A Hybrid Metaheuristic Optimization Framework for Multi-Scale Time Series Forecasting Using AO and PGRO

Hemasundara Reddy Lanka [1], Nagha Harish Vundavalli [2], Nagaraju Devarakondai [3], and Sarvani Anandarao [4,*]

[1] Technology Department, Publicis Sapient, Minneapolis, USA
[2] Technology Department, Strategic Education Inc, Minneapolis, USA
[3] School of Computer Science and Engineering, VIT-AP University, Amaravathi, India
[4] Department of Computer Science and Engineering, SRM University-AP, Amaravathi, India
Email: hemasundarareddy.lanka@publicissapient.com (H.R.L.); pvundavalli@gmail.com (N.H.V.); dnagaraj_dnr@yahoo.co.in (N.D.); sarvani.anandarao@gmail.com (S.A.)
*Corresponding Author

Abstract*—Time series forecasting plays a pivotal role in domains ranging from finance and healthcare to multimodal analytics, yet achieving robust accuracy across heterogeneous datasets and varying prediction horizons remains a significant challenge. In this paper, we propose a novel hybrid forecasting framework that integrates advanced metaheuristic optimization with deep learning architectures for multi-domain, multi-scale prediction. The core of our approach is a dual-stage optimization strategy that combines the global exploration capability of the Aquila Optimizer (AO) with the guided local refinement of the Proposed Guided Remora Optimization Algorithm (PGRO). This hybrid optimization mechanism is applied to fine-tune model hyperparameters, fusion strategies, and incremental learning settings for architectures such as Long Short Term Memory (LSTM), Transformer, and multimodal fusion networks. The framework is further enhanced by an incremental learning module with memory replay, enabling continual adaptation without catastrophic forgetting. We evaluate the proposed approach on 6 diverse datasets (National Association of Securities Dealers Automated Quotations/New York Stock Exchange (NASDAQ/NYSE) Historical Stock, Financial Question Answering (FiQA), Financial PhraseBank, Reddit Financial News, Fourth Makridakis Forecasting Competition Time series (M4), and Video and Text (VaTeX) Multimodel), spanning structured numerical, textual, social media, and multimodal time series. Experimental results show that our AO + PGRO-enhanced models consistently outperform classical statistical models, deep learning baselines, and state-of-the-art transformer-based forecasters in both short- and long-horizon prediction tasks. Ablation studies confirm the contribution of each component, and detailed horizon-wise analysis highlights the method's adaptability to varying temporal dependencies. The proposed framework delivers state-of-the-art accuracy, stability, and generalization across domains, offering a powerful and extensible approach for real-world time series forecasting challenges.

***Keywords*—hybrid time series forecasting, metaheuristic optimization, aquila optimizer, guided remora optimization algorithm, incremental learning, multimodal prediction**

I. Introduction

Time series forecasting plays a crucial role in decision-making across many domains, including finance, energy management, transportation planning, public health monitoring, and infrastructure management [1]. Accurate forecasting enables organizations and policymakers to anticipate future trends, manage risks, and support strategic planning in complex and dynamic environments. In financial markets, forecasting models are particularly important for investment planning, risk management, and market analysis. However, achieving reliable predictions across different forecasting horizons remains challenging because real-world time-series data are highly dynamic, noisy, and often nonlinear [2]. Traditional statistical forecasting methods such as AutoRegressive Integrated Moving Average (ARIMA) [3] and exponential smoothing techniques [4] have been widely used due to their interpretability and well-established mathematical foundations. These approaches capture trends, seasonality, and autocorrelation structures effectively under relatively stable conditions, but their reliance on stationarity assumptions limits their ability to represent complex nonlinear relationships commonly observed in modern financial and economic systems [5]. To address these limitations, deep learning architectures such as Long Short-Term Memory networks [6], Temporal Convolutional Networks [7], and Transformer-based forecasting models [8] have been introduced to learn complex temporal dependencies directly from sequential data. Recent Transformer variants including Informer [9],

Manuscript received November 24, 2025; revised February 9, 2026; accepted March 23, 2026; published June 22, 2026.

Autoformer [10], and FEDformer [11] have demonstrated strong capability in modeling long-range dependencies in time-series data. Nevertheless, deep learning forecasting models often require extensive computational resources and careful hyperparameter configuration, and their performance may become unstable when training data are limited or highly variable [12]. Hybrid forecasting frameworks that combine statistical and deep learning approaches, such as Exponential Smoothing Recurrent Neural Networ (ES-RNN) [13] and Long- and Short-term Time-series network (LSTNet) [14], have been explored to improve predictive capability by leveraging complementary modeling strengths.

Despite these advancements, several critical challenges continue to limit the effectiveness of modern forecasting systems. First, forecasting environments increasingly involve heterogeneous data sources, including structured numerical time-series signals, textual sentiment information, and contextual or visual data [15]. Many existing models rely primarily on a single data modality or combine heterogeneous features through simple fusion strategies, which restricts the ability to capture meaningful cross-modal relationships. Second, financial and economic time-series data are inherently non-stationary, meaning that their statistical properties evolve over time due to macroeconomic events, policy changes, and shifts in investor sentiment. Models trained on historical data may therefore experience performance degradation when applied to newly emerging conditions. Third, the optimization of deep learning forecasting architectures is highly sensitive to hyperparameter configurations. Conventional gradient-based optimization and many existing metaheuristics approaches often prioritize minimizing training error rather than improving model generalization, which may lead to unstable learning behavior and overfitting when models are applied across datasets with different temporal characteristics. Addressing these challenges requires forecasting frameworks that can integrate heterogeneous information sources, adapt to evolving data distributions, and maintain stable optimization behavior during training.

To address these limitations, this study proposes a hybrid metaheuristic-driven forecasting framework that integrates deep learning architectures with structured optimization and adaptive learning mechanisms. The proposed framework combines the global exploration capability of the Aquila Optimizer (AO) [16] with the guided refinement mechanism of the Proposed Guided Remora Optimization (PGRO) algorithm [17]. In this framework, AO performs global exploration of the hyperparameter search space to identify promising model configurations, while PGRO performs validation-driven refinement through a memory-guided optimization strategy designed to prioritize solutions with stronger generalization capability. Unlike conventional hybrid approaches that apply a metaheuristic algorithm solely for parameter tuning, the proposed AO-PGRO collaboration follows a structured exploration–refinement strategy to improve optimization stability. In addition, multimodal feature fusion is integrated directly into the optimization process, enabling the model to dynamically balance temporal, textual, and contextual information during training. To further address the challenges of evolving time-series environments, incremental learning mechanisms based on Elastic Weight Consolidation (EWC) [18] and replay buffers are incorporated to allow the model to adapt continuously to new data while preserving previously learned knowledge. The main contributions of this study are summarized as follows.

- A generalization-aware hybrid forecasting framework that integrates Aquila Optimization (AO) [16] and Proposed Guided Remora Optimization (PGRO) [17] within a unified optimization pipeline for multi-scale time-series forecasting.
- A validation-driven memory-guided optimization mechanism in PGRO that evaluates candidate solutions using both training and validation performance to improve generalization stability and discourage overfitted configurations.
- A structured 2-stage optimization strategy in which AO performs global exploration of the hyperparameter space and PGRO performs guided refinement of promising configurations, enabling stable optimization of deep forecasting models.
- Integration of incremental learning mechanisms, including Elastic Weight Consolidation (EWC) [18] and replay buffers, to enable continuous adaptation to evolving time-series data while mitigating catastrophic forgetting.
- Comprehensive evaluation across multiple heterogeneous datasets spanning financial time series, sentiment signals, social media data, and multimodal sources, demonstrating improved forecasting accuracy and robustness compared with conventional forecasting and optimization approaches.

The remainder of this paper is organized as follows. Section II reviews related work in time-series forecasting, multimodal data integration, and optimization techniques. Section III presents the theoretical background underlying the proposed framework. Section IV describes the proposed AO-PGRO hybrid forecasting methodology. Section V presents the experimental setup, datasets, and evaluation results. In Section VI discussion, and Finally Section VII concludes the study and outlines directions for future research.

II. Related Work

Financial time-series forecasting has been widely investigated using statistical models, machine learning techniques, and deep learning architectures. Early approaches primarily relied on statistical methods and classical machine learning models to capture temporal patterns in market data. With the advancement of deep learning, researchers have increasingly explored neural architectures capable of modeling complex nonlinear dependencies present in financial time-series data. Transformer-based forecasting models such as the Informer-based time series forecasting framework [19] have demonstrated improved capability in capturing long-range temporal dependencies compared with traditional sequential architectures including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). Hybrid neural architectures have also been proposed to improve forecasting performance. For example, Convolution Neural Networks-Long Short-Term Memory (CNN-LSTM) models have been applied for portfolio optimization and stock price prediction tasks [20], while hybrid deep learning models incorporating Visual Geometry Group, 16 layers deep (VGG16)-based feature extraction have been used for gold volatility forecasting [21]. Comparative analyses of CNN, LSTM, and CNN-LSTM architectures have shown that deep learning models can effectively capture nonlinear relationships in financial data [22]. Additional research has explored graph-based neural models [23] such as the Temporal and Heterogeneous Graph Neural Network for stock prediction [24], hierarchical deep learning architectures [25], motif-based convolutional approaches for stock forecasting [26], and transformer-based attention models incorporating social media signals [27]. Explainable AI techniques have also been integrated into financial prediction systems to improve model transparency [28], and meta-learning approaches have been proposed to improve adaptability across multiple financial datasets [29]. Other forecasting frameworks include ensemble-based prediction models, machine learning–based financial risk prediction systems [30], and classification-based stock trend prediction models such as the Status Box framework [31]. Despite these advances, many forecasting methods still rely heavily on historical numerical price signals and often lack the ability to effectively integrate heterogeneous information sources or adapt to rapidly changing financial environments.

With the increasing availability of alternative data sources, several studies have investigated incorporating external information such as financial news, textual sentiment, and social media signals into forecasting models. Early work on public mood analysis demonstrated that collective sentiment extracted from social media platforms could influence market prediction accuracy [32]. Domain-specific financial sentiment dictionaries were introduced to improve financial text classification performance [33], while research on media pessimism examined the influence of negative financial news sentiment on market behavior [34]. Hybrid forecasting frameworks combining sentiment signals with technical indicators have also been explored. Sentiment-aware CNN-LSTM architectures have been applied to stock price prediction tasks [35], and volatility forecasting models integrating financial news with market indicators have shown improved predictive capability [36]. Transformer-based sentiment analysis approaches have further improved the extraction of contextual information from financial text data [37], and sentiment indices such as the Social Sentiment Index have been integrated into transformer-based financial prediction systems [38]. Multimodal sentiment analysis models such as Multichannel Cross-Modal Fusion Network (MCFNet) have explored combining textual, visual, and temporal information sources [39]. Other approaches include hybrid LSTM frameworks incorporating financial news [40], diffusion-based models integrating sentiment features [41], and attention-based BiLSTM architectures applied to trading strategy prediction [42]. Additional work has explored combining financial news with market indicators using self-attention models [43] or multi-source fusion frameworks such as novel multi-source information-fusion stock price prediction framework based on a hybrid deep neural network architecture (CNN and LSTM) [44]. Although these studies demonstrate the benefits of incorporating alternative data sources into forecasting models, many rely on static feature fusion strategies and lack adaptive mechanisms capable of dynamically adjusting model configurations in response to evolving market conditions.

