.0000	0.6926	0.5161	0.3275	-0.0939	-0.2772	0.6563	0.6698	0.3648
Financial Services	-0.0455	0.6926	1.0000	0.8645	0.4518	0.1622	-0.0193	0.8568	0.9335	0.6694
Real Estate	0.1720	0.5161	0.8645	1.0000	0.6070	0.2811	0.3497	0.7224	0.8077	0.6041
Financials	0.8671	0.3275	0.4518	0.6070	1.0000	0.2843	0.6526	0.3693	0.3868	0.0825
Health Care Services	0.2375	-0.0939	0.1622	0.2811	0.2843	1.0000	0.5374	0.3773	0.2815	0.4138
Educational Services	0.7275	-0.2772	-0.0193	0.3497	0.6526	0.5374	1.0000	-0.0001	-0.0379	-0.0442
Hotels and Tourism	-0.0571	0.6563	0.8568	0.7224	0.3693	0.3773	-0.0001	1.0000	0.9423	0.8638
Transportation	-0.0821	0.6698	0.9335	0.8077	0.3868	0.2815	-0.0379	0.9423	1.0000	0.8284
Technology and Communications	-0.2780	0.3648	0.6694	0.6041	0.0825	0.4138	-0.0442	0.8638	0.8284	1.0000

TABLE VII
ERROR RATE AMONG ALL IMPUTATION METHODS

		Imputation Method					
Evaluation Criteria	Missing %	RF	R	SVR	LSTM	Bi-LSTM	Modified Attention Bi-LSTM
MAE	10%	1.04947	6.70172	15.32649	1.01987	1.00697	0.95359
	20%	2.25394	15.23852	14.31507	2.01063	1.98418	1.89182
	30%	4.08515	26.60835	10.73324	3.14900	3.02171	2.90371
	40%	6.21976	38.23798	6.26303	4.05963	4.05269	3.85694
	50%	9.52158	53.04006	3.81529	5.25372	5.08916	4.85815
MAPE	10%	0.06425	0.46205	0.76074	0.05971	0.05896	0.05517
	20%	0.14033	1.04904	0.69823	0.12265	0.12152	0.11381
	30%	0.25249	1.72579	0.47570	0.18918	0.18264	0.17217
	40%	0.39654	2.50021	0.26119	0.25038	0.24536	0.22725
	50%	0.58634	3.40291	0.15643	0.31275	0.30567	0.28535
MPE	10%	0.00456	0.06338	0.32718	0.00671	0.01093	0.00200
	20%	0.00931	0.15708	0.49337	0.01010	0.02673	0.00051
	30%	0.01212	0.21831	0.35262	0.02916	0.03474	0.00512
	40%	0.03908	0.40533	0.18989	0.01950	0.06102	0.00674
	50%	0.00978	0.59584	0.10311	0.04610	0.06378	0.01190
RMSE	10%	0.24013	0.58173	10.80395	0.49271	0.46070	0.22854
	20%	0.15593	1.41718	15.21188	0.87458	0.78877	0.55038
	30%	0.36065	1.98509	12.52165	1.62271	1.33826	0.76410
	40%	0.57771	1.19726	7.24863	2.01485	1.79464	1.08982
	50%	1.24506	2.44852	4.49474	2.74584	2.47492	1.28805

Fig. 11 represents the MAE of 2940 records of the Amman stock market. The six imputation methods (RF, R, SVR, LSTM, BiLSTM, and Attention LSTM) were compared. The results showed that the Attention LSTM method obtained the least MAE, followed by the BiLSTM and LSTM methods. Moreover, MAE increases with the increase in the missing data percentage.

