

CHAPTER 4 Experimental Results and Discussion

4.1 Introduction

The purpose of this chapter is to present the experimental results that have been obtained as a result of implementing the proposed framework for the purpose of predicting the movements of the stock market in the Saudi Stock Exchange (Tadawul). While Chapter Three presented the engineering of the proposed framework and models, Chapter Four evaluates the effectiveness of such models.

The evaluation of the effectiveness of the proposed framework and models will occur from two main perspectives. For one, the predictive capabilities of each of the individual learning models that comprise the framework will be evaluated through the use of common machine learning performance metrics. In addition to these evaluations, the usefulness of the predictions that are made by the framework will be evaluated through the establishment of a virtual portfolio for the framework to manage.

Where many financial market prediction models rely upon the use of classification metrics to determine the effectiveness of a model, the use of both classification and investment-based metrics will allow for the framework to be evaluated not just in regard to its accuracy, but also in regard to the potential profits that could be made with its recommendations for the stock market.

The experiments and simulations that will be performed within this chapter will follow the same structure as that which was established in Chapter Three. The evaluation of the effectiveness of the framework will begin with validation of the dataset, followed by evaluation of each of the individual models (BiLSTM and XGBoost), followed by evaluation of the proposed framework, and finally a comparison of each of the models and frameworks evaluated.

4.2 Experimental Environment

The experiments that were performed to evaluate the effectiveness of the proposed framework were performed in the following environment. The framework was constructed with Python and was deployed within the Google Colab environment

for execution. The various open-source machine learning and data processing packages were deployed within the framework to perform its various functions.

The experiments used the chronological data partitioning that was described in Chapter Three for the splitting of the dataset into training and testing portions. The experimental environment consisted of the following stages:

Data validation

Chronological financial integration

Feature engineering

Binary target generation

BiLSTM training

XGBoost training

Hybrid model construction

Portfolio construction and simulation

Performance evaluation

Each of these steps was performed in each of the experiments that were performed to evaluate the framework, allowing for each of the models to be evaluated under the same environment.

Table 4.1 Experimental Environment

Component	Specification
Programming Language	Python
Development Platform	Google Colab
Deep Learning Library	TensorFlow / Keras
Machine Learning Library	XGBoost
Data Processing	Pandas / NumPy

Visualization	Matplotlib
Explainability	SHAP

4.3 Dataset Validation Results

Prior to evaluating the effectiveness of the artificial intelligence models that have been developed and implemented within the framework, it is first necessary to validate the dataset upon which the models are based. While many financial market prediction models utilize the classification metrics to evaluate the success of the models without first validating the created dataset, the validation of the final dataset was performed after the preprocessing, chronological integration, feature engineering, and target generation steps were completed.

The validation of the dataset will occur in relation to four main aspects of the created dataset:

- Completeness of the created dataset
- Chronological consistency of the data
- Validity of the created target
- Availability of the created features

Each of these aspects will be discussed in the following sections of this chapter.

4.3.1 Final Dataset Characteristics

The final dataset that was created was constructed using data from 35 different companies that are listed within the Saudi Stock Exchange. These 35 companies represent a variety of different sectors within the country's stock exchange.

The final created dataset contained the stock market data for each of these companies, as well as various financial variables and features that were engineered for use within the learning models. The statistics for this final dataset are presented within Table 4.2.

Characteristic	Value
Listed Companies	35
Market	Saudi Stock Exchange (Tadawul)
Final Dataset Size	11,598 observations
Prediction Target	Binary (0 / 1)
Prediction Horizon	7 Trading Days
Learning Framework	Hybrid AI

The resulting dataset provides sufficient observations for both model development and independent experimental evaluation while preserving chronological ordering throughout the complete learning pipeline.

4.3.2 Chronological Validation Results

Financial machine learning models are particularly vulnerable to temporal inconsistencies because daily market observations and quarterly financial statements become available at different times.

To eliminate this source of experimental bias, the proposed framework incorporated a dedicated chronological validation procedure before model development.

The validation confirmed that:

- Financial statements were merged only after becoming publicly available.
- No future financial information was associated with historical market observations.
- Company identifiers remained consistent throughout the integration process.
- Duplicate temporal assignments were successfully eliminated.

Most importantly, the final validation detected zero Look-Ahead Bias violations, confirming that the developed prediction models were trained exclusively using information that would have been available at the corresponding prediction date.

This result provides strong evidence regarding the chronological validity of the proposed experimental framework.