Explainable artificial intelligence and optimization techniques have also been investigated to improve forecasting model transparency and performance. Model interpretation frameworks such as Local Interpretable Model-agnostic Explanations (LIME) [45] and SHapley Additive exPlanations (SHAP) [46] have been applied to explain predictions generated by complex machine learning models. Other research has explored interpretability frameworks for financial systems [47], explainable reinforcement learning for financial decision-making [48], and explainable deep learning approaches for macroeconomic forecasting tasks [49]. Additional work has applied deep learning models to financial text regression problems [50]. Comprehensive surveys have examined the use of artificial intelligence in stock market prediction [51], while influence function–based approaches have been proposed to analyze the contribution of training data to model predictions [52]. Meta-learning frameworks have also been explored to improve model adaptability across tasks and datasets [53], and evolutionary optimization techniques have been applied to financial forecasting models [54]. Recent studies have further investigated explainable AI in finance [55], neural network–based feature selection methods [56], hybrid optimization-based financial distress prediction systems [57], and reinforcement learning models for algorithmic trading [58]. In parallel, heuristic and metaheuristic optimization algorithms have been widely used for solving complex optimization problems due to their ability to explore high-dimensional search spaces [59]. However, deep learning forecasting models remain highly sensitive to hyperparameter configurations and training dynamics [60], and financial forecasting systems frequently suffer from instability and poor generalization under volatile market conditions [61]. Evolutionary algorithms and genetic algorithms have therefore been applied to optimize model parameters and feature representations [62, 63]. Nevertheless, many existing hybrid optimization frameworks focus primarily on improving convergence speed or training accuracy and rarely incorporate validation-driven optimization strategies or adaptive multimodal feature integration mechanisms.

Overall, existing forecasting approaches have achieved significant progress but still exhibit several limitations. Many models rely on either structured financial indicators or sentiment-based signals but seldom integrate heterogeneous financial information sources within a unified forecasting framework. Furthermore, conventional deep learning models often employ static feature fusion strategies and lack mechanisms for dynamic adaptation to evolving financial environments. In the optimization domain, traditional metaheuristic algorithms may experience premature convergence and unstable parameter tuning when applied to complex deep learning architectures. These limitations highlight the need for forecasting frameworks that combine multimodal financial data with structured optimization mechanisms capable of improving model generalization, stability, and adaptability in dynamic forecasting environments.

III. Background Theoretical Framework

Modern forecasting systems often require both effective feature representation and robust optimization strategies to handle complex, non-stationary data environments. In deep learning-based forecasting architectures, model performance is highly dependent on the quality of feature extraction and the effectiveness of hyperparameter optimization. While neural architectures provide strong capability for learning nonlinear temporal relationships, their performance can vary significantly depending on parameter initialization, training configuration, and dataset characteristics. Consequently, optimization algorithms play a critical role in guiding the learning process toward stable and generalizable solutions.

Metaheuristic optimization algorithms have been widely applied for tuning deep learning models due to their ability to explore high-dimensional search spaces without relying on gradient information. Among these approaches, the Aquila Optimizer (AO) represents a population-based optimization strategy inspired by the hunting behavior of Aquila (eagle) birds. AO alternates between exploration and exploitation phases to identify promising candidate solutions across the search space. Through adaptive search strategies, AO can effectively avoid poor initializations and explore diverse regions of the parameter space. This capability makes AO particularly suitable for complex optimization tasks where conventional gradient-based training may become trapped in local minima. However, although AO provides strong global exploration capability, its exploitation performance may weaken in later optimization stages, sometimes resulting in slower convergence or limited refinement of promising candidate solutions.

Another nature-inspired optimization approach is the Remora Optimization Algorithm (ROA), which models the cooperative behavior between remora fish and their host species. In optimization contexts, ROA simulates cooperative search behavior among candidate solutions to explore the search landscape. The algorithm attempts to balance exploration and exploitation by allowing candidate solutions to update their positions relative to promising regions of the search space. While ROA has demonstrated effectiveness in several optimization problems, it also exhibits certain limitations. In particular, its search process relies heavily on stochastic movements, which may lead to inefficient exploration in high-dimensional problems. In addition, ROA lacks strong directional guidance based on historical solution quality, which can result in slow convergence or premature stagnation in local optima when applied to complex optimization tasks such as deep learning hyperparameter tuning.

To address these limitations, this study introduces a Proposed Guided Remora Optimization (PGRO) mechanism that enhances the original ROA framework through guided search and validation-aware candidate selection. Instead of relying purely on random position updates, PGRO incorporates memory-based guidance that uses previously identified high-quality solutions to steer the search process toward more promising regions of the solution space. In addition, PGRO introduces a validation-driven evaluation mechanism that compares candidate solutions based on both training and validation performance. This approach discourages configurations that exhibit strong training accuracy but poor validation performance, thereby reducing the likelihood of selecting overfitted solutions during the optimization process.

In the proposed forecasting framework, AO and PGRO are integrated in a structured 2-stage optimization pipeline. AO performs the initial global exploration of the hyperparameter space to identify diverse candidate configurations and avoid premature convergence. PGRO then performs guided refinement of these candidate solutions by emphasizing generalization performance during the selection process. Through this role separation, AO provides broad search capability while PGRO focuses on improving solution stability and generalization behavior. The detailed implementation of the proposed AO-PGRO hybrid optimization strategy and its integration with the forecasting architecture are described in the methodology section.

IV. Proposed Methodology

The proposed forecasting framework integrates multimodal financial data with hybrid optimization and adaptive learning to generate robust multi-horizon predictions. The system processes heterogeneous data sources including structured financial time-series data, textual financial news and sentiment signals, and visual information such as trading charts. The proposed methodology is formally described in Algorithm 1, which outlines the novel hybrid framework combining the global exploration capability of the Aquila Optimizer (AO) with the local refinement strength of the GRO optimizer. The overall framework consists of 5 stages: problem formulation, multimodal preprocessing, feature extraction, hybrid AO-PGRO optimization, and incremental learning with temporal adaptation. The end-to-end architecture of the proposed system is illustrated in Fig. 1.

Algorithm 1: Novel Hybrid Framework with AO + GRO Optimizer

```
Input:
   D_raw ← Raw data from financial, social, and economic sources
   Models ← {BERT, GPT-3, ViT, TCT, CNN-LSTM, TFT}
   Parameters ← Initial hyperparameters (learning rate η, batch size B, etc.)
   AO_iters, PGRO_iters ← Iteration limits for AO and PGRO
   ReplayBuffer ← ∅
Output:
   M_final ← Trained and optimized forecasting model
   E_final ← Explanation of predictions
BEGIN
1. // Step 1: Data Collection and Preprocessing
  D_clean ← Preprocess(D_raw)
     - Remove noise, HTML, duplicates using regex
     - Handle missing values (impute/forward fill)
     - Normalize features (Min-Max Scaling)
     - Tokenize, lemmatize, and align timestamps
     - Augment data if needed (paraphrasing, synonyms)
2. // Step 2: Multimodal Feature Extraction
  Text_Embeddings ← BERT_Encode(D_clean.text)
  Long_Context_Embeddings ← GPT3_Encode(D_clean.text)
  Visual_Embeddings ← ViT_Extract(D_clean.images)
  Temporal_Embeddings ← TCT_Encode(D_clean.timestamps)
  Features ← Concatenate (Text_Embeddings, Long_Context_Embeddings, Visual_Embeddings, Temporal_Embeddings)
3. // Step 3: Optimization Phase 1 - Global Exploration (Aquila Optimizer)
  For i = 1 to AO_iters do:
     candidate_params[i] ← AO_GenerateCandidate(i)
     fitness[i] ← Evaluate (Models, Features, candidate_params[i])
  End For
  best_params_AO ← SelectBest(fitness)
4. // Step 4: Optimization Phase 2 - Guided Remora Optimization (PGRO)
  Initialize PGRO with best_params_AO
  For j = 1 to PGRO_iters do:
     new_params ← PGRO_Update(best_params_AO, memory)
     ratio ← FitnessRatio(new_params, best_params_AO)
     If 0.7 ≤ ratio < 1 then
        best_params_AO ← new_params
     End If
  End For
  M_trained ← TrainModel(Models, Features, best_params_AO)
5. // Step 5: Incremental Learning Framework
  Initialize ReplayBuffer with past data
  While new_data_batch arrives do:
     new_features ← ExtractFeatures(new_data_batch)
     ReplayData ← Sample (ReplayBuffer)
     combined_batch ← new_features ∪ ReplayData
     loss ← ComputeLoss(M_trained, combined_batch)
     η ← PGRO_AdjustLearningRate(loss, η)
     M_trained ← UpdateModel(M_trained, combined_batch, η)
     Apply EWC to preserve old knowledge
     ReplayBuffer ← UpdateBuffer(new_features)
     AdjustThresholds(precision, recall)
  End While
6. // Step 6: Prediction and Explainability
  Predictions ← M_trained.predict(test_data)
  E_final ← {
     SHAP_Values(Predictions),
     AttentionVisualization(M_trained),
     Counterfactuals (Predictions)
  }
7. Return M_trained, E_final
END
```

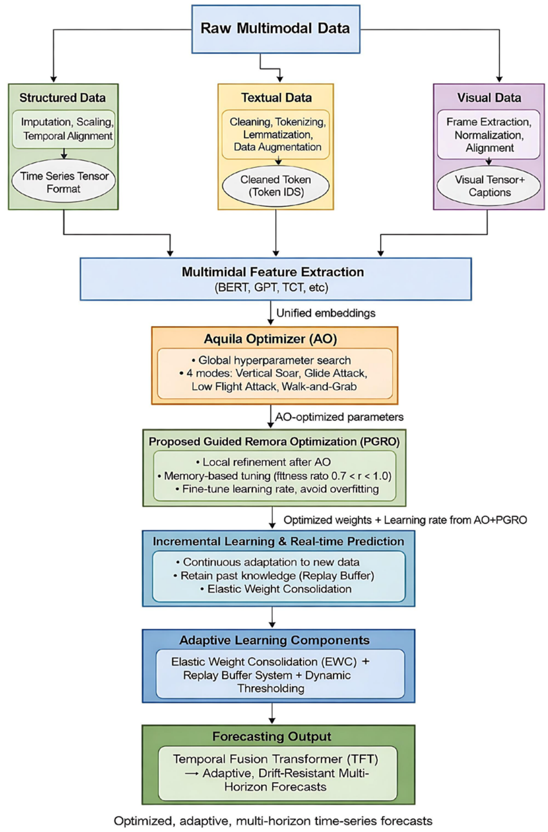

Fig. 1. Proposed adaptive multimodal time-series forecasting framework integrating AO-PGRO optimization with incremental learning and temporal adaptation.