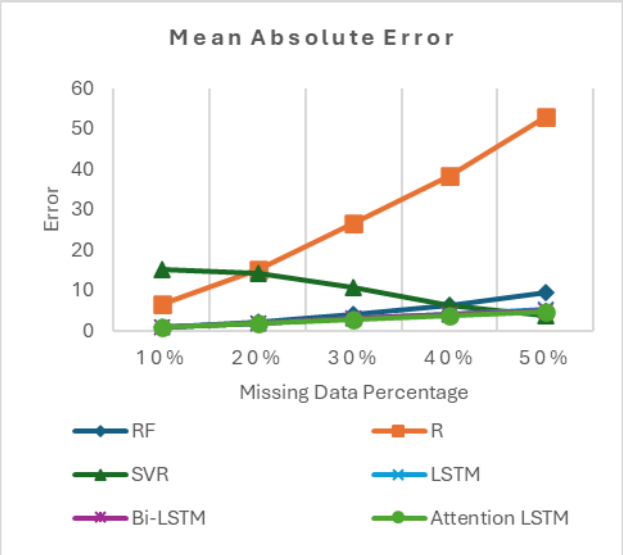

Fig. 11. MAE of 2940 Records of Amman Stock Market

Fig. 12 shows the RMSE of 2940 records of Amman stock market. The six imputation methods (RF, R, SVR, LSTM, BiLSTM, and Attention LSTM) were compared.

The results showed that the RF method obtained the least RMSE with a slight superiority of the attention LSTM method, while the SVR exhibits a high RMSE. Futhermore, the SVR method starts to obtain less RMSE after 20% of missing ratio.

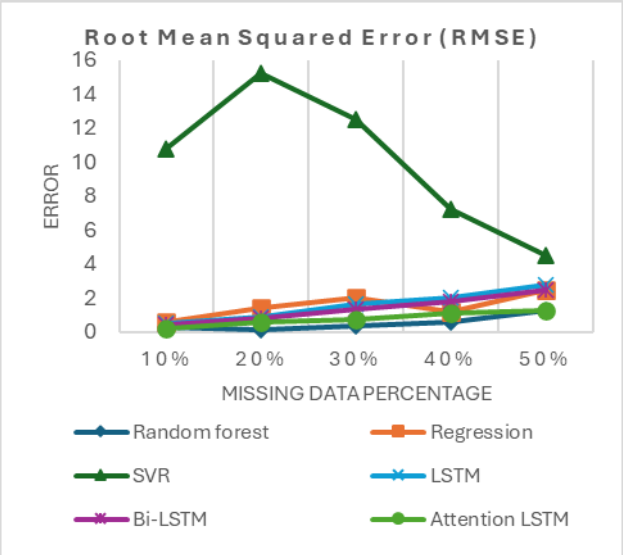

Fig. 12. RMSE of 2940 Records of Amman Stock Market

Fig. 13 shows the MPE of 2940 records of Amman stock market. The six imputation methods (RF, R, SVR, LSTM, BiLSTM and Attention LSTM) were compared. The results demonstrated that the Attention LSTM method obtained the least MPE with a slight superiority of the RF method. Meanwhile, the R method starts to obtain a massive error after 20% of data missing. By contrast, the SVR starts to receive less error after 20% of missing data.

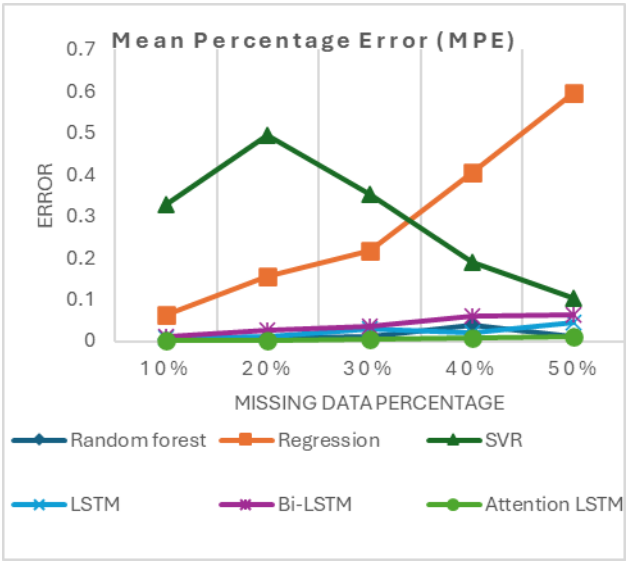

Fig. 13. MPE of 2940 Records of Amman Stock Market

Fig 14 shows the MAPE of 2940 records of Amman stock market. The six imputation methods (RF, R, SVR, LSTM, BiLSTM, and Attention LSTM) were compared. The results indicated that the attention LSTM methods obtained the least MAPE from 10% to 40% of missing data. Meanwhile, the SVR performed better at 50% of missing data. Furthermore, the regression method exhibits an increasing error from 10% to 50%.