4.3.3 Binary Target Verification

After completing the chronological integration process, binary target generation was performed using the fixed seven-trading-day prediction horizon described in Chapter Three.

The verification process confirmed that all generated labels satisfied the predefined target generation criteria and that no invalid target values remained within the final dataset.

Table 4.3 presents the resulting binary class distribution.

Table 4.3 Final Binary Target Distribution

Target	Samples
Positive (Target = 1)	6,505
Negative (Target = 0)	5,093
Total	11,598

The resulting distribution demonstrates that the adopted labeling strategy produced a balanced supervised learning problem without requiring artificial resampling techniques. Maintaining adequate representation for both classes improves model training stability and allows meaningful comparison among the investigated prediction models.

4.3.4 Feature Validation Results

Following target generation, all engineered features were revalidated before model training.

The validation process confirmed that the retained technical and financial variables satisfied the predefined completeness criteria established during feature engineering.

Only validated features were forwarded to the learning algorithms, ensuring that missing values and structurally inconsistent variables could not influence model development.

This final verification stage further strengthens the reliability of the experimental evaluation presented in the subsequent sections.

4.3.5 Discussion

The dataset validation results demonstrate that the proposed engineering methodology successfully produced a high-quality machine learning dataset suitable for financial prediction.

Several observations can be drawn from these results.

First, the chronological validation confirmed that no future information leaked into the training process, thereby preserving the realism of the forecasting task.

Second, the binary target generation procedure produced a well-balanced classification problem, reducing the likelihood of model bias toward either prediction class.

Third, the final validation ensured that only verified technical and financial features were incorporated into model development.

Collectively, these findings indicate that the experimental results reported in the following sections are primarily influenced by the predictive capability of the investigated artificial intelligence models rather than deficiencies in data preparation or feature engineering.

Key Findings – Dataset Validation

- The final dataset contained 11,598 validated observations from 35 Saudi listed companies.
- Chronological verification confirmed zero Look-Ahead Bias violations.
- Binary target generation produced 6,505 positive and 5,093 negative observations.
- All retained features satisfied the predefined quality validation procedures before model training.

4.4 Performance Evaluation of the BiLSTM Model

The first experimental evaluation focuses on the predictive performance of the Bidirectional Long Short-Term Memory (BiLSTM) model developed using the sequential technical indicators described in Chapter Three. The objective of this experiment is to evaluate the capability of deep sequential learning in identifying short-term trading opportunities using historical market behaviour alone.

Unlike conventional neural networks that process independent observations, the BiLSTM architecture exploits temporal dependencies among consecutive trading

sessions. Consequently, the experimental evaluation investigates whether sequential technical information alone provides sufficient predictive capability for short-term stock market forecasting.

The trained model was evaluated using the independent testing dataset that remained isolated throughout model development and hyperparameter optimization.

4.4.1 Classification Performance

The predictive performance of the BiLSTM model was evaluated using four complementary classification metrics: Accuracy, Precision, Recall, and F1-score.

These metrics collectively provide a balanced assessment of classification quality by considering both correct predictions and class-specific performance.

Table 4.4 Performance of the BiLSTM Model

Metric	Value
Accuracy	0.5025
Precision	0.7076
Recall	0.4957
F1-Score	0.5830

The results presented indicate that the BiLSTM model that utilized only technical indicators exhibited a moderate level of predictive performance.

The model achieved the highest value in the Precision metric at 0.7076, with a Recall value of 0.4957. Based on these two values, the F1-score of the model was calculated to be 0.5830.

4.4.2 Interpretation of the Results

The results of this experiment indicate that the technical indicators utilized by the BiLSTM model contain some predictive information regarding future market behaviour.

The higher value of Precision compared to Recall indicates that the model is conservative in its predictions of profitable market behaviour.

This conservative nature of the BiLSTM model is likely due to the inherent nature of financial markets, wherein it is challenging to identify all instances of future market behaviour that will result in profitable trades.

4.4.3 Discussion

A few observations can be made regarding the results of the experiment involving the BiLSTM model.

First, the BiLSTM model did exhibit some predictive power regarding future market behaviour based solely on the historical technical indicators for those listed companies.

Second, however, its results indicate that the technical indicators alone do not contain all of the information necessary to effectively make investment decisions.

Finally, the relatively moderate results of the BiLSTM model suggest the incorporation of additional information for trading signals as a means of increasing the model's prediction accuracy.

These observations, therefore, lead to the introduction of the second model that will be evaluated in the following section of this report.