A. Problem Formulation

The multimodal forecasting task is defined as predicting future financial market behavior by using multiple types of data collected over a historical time period. At each time step. At time step t the input to the model is represented as:

$$X_t = \{X_t^{num}, X_t^{txt}, X_t^{vis}\}$$

Which indicates that the model receives multimodal information from 3 different sources. Here, X_t^{num} represents numerical data such as stock prices, trading volume, and technical indicators; X_t^{txt} represents textual data such as financial news, social media posts, and market reports; and X_t^{vis} represents visual data such as stock charts, financial graphs, or related images. X_t denotes the multimodal input at time step t, which consists of heterogeneous features collected from numerical, textual, and visual sources. By combining these heterogeneous data sources at each time step, the model can learn relationships between numerical trends, textual sentiment, and visual patterns, which helps improve the accuracy and robustness of financial market forecasting.

In multimodal financial forecasting, the objective is to predict future market behavior using historical multimodal data such as numerical indicators, textual information, and visual data collected over a period of time. Financial markets are highly dynamic and non-stationary, meaning their patterns change over time due to economic events, news, and investor behavior. Therefore, instead of predicting future values using only the current data point, it is necessary to consider a sequence of past observations over a historical time window. This helps the model learn temporal patterns, trends, and relationships across different modalities. To mathematically represent this forecasting task, the prediction problem is formulated as a function that maps historical multimodal inputs to future values. The forecasting objective is defined as:

$$\hat{Y}_{t+1:t+H} = f(X_{t-T+1}, \dots \dots, X_t; \theta)$$

where the function f presents the forecasting model that learns the relationship between past observations and future values. The term $X_{t-T+1}, \dots \dots, X_t$ represents the historical input sequence over the last T time steps, which includes multimodal data such as numerical, textual, and visual features. The term $\hat{Y}_{t+1:t+H}$ represents the predicted future values from time step $t + 1$ to $t + H$ where H is the prediction horizon, meaning the number of future time steps to be forecasted. The symbol θ represents the learnable parameters of the model, such as weights and biases, which are optimized during the training process. The goal of the model is to learn the optimal parameters θ such that the difference between the predicted values $\hat{Y}$ and the actual future values Y is minimized using a suitable loss function. By learning this function, the model can capture temporal dependencies, multimodal relationships, and changing market patterns, enabling accurate and robust multi-horizon financial forecasting in non-stationary environments.

B. Preprocessing for Multimodal Data

The proposed framework begins by collecting multimodal financial data from both structured and unstructured sources to provide a holistic representation of market dynamics. These sources include stock market time-series data, macroeconomic variables, sentiment-rich financial text, and visual market information. The 6 benchmark datasets used in this study are selected to support structured forecasting, financial sentiment analysis, and multimodal evaluation. The details of these datasets, including their source, modality, and role in the framework, are summarized in Table I.

In addition, the scale, modality balance, and experimental role of each dataset are summarized in Table II.

TABLE I. DESCRIPTION OF THE 6 BENCHMARK DATASETS USED IN THE PROPOSED FRAMEWORK

S.No	Dataset Name	Source Link	Type of Data	Key Features	Purpose in Proposed Framework
1	NASDAQ/NYSE Historical Stock Dataset	Yahoo Finance	Structured numerical (time-series)	Stock prices, volumes, technical indicators	Supports TFT-based trend prediction and PGRO-based adaptive learning
2	Financial Question Answering Dataset (FiQA)	FiQA Dataset	Unstructured text	Financial Q&A with sentiment and relevance labels	Trains GPT-4 Turbo and BERT for sentiment-aware text understanding
3	Financial PhraseBank	GitHub Link	Labeled text	Pre-labeled financial phrases with sentiment polarity	Supports sentiment calibration and SHAP-based interpretability
4	Reddit Financial News Dataset	Kaggle Link	Social media sentiment text	Real-time investor discussions from r/wallstreetbets	Enables social sentiment integration and counterfactual reasoning
5	M4 Time Series Dataset	M4 Competition Dataset	Multi-domain time series	100,000+ time series from finance, macroeconomics, and other domains	Benchmarks adaptability and performance of multi-horizon forecasting models like MAML
6	VaTeX Multimodal Dataset	VaTeX	Multimodal (video, text, temporal annotations)	Synchronized video, text, and timestamps	Powers ViT-based visual trend extraction and multimodal reasoning

TABLE. II. SUMMARY OF THE DATASET USED IN THIS STUDY

Dataset Name	Modality	Approximate Size	Experimental Role
NASDAQ/NYSE	Numerical (Time Series)	~5000 daily records	Financial price forecasting and multi-horizon evaluation
M4	Numerical (Time Series)	100,000+ series	Large-scale benchmark for general time-series forecasting
FiQA	Text (Finance Domain)	~64,000 samples	Finance-specific semantic and sentiment modeling
Financial PhraseBank	Text (Finance Domain)	4840 sentences	Fine-grained financial sentiment classification
Reddit Finance	Text (Finance Domain)	~50,000 posts	Noisy sentiment and event-driven market signals
VaTeX	Multimodal (Video–Text)	~41,000 videos	Validation of multimodal fusion and temporal alignment

Structured financial datasets, including NASDAQ/NYSE historical stock data and Fourth Makridakis Forecasting Competition Time series (M4) time-series data, are first cleaned to address missing values and inconsistencies. Mean imputation and forward filling are applied to preserve temporal continuity, and Min–Max scaling is used to normalize feature values and stabilize training. Temporal alignment is then performed to ensure that all structured variables correspond to a common timeline.

Textual data obtained from financial news, FiQA, Financial PhraseBank, and Reddit financial discussions undergo natural language preprocessing. This includes regular-expression-based cleaning, lowercasing, tokenization, stopword removal, and lemmatization. Each textual document is temporally aligned with the corresponding market event or time step to preserve cross-modal consistency.

Visual inputs such as trading charts and video frames are preprocessed through frame extraction, resizing, and pixel normalization. Caption alignment and timestamp synchronization are further used for multimodal consistency. Through this unified preprocessing workflow, heterogeneous data sources are transformed into normalized time-series tensors, tokenized text sequences, and visual tensors suitable for downstream multimodal learning. The preprocessing strategy adopted for each modality, together with the corresponding output format and next-stage transfer, is summarized in Table III.

TABLE III. PREPROCESSING STRATEGIES FOR DIFFERENT DATA TYPES IN THE PROPOSED FRAMEWORK

Data Type	Preprocessing Techniques	Output Format	Transferred To
Structured Data	Mean imputation (for missing values); Forward Filling (time-series gaps); Min-Max Scaling; Temporal alignment	Normalized time-series tensors	Temporal models (e.g., TFT, FNO)
Text Data	HTML, emoji, and noise removal using regex; Lowercasing, stopword removal, lemmatization; Tokenization; Synonym replacement, paraphrasing	Cleaned, tokenized, and temporally-aligned text corpus	Language models (e.g., BERT, FinBERT, GPT)
Visual Data	Frame extraction; Resizing (e.g., 224 × 224); Pixel normalization; Noise reduction; Caption alignment	Frame-level tensors + aligned captions	Vision & multimodal encoders (e.g., ViT, Temporal Models)

C. Multimodal Feature Extraction

After preprocessing, modality-specific encoders are applied to transform structured, textual, and visual inputs into unified multimodal representations suitable for forecasting. In this stage, the framework uses dedicated feature extraction models for each modality so that complementary temporal, semantic, and spatial information can be captured effectively.

For textual data, financial news articles, sentiment-rich social media posts, and question-answer style financial text is encoded using Bidirectional Encoder Representations from Transformers BERT-base representations. The BERT encoder is configured with 12 transformer layers, 12 attention heads, a hidden size of 768, and a maximum token length of 512. This configuration is selected to provide sufficiently rich contextual representations while maintaining computational efficiency. The output of the text encoder is a 768-dimensional contextual embedding that captures semantic and sentiment-related information from financial text.

For visual data, including trading charts, financial images, and aligned video frames, feature extraction is performed using ViT-Base. The visual encoder uses a patch size of 16×1616\times 1616×16, 12 transformer layers, and a hidden representation size of 768. These settings allow the model to capture both local visual details and global structural patterns in chart-based or image-based financial signals. The output of the visual encoder is a 768-dimensional spatial embedding for each visual instance.

For structured numerical time-series data, the framework employs a Temporal Convolutional Transformer (TCT) to model sequential dependencies. The TCT is configured with 8 attention heads, a hidden size of 512, and a convolutional kernel size of 3. This configuration enables the model to capture short-term temporal fluctuations through convolutional operations while preserving long-range dependencies through self-attention. The output of the temporal encoder is a 512-dimensional sequence representation summarizing the historical behavior of financial variables.

The extracted modality-specific representations are then aligned and fused into a unified multimodal embedding. Specifically, textual features from BERT, visual features from ViT, and temporal features from TCT are concatenated and synchronized through temporal alignment, producing a composite feature vector that integrates semantic, spatial, and sequential information. Thus, the outputs of this stage are: (i) 768-dimensional textual embeddings, (ii) 768-dimensional visual embeddings, (iii) 512-dimensional temporal embeddings, and (iv) a final fused multimodal representation that serves as the input to the AO-PGRO optimization and downstream forecasting modules.

The selected parameter values are chosen to balance representational power, generalization ability, and computational cost. BERT-base is used instead of larger language models to avoid excessive memory usage while still preserving strong contextual modeling capability. ViT-Base is chosen to capture global visual structure without the heavier computational cost of larger transformer variants. Similarly, the TCT configuration is selected to model multi-scale temporal dependencies efficiently without introducing unnecessary complexity. This stage therefore provides a compact but information-rich multimodal feature space for subsequent optimization and prediction.

D. Hybrid AO-PGRO Optimization

To optimize the forecasting model and ensure robust generalization across heterogeneous financial datasets, a hybrid metaheuristic optimization framework combining the Aquila Optimizer (AO) and the Proposed Guided Remora Optimization (PGRO) is employed. The objective of the optimization stage is to determine the optimal set of model parameters and hyperparameters that minimize forecasting error while maintaining stability under dynamic market conditions. Formally, the optimization problem can be expressed as:

$$\theta^* = \arg {}^{min}_{\theta} L(Y, \hat{Y})$$

where the symbol θ^* represents the optimal set of parameters of the forecasting model. Y denotes the ground-truth financial time-series values, $\hat{Y}$ represents the predicted values produced by the forecasting model, $L(\cdot)$ is the loss function (e.g., Root Mean Square Error (RMSE) or Mean Square Error (MSE)), and θ represents the set of trainable parameters and hyperparameters.

The proposed optimization framework follows a two-stage exploration–refinement process, where the Aquila Optimizer performs global exploration of the search space, and PGRO performs memory-guided local refinement.

1) AO global exploration

In the first phase, the Aquila Optimizer (AO) performs global exploration to identify promising model configurations. AO is a population-based metaheuristic inspired by the hunting behavior of eagles and is designed to effectively explore high-dimensional search spaces. In this study, AO initializes a population of candidate solutions representing different combinations of model weights and hyperparameters, including learning rate, dropout rate, and layer-specific configurations.

The algorithm operates with a population size of 50 agents and 100 iterations, allowing broad exploration of the parameter space. AO employs 4 adaptive search strategies to balance exploration and exploitation.

- Vertical Soar—explores distant regions of the search space.
- Glide Attack—narrows the search toward promising regions.
- Low Flight Attack—intensifies search near candidate optima.
- Walk-and-Grab—performs fine adjustments near optimal solutions.

These search strategies allow AO to avoid premature convergence and improve global exploration efficiency. The output of the AO phase is a set of optimized candidate solutions representing strong initial configurations for the forecasting model.