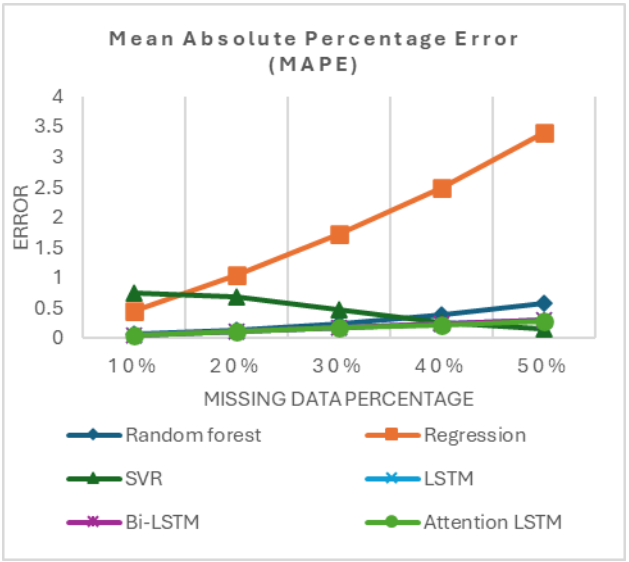

Fig 14. MAPE of 2940 Records of Amman Stock Market

Fig. 15 shows the MPE of 2940 records of Amman stock market for MNAR dataset. The two imputation methods (BiLSTM and Attention LSTM) were compared. The results showed that the attention LSTM methods obtained the least MPE from 10% to 50% of missing data.

Fig. 16 shows the MAPE of 2940 records of Amman stock market for the MNAR dataset. The two imputation methods (BiLSTM and Attention LSTM) were compared, and the results indicated that the attention LSTM methods obtained the least MAPE from 10% to 50% of missing data.