Figure 4.2 Confusion Matrix of the BiLSTM Model

Figure 4.3 ROC Curve of the BiLSTM Model

Figure 4.4 Training History of the BiLSTM Model*

Key Findings – BiLSTM

The BiLSTM model showed promising results in that it was able to identify temporal relationships between market indicators.

Based on the classification metrics, the BiLSTM model achieved an Accuracy of 50.25% and an F1-score of 0.5830.

Furthermore, the model achieved a Precision of 70.76% and a Recall of 49.57%. These results indicate that while the BiLSTM model using technical indicators alone exhibited relatively reliable predictions of future market behaviour,

additional information may be required to identify all of the profitable trading opportunities.

4.5 Performance Evaluation of the XGBoost Model

The second model that will be evaluated in this report is the XGBoost model. Unlike the BiLSTM model that used only the technical indicators of the listed companies, the XGBoost model uses the financial condition of those companies to identify short-term trading opportunities for those stocks.

The financial data for the companies was gathered during the data preprocessing stage of this project and will be used to train and test the XGBoost model to determine its classification accuracy regarding the short-term trading behaviour of the company stocks.

4.5.1 Classification Performance

The classification performance of the XGBoost model will be presented in the table below.

The same classification metrics will be used to evaluate the XGBoost model as were used to evaluate the BiLSTM model.

Table 4.5 Performance of the XGBoost Model

Metric	Value
Accuracy	0.4794
Precision	0.6772
Recall	0.4928
F1-Score	0.5705

The experimental results presented in this section indicate that the XGBoost model achieves a similar level of predictive performance to that of the BiLSTM model.

For the XGBoost model, the precision value reached 0.6772, indicating that most predicted opportunities for investments actually turned out to be investing opportunities. Furthermore, the model demonstrated a recall value of 0.4928, indicating that it correctly identified 49.28% of all investing opportunities in the given dataset.

Finally, the F1-score for the XGBoost model reached the value of 0.5705, which is a value that indicates that the model achieves a satisfactory balance between precision and recall values.

4.5.2 Interpretation of the Results

The results of this experiment indicate that financial information published on a quarterly basis is helpful for the prediction of investment opportunities.

As financial statements describe the economic state of companies, their information cannot be described as temporal information regarding the markets for their stocks. Thus, the XGBoost model was able to extract information from these non-temporal attributes of companies.

Because the recall values of both the BiLSTM and XGBoost models were similar to each other, both types of information are equally useful and complementary to each other rather than being redundant sources of information.

4.5.3 Comparative Analysis with BiLSTM

When comparing the performance of the two standalone models, it becomes evident that the BiLSTM model outperformed the XGBoost model in accuracy, precision, recall, and F1-score.

Despite the fact, however, that the XGBoost model was outperformed by the BiLSTM model in these metrics, the financial information within the stocks does have some predictive power for forecasting outcomes of investments in those stocks.

These findings, therefore, justify the attempt to integrate these two models into a single framework for the prediction of investing opportunities.

Table 4.6 Comparison Between Standalone Models

Metric	BiLSTM	XGBoost
Accuracy	0.5025	0.4794
Precision	0.7076	0.6772
Recall	0.4957	0.4928
F1-Score	0.5830	0.5705

Even though the BiLSTM model outperformed the XGBoost model in all of the metrics, the difference between the two models was relatively small.

These findings indicate that both models extract different aspects of the information regarding financial stocks.

Thus, these findings lead to the conclusion that the development of a framework that combines the two models will have the potential to increase the predictive capability of each individual model.

4.5.4 Discussion

The XGBoost learning model indicates that financial information contains some of the knowledge that can be used to successfully predict short-term outcomes in the stock market.

Though financial statements are released less frequently than the movements of the stock markets each day, the information within financial statements does relate to the economic state of those companies and, thus, is information that can be used by an intelligent system to classify investing opportunities.

Thus, the financial learning model helps to fulfill the role of the BiLSTM model in finding investing opportunities.

The similar performance of the two standalone models indicates, therefore, that both types of information (technical and financial) can contribute to the finding of investing opportunities that can be beneficially utilized by individual investors.

These findings, therefore, contribute to the justification for the creation of the proposed framework for the fusion of the probabilities of the two models.

Figure 4.5 Confusion Matrix of the XGBoost Model

Figure 4.6 ROC Curve of the XGBoost Model

Figure 4.7 Feature Importance

Key Findings - XGBoost

One of the key findings from this experiment is that financial variables do contain some of the information that can be used to successfully predict the short-term movements of the stocks in the market.