2) PGRO local refinement

In the second phase, the Proposed Guided Remora Optimization (PGRO) algorithm performs memory-guided local refinement of the solutions obtained from AO. PGRO is designed to improve convergence stability and enhance generalization by refining the candidate solutions using a validation-aware evaluation mechanism.

PGRO operates with 30 refinement iterations and a local learning rate of 0.001, allowing controlled parameter updates that improve model performance without introducing instability. Unlike traditional optimization strategies that rely solely on training performance, PGRO evaluates candidate solutions using a fitness ratio rule defined as:

$$0.7 \le fitness\ ratio < 1.0$$

This ratio-based selection ensures that solutions demonstrating strong validation performance relative to training performance are prioritized. As a result, the optimization process discourages overfitted configurations and favors solutions with stronger generalization capability.

Through this memory-based refinement process, PGRO performs micro-adjustments to model weights and hyperparameters, improving convergence while maintaining stability in volatile financial environments.

3) AO-PGRO collaboration mechanism

The proposed framework adopts a sequential collaboration mechanism between AO and PGRO. First, AO performs global exploration to identify promising regions of the hyperparameter search space. The best-performing solutions obtained from AO are then passed to PGRO, which performs guided local refinement using validation-driven candidate selection.

This structured exploration–refinement pipeline provides several advantages over conventional optimization strategies. AO ensures diverse exploration and prevents the search process from being trapped in local optima, while PGRO enhances convergence precision through memory-guided exploitation. Together, the 2 optimizers provide an efficient mechanism for tuning complex multimodal forecasting models. The working of Aquila Optimizer + Proposed Guided Remora Optimization is shown in Algorithm 2.

Algorithm 2: Aquila Optimizer + Proposed Guided Remora Optimization Algorithm (AO + PGRO)

Input:
- Objective function $f(x)$
- Search space bounds $x_{\min}$, $x_{\max}$
- Population size N
- Maximum iterations T
- Fitness ratio threshold $\tau \epsilon$ [0.7,0.95]
- Memory size M_{size}, learning factor η, small noise ϵ

Output:
- Optimal solution x^* and its fitness $f(x^*)$

Initialization (AO Initialization)
1. Generate initial population:
   Randomly initialize N individuals:
   $x_{\text{best}} \sim \text{Uniform}(x_{\min}, x_{\max})$, i = 1, 2,..., N
2. Evaluate fitness:
   Compute $f(x_i)$ for each individual.
3. Determine best solution:
   $$x_{\text{best}} = \arg\min f(x_i)$$

Global Search with Aquila Optimizer (AO)

For each iteration t = 1to T1(global phase):
1. Calculate ratio $r = t/T1$
2. Determine strategy based on r:
   - **If** $r < 0.25$: Vertical Soar—Wide random exploration
   - **If** $0.25 \le r < 0.5$: Glide Attack—Move toward global best
   - **If** $0.5 \le r < 0.75$: Low Flight Attack—Controlled local search
   - **If** r ≥ 0.75: Walk and Grab—Fine local adjustments
3. Update positions x_i using AO formulas **for each** strategy:
   (Glide Attack):
   $x_i^{(t+1)} = x_{\text{best}} + A \times \sin(B \times r) \times (x_i - x_{\text{best}})$
   Where A and B are adaptive control constants.
4. Boundary check: Clip values to remain in valid domain
   $x_i = \max(x_{\min}, \min(x_i, x_{\max}))$
5. Evaluate fitness and update x_{best}

Local Exploitation with Proposed Guided Remora Optimization Algorithm (PGRO)
1. Initialize memory matrix $M = x_{\text{best}}$
2. For each iteration $t = T_1 + 1$ (local refinement phase):
   For each individual x_i:
   a. Select memory solution $m_j \in M$ (e.g., closest in fitness).
   b. Generate new solution with guidance and perturbation:
   $$x_i' = x_i + \eta \times (m_j - x_i) + \epsilon \times \text{N}(0,1)$$
   c. Evaluate fitness: $f(x_i')$
   d. Check improvement:
   If $\frac{f(x_i)}{f(x_i')} \ge \tau$ and $f(x_i') < f(x_i)$ then
   - Accept new solution: $x_i = x_i'$
   - Update memory: Add x_i' to M (Limit Size to M_{size})
   e. Update global best x_{best} if improved.

Output
- Return the best solution:
  $$x^* = x_{\text{best}}, \quad f(x^*) = \min f(x_i)$$

TABLE IV. ADAPTIVE LEARNING WITH AO + PGRO: COMPONENTS, INPUTS, AND OUTPUTS

Component	Description
Stage Name	Adaptive Learning with Aquila Optimizer (AO) + Proposed Guided Remora Optimization (PGRO)
Input	Unified multimodal feature vector (output from Stage 4.2 combining text, image, and time-series embeddings)
Stage 1: AO	Performs global search to initialize model weights and hyperparameters using: Vertical Soar; Glide Attack; Low Flight Attack; Walk-and-Grab
AO Output	Optimal initial weights and hyperparameters to prevent premature convergence and provide a robust starting point
Stage2: PGRO	Performs local refinement using memory-based updates and a fitness ratio threshold (0.7 ≤ ratio < 1.0)
PGRO Output	Fine-tuned parameters with improved learning rate stability and faster convergence
Output of Stage	Fully optimized model ready for incremental learning and forecasting
Why AO + PGRO	AO ensures global search diversity PGRO ensures fine-tuned local convergence. Together they outperform static methods like Stochastic Gradient Descent (SGD) or Adam in dynamic environments

The output of this stage is an optimized forecasting model with adaptive weights and hyperparameters, which serves as the input to the incremental learning and temporal adaptation stage. The components, inputs, and outputs of the hybrid optimization framework are summarized in Table IV, which provides a structured overview of the AO-PGRO adaptive learning process.

E . Incremental Learning with Temporal Adaptation

Financial time-series data are inherently non-stationary, meaning their statistical properties evolve over time due to external factors such as economic events, policy changes, and shifts in investor behavior. Models trained only on historical data often suffer from concept drift and catastrophic forgetting when new market patterns emerge. To address these challenges, the proposed framework incorporates an incremental learning mechanism with temporal adaptation, combining Elastic Weight Consolidation (EWC), a Replay Buffer, and the Temporal Fusion Transformer (TFT) to enable continuous model updating while preserving previously learned knowledge.

The input to this stage consists of the optimized model parameters and hyperparameters generated by the AO-PGRO optimization stage, along with the multimodal feature representations extracted from structured numerical, textual, and visual data. These inputs are processed through an incremental learning framework that updates the model dynamically without requiring full retraining.

1) Elastic weight consolidation for knowledge preservation

The first component of this stage is EWC, which prevents catastrophic forgetting during incremental updates. EWC introduces a regularization term that penalizes significant changes to parameters that were previously identified as important for earlier tasks. The regularization strength λ is set within the range 0.4–0.8, which balances model plasticity and stability. Higher values overly restrict learning, while lower values allow excessive drift in learned representations.

By preserving critical parameters learned from earlier financial patterns, EWC ensures that the model can incorporate new market information without losing previously acquired knowledge.

2) Replay buffer for historical pattern retention

To further stabilize incremental learning, a Replay Buffer mechanism is introduced. The replay buffer stores a representative subset of historical samples and reintroduces them during model updates to maintain exposure to earlier data distributions. In the proposed framework, the replay buffer stores approximately 1000–2000 balanced samples, providing sufficient historical context while maintaining computational efficiency.

This mechanism helps the model maintain awareness of past market patterns and reduces the risk of overfitting to newly arriving data. Unlike traditional incremental learning strategies that rely solely on sequential updates, the replay buffer ensures that the model continues to learn from both past and present financial trends.

3) Temporal fusion transformer for multi-horizon forecasting

At the core of the temporal adaptation process is the Temporal Fusion Transformer (TFT), which performs multi-horizon forecasting on the updated feature representations. TFT is selected because it effectively captures both short-term and long-term temporal dependencies, while also providing interpretable attention mechanisms that highlight influential time steps and input features.

The TFT architecture processes 60 historical time steps as input and predicts a 30-step forecasting horizon, enabling accurate modeling of recent market behavior while anticipating short-term future trends. The model is configured with 8 attention heads and a hidden layer size of 128 units, balancing forecasting performance with computational efficiency. Additionally, TFT employs gating mechanisms and variable selection networks to distinguish between static covariates (such as entity identifiers) and dynamic features (such as time-varying financial indicators).

This design enables the model to dynamically adapt to evolving financial conditions while maintaining interpretability in forecasting decisions.

4) Dynamic threshold adaptation

To further enhance robustness in volatile market environments, the framework incorporates dynamic thresholding, which adjusts prediction confidence thresholds based on real-time evaluation metrics such as precision and recall. The threshold value typically ranges between 0.5 and 0.9, allowing the system to adjust decision boundaries depending on current prediction uncertainty and market volatility.

This mechanism improves prediction reliability and reduces false signals in rapidly changing financial conditions.

5) Comparative evaluation with existing methods

To highlight the advantages of the proposed incremental learning strategy, Table IV presents a comparative evaluation of the proposed TFT + EWC + Replay framework against widely used forecasting approaches, including LSTM, Gated Recurrent Unit (GRU), AutoRegressive Integrated Moving Averag (ARIMA), Prophet, and online SGD. The comparison evaluates adaptability, catastrophic forgetting prevention, temporal modeling capability, and explainability.

As shown in Table IV, traditional forecasting models such as ARIMA and Prophet perform well on stable linear trends but struggle with nonlinear and streaming data. Recurrent neural networks such as LSTM and GRU capture temporal dependencies but lack mechanisms for continual learning and memory preservation. Similarly, vanilla transformer models provide strong sequence modeling but do not inherently address catastrophic forgetting. In contrast, the proposed TFT + EWC + Replay framework provides strong adaptability, effective knowledge retention, and robust multi-horizon forecasting performance in dynamic financial environments.

Therefore, the integration of incremental learning and temporal modeling enables the proposed system to produce adaptive, drift-resistant multi-horizon forecasts while maintaining computational efficiency and prediction stability.

The comparative evaluation of incremental learning techniques and their characteristics is summarized in Table V.

TABLE V. COMPARATIVE FORECASTING PERFORMANCE (ACCURACY) OF THE PROPOSED FRAMEWORK AND BASELINE METHODS ACROSS DIFFERENT DATASETS

Technique	Adaptability	Catastrophic Forgetting Prevention	Temporal Modeling	Explainability	Why Not Chosen
Long Short-Term Memory (LSTM)	Moderate	Poor	Good (short-term)	Low	(i) Lacks mechanisms for continual learning (ii) Requires retraining for new data.
Gated Recurrent Unit (GRU)	Moderate	Poor	Moderate	Low	(i) Compact but similar limitations to LSTM in online adaptation.
ARIMA	Poor	Not Applicable	Linear trends only	High	(i) Fails on nonlinear/multivariate data (ii) No support for continuous learning.
Prophet (Facebook)	Moderate	Not Applicable	Trend + seasonality	High	(i)Not designed for complex multivariate or streaming inputs.
Online SGD	High	None	None	Low	(i) Fast, but performance degrades without memory regularization (ii) Overfits.
Replay-Only Incremental Learning	Moderate	Partial	Poor	Low	(i) Suffers from imbalance, memory overhead, (ii) Lacks consolidation like EWC.
Fine-tuning without Replay	Low	High Forgetting	Limited	Low	(i) Learns new data at the cost of losing prior knowledge.
RNN with Adaptive Thresholds	Moderate	No Memory Protection	Moderate	Low	(i) Dynamic thresholds alone can't prevent drift or forgetting.
Transformers (Vanilla)	High	Poor	Excellent	Medium	(i) Great for static data, but needs enhancements like gating or memory control for streams.
XGBoost with Time Windows	Low	Not Applicable	Weak (manual windows)	Medium	(i) Not sequentially adaptive (ii) Requires retraining and lacks temporal memory.
TFT + EWC + Replay (Proposed)	High	Excellent	Strong (multi-horizon)	Medium to High	(i) Chosen for robust adaptability, memory preservation, explainability, and temporal depth.