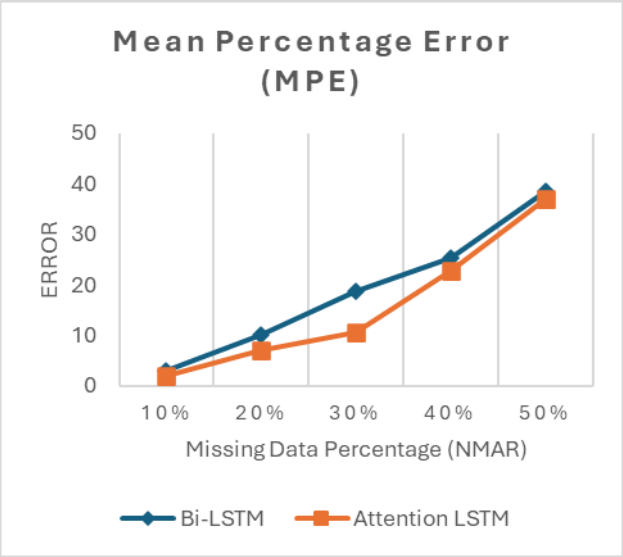

Fig. 15. MPE for the MNAR Dataset

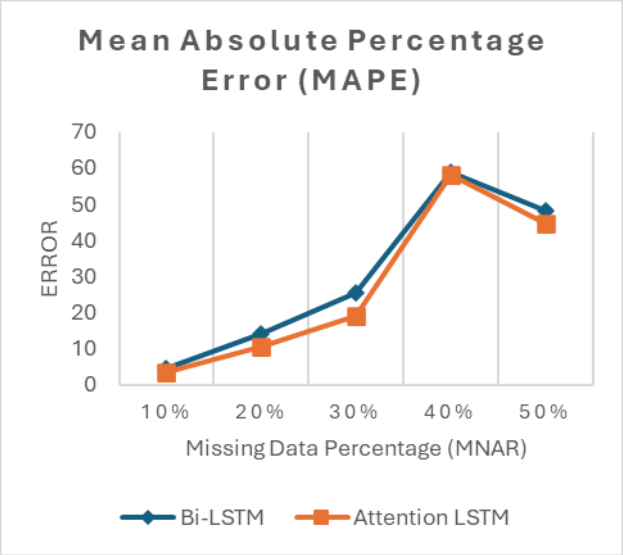

Fig. 16. MAPE for the MNAR Dataset

X. Conclusion

In this study, a novel imputation method, Attention Bi-LSTM, was proposed for addressing missing data in financial markets, specifically applied to the Amman Stock Market dataset. The performance of the proposed model was compared against five other imputation techniques Random Forest, Regression, SVR, LSTM, and Bi-LSTM—using key evaluation metrics such as MAE, RMSE, MPE, and MAPE. The results demonstrated that the Attention Bi-LSTM outperformed these traditional methods, especially in scenarios involving higher percentages of missing data (up to 50%).

The integration of the attention mechanism into the Bi-LSTM architecture allowed the model to dynamically focus on the most relevant portions of the time series data, improving the precision of imputations. Notably, the proposed model yielded the lowest MAE and MAPE values across all imputation methods, and its performance advantage increased as the percentage of missing data rose. This result underscores the resilience of the Attention Bi-LSTM in handling large-scale missing data problems, making it highly suitable for financial datasets where missingness is a common issue due to various factors like transmission failures or recording errors.

From a computational standpoint, the Attention Bi-LSTM was able to balance accuracy with efficiency, achieving higher accuracy while maintaining manageable computational complexity. This is particularly relevant for real-time applications in stock markets, where imputation speed and accuracy are both critical.

However, some limitations remain. While the Attention Bi-LSTM outperformed in most cases, the Random Forest method occasionally exhibited better performance in terms of RMSE, especially at lower missing data ratios. Additionally, the SVR model displayed improved accuracy when dealing with higher proportions of missing data (close to 50%), which suggests that certain traditional machine learning models might still have merit in specific scenarios, particularly when computational efficiency is paramount.

XI. Future Work

Moving forward, this research can be expanded in several important directions. One key avenue for future work involves incorporating additional financial datasets from other markets, such as commodities like gold and silver, as well as digital currencies like Bitcoin and Ethereum. This extension would provide a broader validation of the proposed model's effectiveness across different financial domains.

Additionally, exploring hybrid models that combine the strengths of the Attention Bi-LSTM with other machine learning or statistical techniques, such as Random Forest, could further enhance imputation accuracy [39]. These hybrid approaches might leverage the complementary strengths of different methods, especially when dealing with datasets that contain mixed missing data mechanisms (MAR, MNAR, etc.).

Another important direction is to investigate the scalability of the Attention Bi-LSTM model on larger datasets, as well as in real-time applications. The model's adaptability to varying data frequencies (e.g., high-frequency trading data) and its efficiency in real-time prediction environments remain areas for further optimization.

Finally, future research should explore the inclusion of advanced attention mechanisms, such as multi-head attention or self-attention models, which have shown promise in natural language processing and could be adapted to further improv imputation tasks in financial time series data.

Acknowledgment

I extend my appreciation to the members of the research team who helped with the writing of this article.