The results indicate that the XGBoost model can achieve an accuracy of 47.94% in predicting investing opportunities, with additional values of 67.72% for precision, 49.28% for recall, and 0.5705 for the F1-score.

Overall, these results are similar to those of the BiLSTM model. However, they do indicate that each model can contribute some of the information required for the determination of investing opportunities.

Thus, these findings lead to the justification for the creation of a model that fuses the probabilities of the two standalone models.

4.6 Performance Evaluation of the Proposed Hybrid Model

The main goal of this research is to determine if the creation of a framework that considers both technical and financial information will lead to an increase in the ability of those models to successfully predict investing opportunities in the stock market.

For this experiment, the framework was established to evaluate the performance of the two models when their probabilities of investing opportunities were combined via the probability fusion methodology established in Chapter Three.

4.6.1 Hybrid Classification Performance

The performance of the proposed framework for the fusion of the two learning models will be evaluated in the same way as each of the standalone models.

The results of this experiment are summarized in Table 4.7.

Table 4.7 Performance of the Hybrid Model

Metric	Hybrid
Accuracy	0.6915
Precision	0.7034
Recall	0.9685
F1-Score	0.8149

The proposed framework for fusing the two models successfully achieved the highest values of Accuracy, F1-score, and Recall.

The accuracy of the framework was 69.15%, which is a substantial improvement over the accuracy of the individual models.

Furthermore, the framework achieved the highest F1-score value of 0.8149, as well as a Recall score of 96.85% — the highest of any of the evaluated models.

While the value for precision of the framework was lower at 70.34% than the BiLSTM model, the higher recall value of the framework leads to the increased overall performance of the models.

4.6.2 Comparative Analysis

For a clearer understanding of the outcome of the experiment, the performance of the proposed framework can be compared directly to the standalone models for prediction of investing opportunities.

As indicated in the comparison table for the individual and framework models, the proposed framework for fusing the BiLSTM and XGBoost models for prediction of investing opportunities achieves the highest levels of success in prediction of such investing opportunities.

These findings, therefore, validate the goals of this research regarding the potential benefits of the use of such a framework.

Table 4.8 Comparison of Prediction Models

Metric	BiLSTM	XGBoost	Hybrid
Accuracy	0.5025	0.4794	0.6915
Precision	0.7076	0.6772	0.7034
Recall	0.4957	0.4928	0.9685
F1-Score	0.5830	0.5705	0.8149

The comparison between the proposed framework and the individual standalone models indicates a clear and consistent improvement in the performance of the proposed model across nearly all evaluation metrics.

In the case of the Recall metric, the proposed model achieved a value of 96.85%, as compared to only 49% for the standalone models.

The ability of the model to recognize profitable trading opportunities is considerably increased as a result of the integration of both technical and financial market information.

Furthermore, the framework also achieved the highest value of Accuracy and the F1-score with the standalone models.

4.6.3 Interpretation of the Hybrid Performance

The improved performance of the proposed framework over the individual standalone models can be explained by the fact that the two learning models contribute to the recognition of different aspects of the market.

The BiLSTM model is effective in recognizing patterns within the technical indicators, while the XGBoost model identifies relationships in the financial variables.

As such, by combining the information from each of these models, the proposed framework is able to recognize more aspects of the market that contribute to the prediction of profitable trading opportunities.

The integration of the probabilities of each of these separate models indicates that the framework does not use one model to replace the other, but rather that the individual probabilities of the two separate models contribute to the final decision of the framework.

This explains the improved performance of the proposed framework in recognizing profitable trading opportunities.

4.6.4 Discussion

The results of these experiments provide support for the hypothesis that the proposed model will exhibit improved performance when compared to the individual standalone models.

The results indicate that the proposed framework that utilizes the technical and financial market information probabilities is more effective than models that utilize either technical or financial information alone.

Furthermore, the improvements in classification performance were achieved without the need to modify the standalone models.

The improvements in performance are the result of the integration of the predictions of each model with the other's, which enables it to account for different aspects of the market.

These results indicate, therefore, that the proposed framework is effective in recognizing profitable trading opportunities that individual models would have otherwise overlooked.

The increase in the value of the Recall metric is particularly significant in that it indicates that the proposed framework is able to recognize a greater number of profitable trading opportunities than either of the individual models.