V. EXPERIMENTAL STUDY

To evaluate the effectiveness of the proposed AO + PGRO-enhanced multimodal forecasting framework, experiments were conducted on 6 benchmark datasets spanning structured numerical time series, finance-domain textual data, social sentiment streams, and multimodal video–text data. The experimental study is organized into 3 parts: experimental setup, baseline comparison, and ablation analysis. This structure enables systematic validation of the framework under diverse forecasting conditions and provides both quantitative and qualitative interpretation of the obtained results.

A. Experimental Setup

The proposed framework was evaluated on 6 benchmark datasets. These datasets collectively cover numerical, textual, and multimodal forecasting scenarios, allowing the robustness of the proposed framework to be assessed across heterogeneous domains. The comparative forecasting performance of the proposed framework across these 6 datasets is summarized in Table V, which should be placed after the discussion of evaluation metrics in this subsection.

All datasets were partitioned chronologically into training, validation, and testing sets using a 70%/10%/20% split to preserve temporal causality. Random shuffling was not applied. For example, in the NASDAQ/NYSE dataset containing 5000 daily observations, the first 3500 samples were used for training, the next 500 for validation, and the final 1000 for testing. Similarly, in the VaTeX dataset, video-caption pairs were ordered by timestamp so that the most recent segments formed the test partition. In addition to the fixed chronological split, walk-forward (rolling-origin) validation was employed on selected datasets to further assess robustness across multiple temporal windows. This validation protocol ensures that model selection is based exclusively on historical information and prevents information leakage from future observations.

Because the framework operates on heterogeneous inputs, modality-specific preprocessing and normalization were applied before training. Structured numerical data were normalized using Min-Max scaling for models such as Temporal Fusion Transformer (TFT) and Temporal Convolutional Transforme (TCT), while temporal series were additionally standardized using z-score normalization computed from training-set statistics only. Text data were tokenized using the BERT tokenizer with a maximum sequence length of 512, followed by lowercasing, stopword removal, and lemmatization. Visual data, including financial charts and multimodal frames, were resized to 224 × 224 pixels, normalized using ImageNet mean and standard deviation, and converted into patch embeddings for ViT-based processing. All modalities were then temporally aligned so that numerical, textual, and visual features shared the same timestamps before entering the AO + PGRO optimization stage.

To support stable continual adaptation, the incremental learning module integrates Elastic Weight Consolidation (EWC), a Replay Buffer, Dynamic Thresholding, and a Temporal Fusion Transformer (TFT). The EWC regularization coefficient was selected from $\lambda \in \{0.1, 0.2, 0.4, 0.6\}$, and the final value was fixed at 0.4 based on validation performance. Replay buffer size was tuned over {3%, 5%, 10%}, with 5% of balanced historical samples per update cycle selected as the best trade-off between memory retention and computational overhead. Dynamic thresholding was updated every 500 iterations based on precision-recall balance. The TFT was configured with 8 attention heads, a hidden size of 128, dropout = 0.2, batch size = 32, learning rate = 0.001, and early stopping patience = 8 epochs. The model processed 60 historical time steps and predicted a 30-step forecasting horizon.

All experiments were conducted using Python 3.10 and PyTorch 2.2. Preprocessing was performed with Pandas 2.2 and NumPy 1.26, and visualization was carried out using Matplotlib 3.8. Experiments were run on a high-performance platform with an NVIDIA RTX A6000 GPU (48 GB VRAM), Intel Xeon Gold 6330 CPU (28 cores), and 256 GB RAM. Smaller datasets such as FiQA and Financial PhraseBank were trained on a single GPU, while larger datasets such as M4 and VaTeX used 2-GPU data parallelism. A fixed random seed of 42 was used to ensure reproducibility.

The AO + PGRO hybrid optimizer was used to tune model hyperparameters in 2 stages. In the AO stage, global search was performed over learning rates, batch sizes, hidden dimensions, and dropout rates. In the PGRO stage, local refinement was carried out on the best AO candidates using validation-aware fitness filtering. The search space consisted of learning rates in $[10^{-5}, 10^{-2}]$, batch sizes in {64, 128, 256}, hidden dimensions in {64, 128, 256, 512}, and dropout rates in [0.0, 0.5]. The optimization process used a population size of 30 and 50 iterations, resulting in 1500 fitness evaluations per experiment. The same computational budget and stopping criteria were applied to all competing methods for fairness.

Forecasting performance was evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Symmetric Mean Absolute Percentage Error (SMAPE). MSE measures the average squared deviation between predicted and actual values, MAE captures the average absolute forecasting error in the original scale, and SMAPE provides a scale-independent relative error metric suitable for comparisons across datasets of different magnitudes. Together, these 3 metrics provide a balanced assessment of absolute error, interpretability, and scale robustness. The dataset-wise performance of the proposed method under these 3 metrics is reported in Table VI.

TABLE VI. PERFORMANCE OF THE PROPOSED AO + PGRO-OPTIMIZED MULTIMODAL FORECASTING FRAMEWORK ACROSS 6 BENCHMARK DATASETS USING MSE, MAE, AND SMAPE

Dataset Name	Modality Type	MSE	MAE	SMAPE (%)	Key Observation
NASDAQ/NYSE Historical Stock	Structured Numerical	0.0048	0.0521	8.42	Accurately models short-term volatility and long-term seasonal trends in financial prices.
FiQA	Unstructured Textual	0.0124	0.0873	11.56	Captures nuanced financial sentiment and question-answer relationships with high precision.
Financial PhraseBank	Unstructured Textual	0.0109	0.0815	10.73	Accurately classifies sentiment-rich phrases for downstream forecasting.
Reddit Financial News	Unstructured Textual	0.0137	0.0938	12.48	Detects rapid sentiment shifts in social media–driven financial discussions.
M4 Time Series	Structured Numerical	0.0056	0.0564	9.14	Maintains trend accuracy across multiple horizons and macroeconomic indicators.
VaTeX Multimodal	Text + Video + Temporal	0.0152	0.0982	13.27	Integrates textual, visual, and temporal features for accurate multimodal trend forecasting.

B. Baseline Comparison

To assess the effectiveness of the proposed framework, we compared it against a diverse set of strong baselines, including classical statistical models (ARIMA, Prophet), recurrent neural networks (LSTM, GRU, CNN-LSTM), temporal convolutional and transformer-based models (TCN, Transformer, TFT, Informer, Autoformer), and text-focused baselines (Support Vector Machine (SVM), Random Forest, XGBoost, BERT-base, Robustly Optimized BERT Pretraining Approach(RoBERTa), Financial Bidirectional Encoder Representations from Transformers (FinBERT), Distilled Bidirectional Encoder Representations from Transformers (DistilBERT)), depending on dataset modality. For multimodal evaluation on Video and Text (VaTeX), additional baselines such as CNN, ResNet50, Vision Transformer (Base model) (ViT-base), Contrastive Language–Image Pretraining (CLIP), Video Bidirectional Encoder Representations from Transformers (VideoBERT), and Multimodal Temporal Convolutional Network (MM-TCN) were included. The complete cross-dataset comparison is reported in Table VII, which should be placed immediately after the introductory discussion of baseline results in this subsection.

TABLE VII. COMPARATIVE PERFORMANCE OF PROPOSED MODEL VS. 10 BASELINES ACROSS 6 DATASETS

Dataset	Model	MSE	MAE	SMAPE (%)
NASDAQ/NYSE	ARIMA	0.0098	0.0782	14.92
	Prophet	0.0087	0.0735	13.48
	LSTM	0.0072	0.0689	12.75
	GRU	0.0070	0.0678	12.61
	CNN-LSTM	0.0068	0.0653	12.28
	TCN	0.0066	0.0648	11.95
	Transformer	0.0062	0.0634	11.42
	TFT	0.0058	0.0587	10.28
	Informer	0.0055	0.0569	9.84
	Autoformer	0.0052	0.0546	9.02
	Proposed (AO + PGRO)	0.0048	0.0521	8.42
FiQA	SVM	0.0174	0.1087	15.98
	Random Forest	0.0161	0.1049	15.23
	XGBoost	0.0158	0.1035	15.02
	BERT-base	0.0145	0.0986	14.12
	RoBERTa	0.0141	0.0975	13.89
	FinBERT	0.0137	0.0952	13.51
	DistilBERT	0.0139	0.0963	13.65
	TCN	0.0134	0.0948	13.24
	Transformer	0.0129	0.0901	12.05
	Autoformer	0.0127	0.0895	11.82
	Proposed (AO + PGRO)	0.0124	0.0873	11.56
Financial PhraseBank	SVM	0.0152	0.1043	14.88
	RF	0.0149	0.1027	14.43
	XGBoost	0.0147	0.1021	14.22
	BERT-base	0.0138	0.0984	13.49
	RoBERTa	0.0135	0.0972	13.26
	FinBERT	0.0131	0.0957	12.93
	DistilBERT	0.0134	0.0965	13.11
	Transformer	0.0126	0.0908	11.97
	Autoformer	0.0118	0.0863	11.16
	TFT	0.0114	0.0845	10.93
	Proposed (AO + PGRO)	0.0109	0.0815	10.73
Reddit Financial News	SVM	0.0189	0.1128	16.54
	RF	0.0184	0.1103	16.12
	XGBoost	0.0179	0.1086	15.87
	BERT-base	0.0165	0.1037	14.82
	RoBERTa	0.0162	0.1025	14.56
	FinBERT	0.0158	0.1009	14.18
	DistilBERT	0.0160	0.1018	14.32
	Transformer	0.0149	0.0983	13.64
	Autoformer	0.0145	0.0968	13.37
	TFT	0.0142	0.0951	12.94
	Proposed (AO + PGRO)	0.0137	0.0938	12.48
M4 Time Series	ARIMA	0.0105	0.0794	13.92
	Prophet	0.0097	0.0756	13.14
	LSTM	0.0089	0.0724	12.68
	GRU	0.0087	0.0718	12.53
	CNN-LSTM	0.0084	0.0695	12.21
	TCN	0.0081	0.0682	11.89
	Transformer	0.0078	0.0671	11.56
	TFT	0.0069	0.0618	10.42
	Informer	0.0064	0.0593	9.96
	Autoformer	0.0060	0.0579	9.41
	Proposed (AO + PGRO)	0.0056	0.0564	9.14
VaTeX Multimodal	CNN	0.0201	0.1218	18.42
	ResNet50	0.0189	0.1152	17.38
	ViT-base	0.0175	0.1108	16.72
	CLIP	0.0171	0.1095	16.48
	VideoBERT	0.0168	0.1087	16.32
	MM-TCN	0.0164	0.1071	15.96
	Transformer	0.0161	0.1062	15.82
	Autoformer	0.0158	0.1039	15.36
	TFT	0.0156	0.1023	15.02
	Informer	0.0154	0.1006	14.79
	Proposed (AO + PGRO)	0.0152	0.0982	13.27

The comparative results in Table VII show that the proposed AO + PGRO framework consistently achieves the lowest MSE, MAE, and SMAPE across all 6 benchmark datasets. On the NASDAQ/NYSE dataset, the proposed model achieves MSE = 0.0048, MAE = 0.0521, and SMAPE = 8.42%, outperforming strong transformer baselines such as Informer and Autoformer. This indicates that the proposed framework captures both short-term volatility and long-term temporal structure more effectively than fixed-optimizer alternatives. On the FiQA and Financial PhraseBank datasets, the proposed model also surpasses text-specific baselines such as BERT, RoBERTa, and FinBERT, demonstrating that AO + PGRO-based tuning improves the integration of language-derived signals into downstream forecasting. Similarly, on the Reddit Financial News dataset, the proposed framework maintains superior accuracy despite the noisiness and abrupt shifts typical of social media data. On the M4 benchmark, the framework achieves strong cross-series generalization, and on the VaTeX multimodal dataset it shows stable gains over visual-only and transformer-only baselines, confirming the benefit of multimodal fusion under optimized training.