References

[1] C. M. Boya, “From efficient markets to adaptive markets: Evidence

from the French stock exchange," *Res. Int. Bus. Financ.*, vol. 49, pp. 156–165, 2019, doi: 10.1016/j.ribaf.2019.03.005.
[2] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, 2020, doi: 10.1007/s00521-019-04212-x.
[3] Mohd Tahir Ismail, and Remal Shaher Al-Gounmeein, "Overview of Long Memory for Economic and Financial Time Series Dataset and Related Time Series Models: A Review Study," IAENG International Journal of Applied Mathematics, vol. 52, no.2, pp261-269, 2022
[4] R. J. A. Little and D. B. Rubin, *Statistical analysis with missing data*, vol. 793. John Wiley \& Sons, 2014. doi: 10.1002/9781119013563.
[5] F. B. Hamzah, F. M. Hamzah, S. F. M. Razali, and H. Samad, "A comparison of multiple imputation methods for recovering missing data in hydrological studies," *Civ. Eng. J.*, vol. 7, no. 9, pp. 1608–1619, 2021, doi: 10.28991/cej-2021-03091747.
[6] E. Di Valentino *et al.*, "In the realm of the Hubble tension - A review of solutions," *Class. Quantum Gravity*, vol. 38, no. 15, p. 153001, 2021, doi: 10.1088/1361-6382/ac086d.
[7] J. R. van Ginkel, M. Linting, R. C. A. Rippe, and A. van der Voort, "Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data," *J. Pers. Assess.*, vol. 102, no. 3, pp. 297–308, 2020, doi: 10.1080/00223891.2018.1530680.
[8] J. S. Murray, "Multiple imputation: A review of practical and theoretical findings," *Stat. Sci.*, vol. 33, no. 2, pp. 142–159, 2018, doi: 10.1214/18-STS644.
[9] D. Adhikari *et al.*, "A Comprehensive Survey on Imputation of Missing Data in Internet of Things," *ACM Comput. Surv.*, vol. 55, no. 7, 2022, doi: 10.1145/3533381.
[10] A. A. El-Sheikh, F. A. Alteer, and M. R. Abonazel, "Four imputation methods for handling missing values in the ardl model: An application on libyan fdi," *J. Appl. Probab. Stat.*, vol. 17, no. 3, pp. 29–047, 2022.
[11] T. Njuguna, "Modelling Stock Market Volatility Using Random By Terry Njuguna a Research Project Submitted in Partial Fulfilment of the Requirement for the Award of Master of Science , Finance At the School of Business , the University of Nairobi November 2021 Declarat," no. November, Nov. 2021.
[12] M. Biswas, A. Shome, M. A. Islam, A. J. Nova, and S. Ahmed, "Predicting stock market price: A logical strategy using deep learning," in *ISCAIE 2021 - IEEE 11th Symposium on Computer Applications and Industrial Electronics*, 2021, pp. 218–223. doi: 10.1109/ISCAIE51753.2021.9431817.
[13] H. Ahn, K. Sun, and K. P. Kim, "Comparison of missing data imputation methods in time series forecasting," *Comput. Mater. Contin.*, vol. 70, no. 1, pp. 767–779, 2021, doi: 10.32604/cmc.2022.019369.
[14] S. Bryzgalova, S. Lerner, M. Lettau, and M. Pelger, "Missing Financial Data," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4106794.
[15] S. M. Ribeiro and C. Leite De Castro, "Time Series Imputation by Nature and by Decomposition," in *2021 IEEE Latin American Conference on Computational Intelligence, LA-CCI 2021*, 2021, pp. 1–6. doi: 10.1109/LA-CCI48322.2021.9769791.
[16] S. Behar and A. Sharma, "An Adaptive Model For Stock Market Forecasting Using Modified Genetic Algorithm," vol. 18, no. 5, pp. 3943–3954, 2021.
[17] D. Sciences, M. Hendrick, and A. Stam, "Asset Pricing and The Applications of Machine Learning in Missing Data Treatment," *D. Sci.*, no. June, Jun. 2022.
[18] R. Qin and Y. Wang, "ImputeGAN: Generative Adversarial Network for Multivariate Time Series Imputation," *Entropy*, vol. 25, no. 1, 2023, doi: 10.3390/e25010137.
[19] A. Y. Areiqat, A. Abu-Rumman, Y. S. Al-Alani, and A. Alhorani, "Impact of behavioral finance on stock investment decisions applied study on a sample of investors at Amman Stock Exchange," *Acad. Account. Financ. Stud. J.*, vol. 23, no. 2, pp. 1–17, 2019.
[20] "AMMAN Stock Exchange." Accessed: May 01, 2024. [Online]. Available: https://www.ase.com.jo/en