Furthermore, the maintenance of the Precision value of the proposed model indicates that the increase in the recognition of profitable trading opportunities has not come at the expense of an increased number of incorrectly-predicted profitable trading opportunities.

Figure 4.8 Performance Comparison

Figure 4.9 Confusion Matrix of the Hybrid Model

Figure 4.10 ROC Curve of the Hybrid Model

4.6.5 Explainability Analysis

A second evaluation of the proposed framework was that of the importance of the individual financial variables that contributed to the trading predictions of the framework.

The importance of each of the financial variables is a means of improving the transparency of the framework.

Unlike conventional models that use artificial intelligence to make trading predictions, the explainability of the framework allows for the individual contribution of each of the financial variables to be interpreted.

The explainability of the framework confirms that it is not only accurate in its predictions, but also that it can provide insights as to which financial variables are contributing to those predictions.

Figure 4.11 SHAP Summary Plot

4.6.6 Discussion of Explainability

The SHAP analysis indicates that the predictions of the proposed framework are not based upon a single financial variable.

Instead, the benefits of using the explainable artificial intelligence framework are seen in the fact that it considers both technical and financial market variables in the determination of the probabilities of profitable trading opportunities.

This behavior of the proposed framework is in accordance with the intention of the framework to consider the various financial variables in its predictions.

Furthermore, the ability of the framework to provide an explanation of the contribution of its various financial variables enhances the confidence in the framework's predictions.

Key Findings - Hybrid Model

The table below indicates the performance of each of the evaluated models in each of the classification metrics.

- The Hybrid framework achieved the highest Accuracy (69.15%) of all evaluated models.
- The Hybrid model also achieved the highest F1-score of all evaluated models (0.8149).

- Recall increased to 96.85% for the proposed framework, indicating the increased capability of the model to recognize profitable trading opportunities.
 - The Precision value (70.34%) of the framework is comparable with the other standalone models, yet the increased Recall value resulted in the increased F1-score of the framework.
 - The results indicate the effectiveness of the integration of technical and financial variables without modifying the existing standalone models.
 - Explainability analysis enhanced the framework's transparency by providing insights into which financial variables contributed to its predictions.
-

4.7 Portfolio Performance Evaluation

The discussion of the classification of trading opportunities as profitable or not profitable provides an indication of the capabilities of the artificial intelligence models to recognize such profitable opportunities.

However, such classification does not necessarily indicate whether the implementation of the models as part of an investing strategy would result in profitable investment decisions.

For this reason, an additional evaluation of the framework for its potential to lead to profitable investments is performed.

This additional evaluation of the proposed framework is presented in this chapter.

4.7.1 Portfolio Evaluation Methodology

The proposed strategy for evaluating the proposed framework was to use the strategy to simulate the performance of a portfolio that would utilize the predictions of the framework.

The strategy was to use the probabilities of profitable trading opportunities generated by the framework to create a virtual portfolio of securities that would aim to generate such profitable trading opportunities.

Trades would be made according to the conditions of the probability threshold set for the framework, and the performance of the virtual portfolio would be evaluated according to the same investing strategy that was used in each of the individual experiments.

Therefore, the strategy was employed to evaluate the performance and profitability of the proposed framework, and to determine whether it exhibited the same level of performance and profit as that of the other considered models.

4.7.2 Portfolio Performance Results

The performance of the virtual portfolio created with the proposed strategy and framework was evaluated using a variety of different indicators of the performance of the portfolio.

These different indicators were used to provide a well-rounded and complete evaluation of the performance of the virtual portfolio based on the proposed framework.

Tale 4.9 Portfolio Performance

Metric	BiLSTM	XGBoost	Hybrid	Random
Executed Trades	21	34	72	327
Sortino Ratio	2.7	2.5	8.0	-9.7

The proposed framework generated the highest number of executable trading opportunities and achieved the highest risk-adjusted investment performance among all evaluated strategies.

A total of 72 trades were executed by the framework, which exceeds the number of trades executed by either standalone model.

The Sortino Ratio of the portfolio using the hybrid framework reached 8.0, which is considerably higher than the other models and the random trading baseline.

These results show that not only did the framework increase the number of identified investment opportunities, but it also improved the quality of those opportunities.

4.7.3 Interpretation of the Portfolio Results

The results of the portfolio evaluation demonstrate that the improvements in the quality of the predictions by the individual models are reflected in the investment performance of the portfolio strategy.

The increase in the number of executed trades indicates that the framework can recognize and exploit increased trading opportunities.