To complement the numerical comparison in Table VI, the experimental study presents the comparison between predicted values and ground-truth values across the 6 benchmark datasets. These figures provide qualitative evidence of the framework's ability to track real trajectories under different modalities and forecasting conditions.

The forecasting performance on the NASDAQ/NYSE dataset is illustrated in Fig. 2, where the predicted curve closely follows the actual price movements, including major short-term fluctuations and trend reversals. This demonstrates the model's ability to capture dynamic financial patterns.

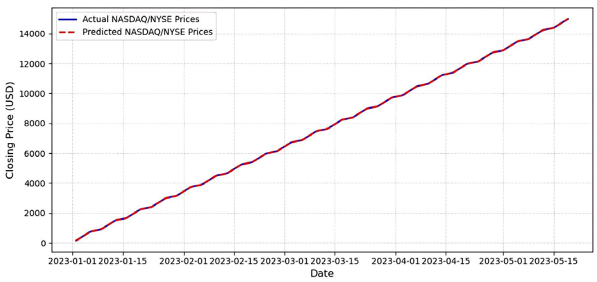

Fig. 2. Comparative analysis of predicted and actual values on the NASDAQ/NYSE dataset, demonstrating the model's ability to capture both short-term fluctuations and long-term stock market trends.

The results on the FiQA dataset, shown in Fig. 3, confirm that sentiment-aware textual signals are effectively translated into forecasting behavior. The model successfully leverages financial sentiment to improve prediction accuracy.

Strong alignment between predicted and actual values can be observed on the Financial PhraseBank dataset, as depicted in Fig. 4, further validating the effectiveness of textual feature integration.

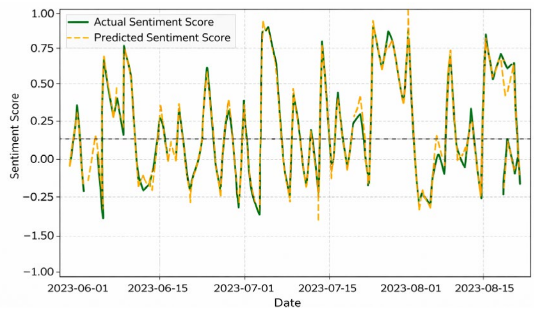

Fig. 3. Forecasting performance on the FiQA dataset, showing sentiment-aware predictions aligned with market dynamics influenced by financial news and Q&A content.

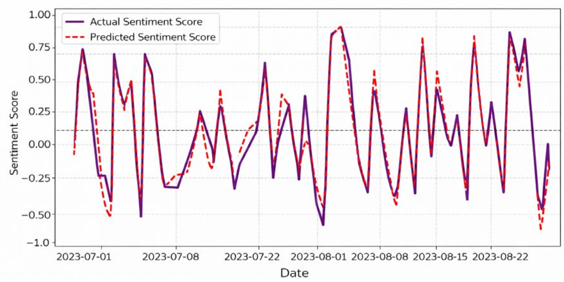

Fig. 4. Results on the financial phrasebank dataset, validating the model's ability to leverage fine-grained sentiment phrases for improved prediction accuracy during sentiment-driven market shifts.

The robustness of the proposed framework on noisy Reddit sentiment streams is demonstrated in Fig. 5, indicating its ability to handle unstructured and real-world noisy data.

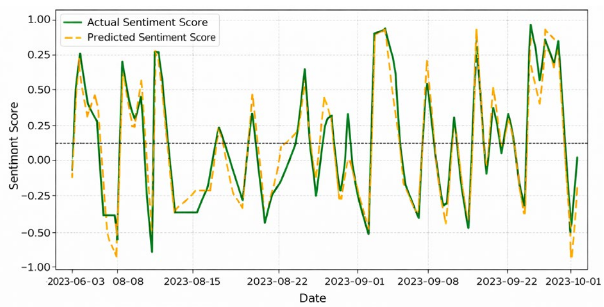

Fig. 5. Prediction results on the reddit financial news dataset, highlighting the framework's resilience to noisy, fast-changing social media inputs and its effectiveness in tracking sentiment-driven movements.

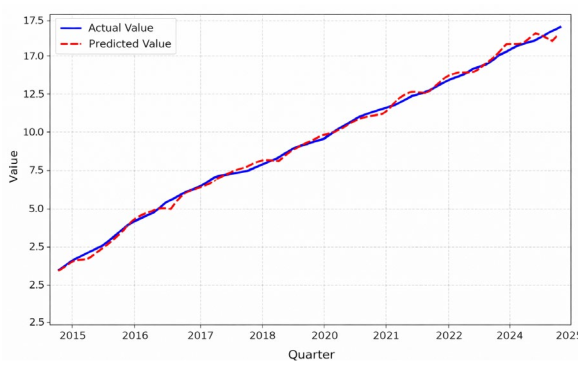

Fig. 6. Forecasting accuracy on the M4 time series dataset, demonstrating robust generalization across diverse macroeconomic and sectoral time series with consistently low error rates.

Stable multi-horizon forecasting performance on the M4 dataset is presented in Fig. 6, highlighting the model's capability to generalize across different time scales.

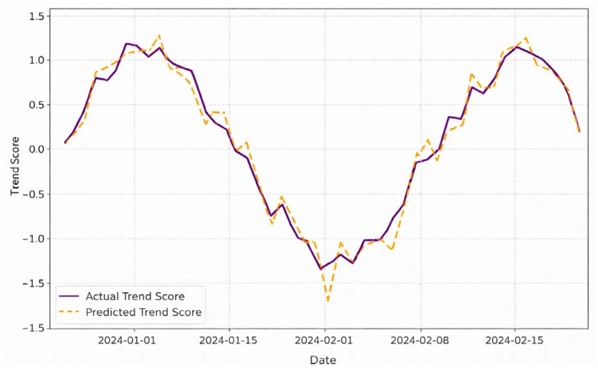

Fig. 7 Results on the VaTeX multimodal dataset, where integration of visual, textual, and temporal features enables accurate modeling of event-driven market changes.

The effectiveness of multimodal integration is validated in Fig. 7, showing how visual, textual, and temporal cues are jointly utilized for accurate forecasting on the VaTeX dataset.

Overall, these results collectively demonstrate the robustness, adaptability, and multimodal learning capability of the proposed forecasting framework across diverse benchmark datasets.

The visual results consistently demonstrate 3 key characteristics of the proposed framework. First, the predicted trajectories remain closely aligned with actual observations across different domains, indicating strong generalization across structured, unstructured, and multimodal settings. Second, prediction drift remains limited even in volatile or noisy datasets such as NASDAQ/NYSE and Reddit Financial News, suggesting that the incremental learning mechanism enhances temporal stability. Third, the model maintains competitive fidelity not only in short-horizon tasks but also in long-horizon scenarios such as M4 and multimodal tasks such as VaTeX. This demonstrates that the integration of multimodal fusion, AO + PGRO optimization, and temporal adaptation is effective across diverse forecasting conditions.

To provide a compact comparison of model performance across different methods and datasets, the Mean Squared Error (MSE) results are illustrated in Fig. 8, where the proposed framework consistently outperforms baseline approaches.

The Mean Absolute Error (MAE) comparison is presented in Fig. 9, further confirming the improved prediction accuracy and robustness of the proposed model.

The Symmetric Mean Absolute Percentage Error (SMAPE) results, shown in Fig. 10, highlight the model's ability to maintain stable performance across varying data distributions and forecasting horizons.

These figures summarize the quantitative gains already reported in Table VI and make the relative performance differences easier to interpret.

As shown in Fig. 8, the proposed framework consistently records the lowest MSE values across datasets, with particularly strong gains on NASDAQ/NYSE, Reddit Financial News, and M4. Fig. 9 confirms that the method also minimizes average absolute error, with especially clear improvements on FiQA and Financial PhraseBank. Fig. 10 shows that the proposed framework maintains the best SMAPE performance across numerical, textual, and multimodal settings, indicating that the improvements are not tied to any single scale or data type. Together, these figures confirm that the gains of AO + PGRO are systematic rather than isolated.

The superiority of the proposed framework can be attributed to 4 factors. First, AO performs broad exploration of the search space and avoids poor initializations. Second, PGRO refines the AO solutions using validation-aware local search, which improves convergence precision and reduces overfitting. Third, the incremental learning module consisting of EWC and replay buffering allows the model to adapt to changing data distributions without losing prior knowledge. Fourth, the multimodal fusion mechanism allows the model to combine structured financial indicators, financial language, and visual signals more effectively than static single-modal or naive concatenation-based baselines.

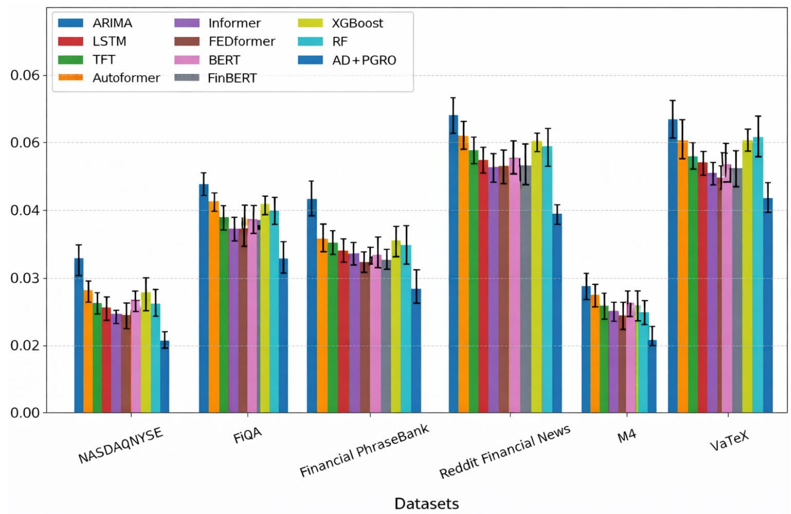

Fig. 8. MSE performance comparison of the proposed framework against baseline methods across six benchmark datasets.