[21] A. F. of Exchanges, "AFE - Arab Federation of Exchanges." 2022. [Online]. Available: www.arab-exchanges.org
[22] world-exchanges, "world-exchanges." 2022. [Online]. Available: www.world-exchanges.org
[23] O. R. Adegboye and E. Deniz Ülker, "Hybrid artificial electric field employing cuckoo search algorithm with refraction learning for engineering optimization problems," *Sci. Rep.*, vol. 13, no. 1, p. 4098, 2023, doi: 10.1038/s41598-023-31081-1.
[24] A. T. Abbas, A. A. Al-Abduljabbar, M. M. El Rayes, F. Benyahia, I. H. Abdelgaliel, and A. Elkaseer, "Multi-Objective Optimization of Performance Indicators in Turning of AISI 1045 under Dry Cutting Conditions," *Metals (Basel).*, vol. 13, no. 1, p. 96, 2023, doi: 10.3390/met13010096.
[25] A. T. Karadeniz, Y. Çelik, and E. Başaran, "Classification of walnut varieties obtained from walnut leaf images by the recommended residual block based CNN model," *Eur. Food Res. Technol.*, vol. 249, no. 3, pp. 727–738, 2023, doi: 10.1007/s00217-022-04168-8.
[26] K. S. M. Li *et al.*, "Wafer Defect Pattern Labeling and Recognition Using Semi-Supervised Learning," *IEEE Trans. Semicond. Manuf.*, vol. 35, no. 2, pp. 291–299, 2022, doi: 10.1109/TSM.2022.3159246.
[27] T. M. Pham, N. Pandis, and I. R. White, "Missing data, part 2. Missing data mechanisms: Missing completely at random, missing at random, missing not at random, and why they matter," *Am. J. Orthod. Dentofac. Orthop.*, vol. 162, no. 1, pp. 138–139, 2022, doi: 10.1016/j.ajodo.2022.04.001.
[28] S. Hong and H. S. Lynn, "Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction," *BMC Med. Res. Methodol.*, vol. 20, no. 1, pp. 1–12, 2020, doi: 10.1186/s12874-020-01080-1.
[29] A. Pathak, S. Batra, and H. Chaudhary, "Imputing Missing Data in Electronic Health Records," in *Lecture Notes in Electrical Engineering*, 2022, pp. 621–628. doi: 10.1007/978-981-19-2828-4_55.
[30] Y. Zhang and P. J. Thorburn, "Handling missing data in near real-time environmental monitoring: A system and a review of selected methods," *Futur. Gener. Comput. Syst.*, vol. 128, pp. 63–72, 2022, doi: 10.1016/j.future.2021.09.033.
[31] Y. Luo, "Evaluating the state of the art in missing data imputation for clinical data," *Brief. Bioinform.*, vol. 23, no. 1, p. bbab489, 2022, doi: 10.1093/bib/bbab489.
[32] M. Liu *et al.*, "Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques," *Artif. Intell. Med.*, vol. 142, p. 102587, 2023, doi: 10.1016/j.artmed.2023.102587.
[33] C. F. Tsai and Y. H. Hu, "Empirical comparison of supervised learning techniques for missing value imputation," *Knowl. Inf. Syst.*, vol. 64, no. 4, pp. 1047–1075, 2022, doi: 10.1007/s10115-022-01661-0.
[34] A. Liguori, R. Markovic, M. Ferrando, J. Frisch, F. Causone, and C. van Treeck, "Augmenting energy time-series for data-efficient imputation of missing values," *Appl. Energy*, vol. 334, p. 120701, 2023, doi: 10.1016/j.apenergy.2023.120701.
[35] Y. Liu, Z. Zhang, and S. Qin, "Deep Imputation-Prediction Networks for Health Risk Prediction using Electronic Health Records," in *Proceedings of the International Joint Conference on Neural Networks*, 2023, pp. 1–9. doi: 10.1109/IJCNN54540.2023.10191793.
[36] J. Ma, J. C. P. Cheng, F. Jiang, W. Chen, M. Wang, and C. Zhai, "A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data," *Energy Build.*, vol. 216, 2020, doi: 10.1016/j.enbuild.2020.109941.
[37] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," *Fractal Fract.*, vol. 7, no. 2, p. 203, 2023, doi: 10.3390/fractalfract7020203.
[38] I. M. Sumertajaya, E. Rohaeti, A. H. Wigena, and K. Sadik, "Vector Autoregressive-Moving Average Imputation Algorithm for Handling Missing Data in Multivariate Time Series," *IAENG Int. J. Comput. Sci.*, vol. 50, no. 2, pp. 727–735, 2023.
[39] H. Wu, S. Li, W. Shi, and S. Du, "FUSAIN: Combining Functional Dependencies and Clustering for Missing Values Imputation," *Eng. Lett.*, vol. 30, no. 2, pp. 513–521, 2022.