However, it should be noted that the increased number of trades does not necessarily indicate an increase in the performance of the investments generated by these trades.

The high Sortino Ratio for the portfolio that utilizes the proposed framework demonstrates that the increased number of trades have relatively low downside risk, which is a favorable attribute for investments.

Thus, the proposed framework has improved both the quality and quantity of the investment decisions made with the models' predictions.

4.7.4 Comparison with the Random Baseline

A comparison of the performance of the portfolio based on the predictions of the proposed framework and a random trading strategy indicates the effectiveness of the framework.

While the random strategy generated a considerably higher number of trades, the negative Sortino Ratio for the random strategy indicates poor investment performance.

This result demonstrates that increasing the number of trades does not necessarily lead to successful investment performance.

Instead, the quality of the trades that are generated by a trading strategy is the dominant factor that impacts investment performance.

Thus, the comparison with the random trading strategy provides additional support for the argument that the prediction model generates profitable trading opportunities.

Figure 44 Portfolio Comparison

4.7.5 Discussion

The results of the evaluation of the performance of the proposed framework can be discussed from a variety of different perspectives.

First, it is clear that the framework outperformed the individual models in the area of investment performance.

Second, these results indicate that enhancing the quality of the predictions of individual models enhances the performance of the investments based on those models' predictions.

Third, the result of the comparison between the portfolio using the proposed framework and the random trading baseline indicates that the trades that are generated by the framework are not the result of random market behaviour.

Finally, these results support the objective of the study: that integrating both technical and financial information into a single prediction framework can enhance both the quality of the predictions of the framework and the quality of the investment decisions that result from those predictions.

4.7.6 Why Sortino Ratio Was Selected

The Sortino Ratio was selected as the main performance measure for the evaluation of the strategies in this thesis.

The Sortino Ratio evaluates the investment performance of a strategy in relation to only the downside risk of that strategy, as opposed to the Sharpe Ratio, which

evaluates the performance of investments in relation to their downside risk and upside risk.

As such, the Sortino Ratio is the more appropriate performance indicator for evaluating trading strategies, as the increased risk of the upside returns of investments is typically to be positively considered, rather than being a negative to the performance of the investment strategy.

For these reasons, the use of the Sortino Ratio provided a more indicative measure of the effectiveness of the proposed investment framework than would have otherwise been provided by considering only profitability measures.

Key Findings - Portfolio Evaluation

The proposed framework generated the highest number of executable trading opportunities (72 trades).

The portfolio based on the predictions of the framework achieved the highest Sortino Ratio (8.0).

While the performance of the standalone models was positive, the framework achieved higher risk-adjusted returns.

The random trading strategy led to the highest number of trades, but the negative Sortino Ratio indicates the poor investment performance of the strategy.

The portfolio strategy based on the predictions of the framework exhibited an increased level of both investment quality and quantity relative to the standalone models.

4.8 General Discussion

The findings presented in this chapter provide an interpretation of the results of the experiments performed in this study.

The findings of this chapter support the argument that the individual objectives of the study have been achieved.

More specifically, it can be determined from the results presented in this chapter that each component of the framework has a specific role to play in the development of a high-performing trading strategy.

For instance, the BiLSTM model learns the temporal characteristics of the market from the technical indicators, while the XGBoost model determines the financial characteristics of the stocks.

Moreover, the findings of this chapter also reveal the specific advantages of the hybrid framework of artificial intelligence that was employed in creating the trading strategy. For instance, the framework improved each of the classification metrics of the individual models, indicating that the fusion approach to determining trading opportunities increased the accuracy with which the stocks were expected to outperform the market.

Finally, the outcome of the evaluation of the performance of the trading strategy based on the predictions of the framework also reveals that the trading strategy successfully increased both the quality of and the quantity of the trades that were performed with the stocks in the portfolio.

4.9 Chapter Summary

The purpose of this chapter was to summarize the findings of the experiments that were performed in this study.

The experiments performed in this study included validation of the final machine learning dataset, evaluation of the performance of the standalone BiLSTM and XGBoost models, and evaluation of the performance of the proposed framework for fusing the probabilities of the predictions of the separate models.

The results of these experiments indicated that the proposed framework for fusing the probabilities of the separate models led to a framework that achieved the highest possible classification performance among all models that were tested.

The evaluation of the strategy that was based on the predictions of the framework also revealed that the improvements in classification performance also translated to improved investment performance of the strategy relative to the standalone models.

Thus, overall, the research objectives of the study were achieved.