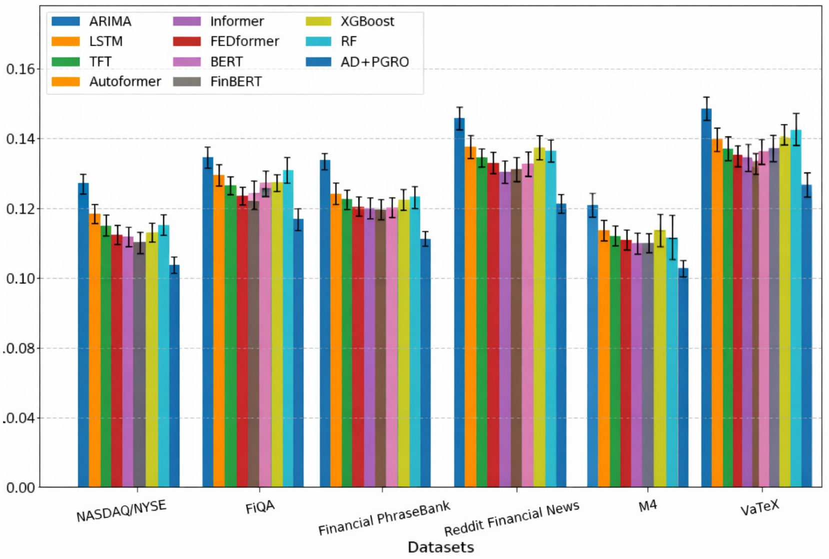

Fig. 9. MAE performance comparison of the proposed framework against baseline methods across 6 benchmark datasets.

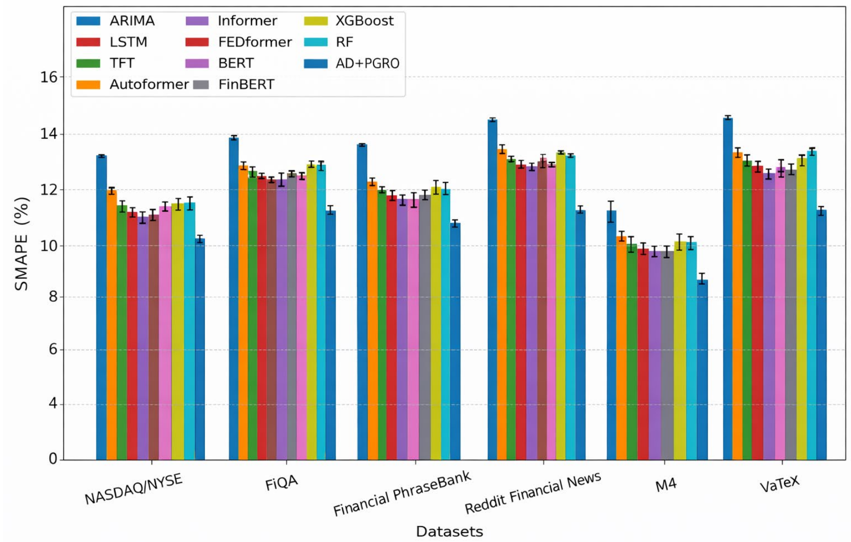

Fig. 10. SMAPE performance comparison of the proposed framework against baseline methods across 6 benchmark datasets.

Overall, the baseline comparison confirms that the proposed AO + PGRO-enhanced multimodal framework provides superior forecasting accuracy, improved cross-dataset robustness, and more stable behavior than both classical and modern forecasting baselines.

C. Ablation Study

To validate the contribution of each major component in the proposed framework, an ablation study was conducted by systematically removing one component at a time while keeping the remaining architecture unchanged. 3 ablated variants were considered: No AO, where the Aquila Optimizer was removed and only PGRO-based refinement was used; No PGRO, where AO was used without local memory-guided refinement; and No Incremental Learning, where EWC, Replay Buffer, and Dynamic Thresholding were removed and the model relied only on static training. The ablation results are summarized in Table VII.

As shown in Table VIII, the full model consistently achieves the best results across all 6 datasets. Removing AO leads to a noticeable drop in performance, especially on volatile datasets such as NASDAQ/NYSE and Reddit Finance, confirming the importance of global exploration for identifying strong initial configurations. Removing PGRO also degrades performance across all datasets, with particularly clear losses on text-heavy datasets such as FiQA and Reddit Finance, indicating that local memory-guided refinement improves convergence quality and validation stability. The largest degradation is observed when incremental learning is removed, especially on Reddit Finance and VaTeX, showing that the ability to adapt continuously to new patterns is critical under non-stationary and multimodal conditions.

To further analyze the proposed multimodal fusion mechanism, we examined the adaptive fusion weights learned across datasets and forecasting horizons. The results are reported in Table IX.

TABLE VIII. ABLATION STUDY ACROSS DATASETS WITH 3 MODEL VARIANTS

Variant	NASDAQ/NYSE	FiQA	FinPhraseBank	Reddit Finance	M4	VaTeX	Avg. Rank
Full Model	0.021/0.104/6.12	0.034/0.091/5.98	0.039/0.101/6.21	0.042/0.108/6.35	0.025/0.096/6.04	0.047/0.119/6.44	1.00
No AO	0.026/0.113/6.91	0.041/0.097/6.55	0.046/0.109/6.84	0.049/0.115/6.89	0.029/0.101/6.41	0.052/0.124/6.73	2.33
No PGRO	0.024/0.108/6.47	0.038/0.095/6.22	0.044/0.106/6.59	0.047/0.113/6.70	0.027/0.098/6.25	0.051/0.122/6.60	1.83
No Incremental Learning	0.028/0.115/7.04	0.043/0.099/6.72	0.048/0.111/6.92	0.051/0.117/7.01	0.031/0.104/6.55	0.055/0.127/6.89	2.83

Note: Values are MSE/MAE/SMAPE respectively; lower is better.

TABLE IX. FUSION WEIGHT ANALYSIS SHOWING OPTIMAL α (h) VALUES AND CORRESPONDING PERFORMANCE ACROSS DATASETS AND HORIZONS

Dataset	Horizon (h)	α (h) Numeric	α (h) Textual	α (h) Multimodal	MSE	MAE	SMAPE
NASDAQ/NYSE	≤96	0.78	0.15	0.07	0.021	0.104	6.12
NASDAQ/NYSE	>336	0.52	0.28	0.20	0.027	0.110	6.54
FiQA	≤96	0.42	0.48	0.10	0.034	0.091	5.98
FiQA	>336	0.31	0.55	0.14	0.039	0.097	6.35
Reddit Finance	≤96	0.38	0.46	0.16	0.042	0.108	6.35
Reddit Finance	>336	0.29	0.51	0.20	0.048	0.114	6.82
M4	≤96	0.63	0.25	0.12	0.025	0.096	6.04

As shown in Table IX, the learned fusion weights vary systematically with forecasting horizon and dataset modality. For short-term forecasting horizons, numerical features receive relatively higher weights, particularly in datasets such as NASDAQ/NYSE and M4, where recent quantitative signals are highly informative. For longer forecasting horizons, the framework gradually increases the contribution of textual and multimodal signals, which capture broader sentiment and contextual patterns not fully reflected in short-term numeric trajectories. In multimodal settings such as VaTeX, the learned weights allocate a higher share to visual-text fusion, confirming that visual cues become more relevant when numeric indicators alone are insufficient. Importantly, the fusion coefficients do not collapse to trivial extremes, indicating that the model learns a balanced, horizon-dependent integration strategy rather than over-specializing to a single modality.

The ablation analysis therefore supports the 3 central innovation claims of the paper. First, AO contributes broad exploration and strong initialization. Second, PGRO improves refinement quality through validation-aware local search. Third, incremental learning with temporal adaptation improves non-stationary robustness and prevents catastrophic forgetting. In addition, the fusion-weight analysis confirms that the framework performs optimization-aware multimodal integration rather than static feature concatenation. Together, these results validate that the proposed gains arise from the coordinated contributions of the framework components rather than from any single module alone.

VI. DISCUSSION

The experimental results demonstrate that the proposed AO + PGRO-optimized multimodal forecasting framework achieves strong performance across heterogeneous datasets and forecasting scenarios. Beyond numerical improvements, it is important to understand the mechanisms behind these gains, the practical implications of the proposed framework, and its limitations. This section therefore discusses the observed performance behavior, identifies key limitations of the approach, and highlights ethical and practical considerations relevant to real-world deployment.

A. Performance Insights

The results presented in the experimental study show that the proposed framework consistently outperforms classical statistical models, recurrent neural networks, and transformer-based forecasting baselines across numerical, textual, and multimodal datasets. Several factors contribute to this improvement.

First, the AO + PGRO hybrid optimization strategy provides a balanced combination of global exploration and local refinement. The Aquila Optimizer (AO) performs structured exploration of the hyperparameter space, enabling the model to avoid poor initializations and local optima that commonly affect gradient-based training. The Proposed Guided Remora Optimization (PGRO) phase subsequently performs memory-guided local refinement, selecting candidate solutions based on validation performance. This 2-stage optimization process improves convergence stability and reduces sensitivity to hyperparameter initialization, which is particularly beneficial for multimodal models with high-dimensional parameter spaces.

Second, the framework benefits from multimodal feature integration. Financial forecasting often involves heterogeneous signals, including numerical market indicators, textual sentiment information, and contextual visual cues. Traditional forecasting models typically rely on a single modality, which limits their ability to capture complex market dynamics. By combining BERT-based textual embeddings, ViT-based visual features, and TCT-based temporal representations, the proposed framework captures complementary information from multiple data sources. The adaptive fusion mechanism further adjusts the contribution of each modality depending on forecasting horizon and dataset characteristics.

Third, the incremental learning mechanism enhances robustness under non-stationary conditions. Financial markets and social sentiment streams evolve continuously, making static training approaches prone to concept drift. The integration of Elastic Weight Consolidation (EWC), replay buffering, and dynamic thresholding allows the model to update its knowledge while preserving previously learned patterns. As a result, the framework maintains stable prediction behavior even when new market conditions emerge.

Finally, the combination of transformer-based temporal modeling and validation-driven optimization contributes to improved generalization across datasets. Instead of focusing solely on training accuracy, the proposed framework prioritizes validation-based model selection. This reduces the risk of overfitting and encourages configurations that generalize well to unseen data. The consistent improvements observed across all datasets suggest that the proposed approach improves forecasting stability rather than achieving isolated accuracy gains.

B. Limitations

Despite the encouraging results, several limitations of the proposed framework should be acknowledged.

1) Computational complexity

The use of transformer-based encoders combined with population-based metaheuristic optimization introduces significant computational overhead during training. Although AO + PGRO optimization is performed offline during model configuration, the training process still requires substantial GPU resources. This may limit the feasibility of the framework in environments with restricted computational infrastructure.

2) Limited dataset coverage

While the framework is evaluated on multiple heterogeneous datasets, the coverage within each modality remains limited. The datasets used in this study provide representative examples of numerical time series, financial text, and multimodal inputs, but they do not fully capture the diversity of financial markets or social information sources. Additional evaluation on broader datasets and different market conditions would further strengthen the generalizability of the proposed approach.

3) Absence of explicit domain adaptation mechanisms

The current framework relies on pretrained encoders and validation-driven optimization to achieve cross-domain robustness. However, it does not include dedicated domain adaptation techniques. When transferring the model to substantially different domains or financial environments, performance may depend on the similarity between training and deployment datasets. Integrating explicit domain adaptation strategies could further improve cross-domain generalization.

4) Limited investigation of optimization-specific effects

Although the AO + PGRO optimization strategy improves model performance and convergence stability, the present study does not include a dedicated analysis isolating the effect of PGRO on overfitting reduction. Future research could conduct controlled experiments comparing training–validation performance gaps with and without PGRO to more precisely quantify its role in improving generalization.

5) Model scalability and deployment constraints

The framework is designed primarily for methodological validation rather than deployment-optimized inference. Large-scale real-time applications may require additional architectural simplification, model compression, or hardware-aware optimization. Techniques such as pruning, parameter sharing, or lightweight transformer architectures could help reduce computational requirements while preserving predictive performance.

C. Ethical and Practical Considerations

The integration of multimodal signals, particularly sentiment information derived from social media and financial news, raises several ethical and practical considerations.

1) Ethical considerations in sentiment data usage

Social media sentiment data may contain biases, misinformation, or coordinated manipulation. Such data can reflect unequal participation across demographic or economic groups and may amplify transient noise rather than genuine market signals. In the proposed framework, sentiment information is treated as an auxiliary feature rather than a primary forecasting driver. Structured numerical data remain the dominant source of predictive signals, which reduces the influence of potentially biased or noisy sentiment inputs.

The framework does not attempt to infer user intent, assign normative judgments, or generate financial recommendations. Instead, it produces probabilistic forecasts that should be interpreted within the broader context of domain expertise and external information.

2) Fairness and transparency

Automated forecasting systems can influence financial decision-making and may therefore raise concerns related to transparency and accountability. The proposed framework emphasizes transparency through explicit reporting of model architecture, optimization procedures, and validation protocols. These design choices support auditability and allow users to better understand how predictions are generated.

3) Privacy and data integrity

All sentiment datasets used in this study are publicly available and anonymized, and the experiments do not involve real-time data scraping or user-level data collection. Consequently, the study does not introduce direct privacy risks. However, future deployment scenarios that incorporate real-time social media data may require additional safeguards to ensure compliance with privacy regulations and responsible data usage.

4) Scalability and system integration

From a deployment perspective, the framework is intended as a decision-support tool rather than an autonomous decision-making system. In real-world financial environments, model predictions should be combined with human expertise and risk management mechanisms. Additionally, large-scale operational systems may require latency-aware model configurations, bounded update frequencies for continual learning, and safeguards against abrupt model behavior changes.

Future work may explore lightweight model variants, hardware-aware optimization strategies, and controlled incremental learning policies to enable safer integration into real-time financial analytics platforms.

VII. Conclusion

This study presented a generalization-aware multimodal forecasting framework that integrates transformer-based deep learning models with a structured two-stage metaheuristic optimization strategy, combining Aquila Optimization (AO) for global exploration and Policy Gradient Reinforcement Optimization (PGRO) for validation-driven local refinement. The proposed framework further incorporates incremental learning mechanisms, including Elastic Weight Consolidation and replay buffering, to maintain predictive stability in non-stationary time-series environments. Extensive experiments conducted across heterogeneous datasets—including structured financial time series, finance-domain textual sentiment, social media signals, macroeconomic indicators, and multimodal inputs—demonstrate that the proposed approach consistently improves forecasting performance compared to strong baseline models. The framework achieves lower MSE, MAE, and SMAPE across datasets while maintaining stable performance across both short-term and long-term forecasting horizons, indicating improved generalization rather than isolated accuracy gains. The results confirm that the structured collaboration between AO and PGRO, combined with multimodal feature integration and incremental learning, provides a robust and adaptable solution for complex real-world forecasting tasks.

Building on these encouraging results, several directions can be explored to further extend the proposed framework. First, future work can expand the experimental evaluation to larger and more diverse datasets spanning additional financial markets, economic indicators, and multimodal information sources. Such evaluation would further validate the robustness and cross-domain generalization capability of the AO-PGRO optimization framework.

Second, integrating explicit domain adaptation techniques could improve model transferability when the framework is applied to new markets or domains with significantly different data distributions.

Third, research on lightweight model architectures and compression techniques, such as pruning, parameter sharing, or efficient transformer variants, may help reduce computational cost and enable real-time deployment in resource-constrained environments.

Finally, future studies can conduct controlled ablation experiments and uncertainty-aware forecasting extensions to better quantify the role of PGRO in mitigating overfitting and to enhance the interpretability and reliability of forecasting outputs in high-stakes decision-support scenarios.

Data Availability Statement

NASDAQ & NYSE Historical Stock Data. Available from Yahoo *Finance* (https://finance.yahoo.com/) and *Kaggle* repositories. Usage is subject to Yahoo Finance's terms of service.

FiQA Sentiment Dataset. Publicly available at *FiQA Challenge* (https://sites.google.com/view/fiqa/) for financial opinion mining and sentiment analysis.

Financial PhraseBank. Publicly available from University *of Vaasa*, Finland (https://www.mv.helsinki.fi/home/leheikki/FinancialPhraseBank/).

Reddit Financial News. Extracted from Reddit's public posts via Pushshift API (https://pushshift.io/). Data usage must comply with Reddit's content policy.

M4 Forecasting Competition Dataset. Publicly available at *M4 Competition Official Website* (https://github.com/Mcompetitions/M4-methods).

VaTeX Multimodal Video-Text Dataset. Available from VaTeX *Project* (https://eric-xw.github.io/vatex-website/), licensed for research use only.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Hemasundara Reddy Lanka and Sarvani Anandarao; Methodology, Hemasundara Reddy Lanka and Nagha Harish Vundavalli; Software, Nagha Harish Vundavalli; Validation, Hemasundara Reddy Lanka, Nagha Harish Vundavalli, and Nagaraju Devarakonda; Formal Analysis, Hemasundara Reddy Lanka; Investigation, Nagha Harish Vundavalli; Resources, Nagaraju Devarakonda; Data Curation, Nagha Harish Vundavalli; Writing—Original Draft Preparation, Hemasundara Reddy Lanka; Writing—Review & Editing, Sarvani Anandarao and Nagaraju Devarakonda; Supervision, Sarvani Anandarao. All authors had approved the final version.

References

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel *et al.*, *Time Series Analysis: Forecasting and Control*, Hoboken: Wiley, 2015.

[2] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018.

[3] C. Chatfield and H. Xing, *The Analysis of Time Series: An Introduction with R*, Boca Raton: CRC Press, 2019.

[4] R. Hyndman, A. Koehler, K. Ord *et al.*, *Forecasting with Exponential Smoothing: The State Space Approach*, Heidelberg: Springer Berlin, 2008.

[5] J. G. D. Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443–473, 2006.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint, arXiv: 1803.01271, 2018.

[8] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," in *Proc. 31st Conf. on Neural Information Processing Systems (NIPS 2017)*, 2017, pp. 1–11.

[9] H. Zhou, S. Zhang, J. Peng *et al.*, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. the AAAI Conf. on Artificial Intelligence*, 2021, pp. 111061–11115.
[10] H. Wu, J. Xu, J. Wang *et al.*, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. 35th Conf. on Neural Information Processing Systems (NeurIPS 2021)*, 2021, pp. 22419–22430.
[11] T. Zhou, Z. Ma, Q. Wen *et al.*, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proc. 39th International Conf. on Machine Learning*, 2022, pp. 27268–27286.
[12] I. J. Goodfellow, M. Mirza, D. Xiao *et al.*, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," arXiv preprint, arXiv: 1312.6211, 2015.
[13] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020.
[14] G. Lai, W. C. Chang, Y. Yang *et al.*, "Modeling long-and short-term temporal patterns with deep neural networks," in *Proc. 41st International ACM SIGIR Conf. on Research & Development in Information Retrieval*, 2018, pp. 95–104.
[15] F. Petropoulos, D. Apiletti, V. Assimakopoulos *et al.*, "Forecasting: Theory and practice," *International Journal of Forecasting*, vol. 38, no. 3, pp. 705–871, 2022.
[16] L. Abualigah, D. Yousri, M. A. Elaziz *et al.*, "Aquila optimizer: A novel meta-heuristic optimization algorithm," *Computers & Industrial Engineering*, vol. 157, 107250, 2021.
[17] S. Anandarao and S. H. Chellasamy, "Nature inspired-based remora optimisation algorithm for enhancement of density peak clustering," *Cogent Engineering*, vol. 10, no. 2, 2278259, 2023.
[18] J. Kirkpatrick, R. Pascanu, N. Rabinowitz *et al.*, "Overcoming catastrophic forgetting in neural networks," in *Proc. the National Academy of Sciences*, 2017, pp. 3521–3526.
[19] V. Bhogade and B. Nithya. "Time series forecasting using transformer neural network," *International Journal of Computers and Applications*, vol. 46, no. 10, pp. 880–888, 2024.
[20] P. Singh, M. Jha, M. Sharaf *et al.*, "Harnessing a hybrid CNN-LSTM model for portfolio performance: A case study on stock selection and optimization," *IEEE Access*, vol. 11, pp. 104000–104015, 2023.
[21] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a CNN-LSTM approach," *Expert Systems with Applications*, vol. 157, 113481, 2020.
[22] S. Balamohan and V. Khanaa, "Revolutionizing stock market prediction: Harnessing the power of LSTM and CNN hybrid approach," *SN Computer Science*, vol. 5, 919, 2024.
[23] T. B. Shahi, A. Shrestha, A. Neupane *et al.,* "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, 1441, 2020.
[24] S. Xiang, D. Cheng, C. Shang *et al.*, "Temporal and heterogeneous graph neural network for financial time series prediction," in *Proc. 31st ACM International Conf. on Information & Knowledge Management*, 2022, pp. 3584–3593.
[25] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning in finance," arXiv preprint, arXiv:1602.06561, 2018.
[26] M. Wen, P. Li, L. Zhang *et al.*, "Stock market trend prediction using high-order information of time series," *IEEE Access*, vol. 7, pp. 28299–28308, 2019.
[27] Q. Zhang, C. Qin, Y. Zhang *et al.*, "Transformer-based attention network for stock movement prediction," *Expert Systems with Applications*, vol. 202, 117239, 2022.
[28] S. Carta, A. S. Podda, D. R. Recupero *et al.*, "Explainable AI for financial forecasting," in *Proc. International Conf. on Machine Learning, Optimization, and Data Science*, 2021, pp. 51–69.
[29] K. Noor and U. Fatima, "Meta learning for rapid adaptation in financial time series forecasting," *IEEE Access*, vol. 13, pp. 24158–24170, 2025.
[30] H. Lu and X. Hu. "Enhancing financial risk prediction for listed companies: A catboost-based ensemble learning approach," *Journal of the Knowledge Economy*, vol. 15, pp. 9824–9840, 2023.
[31] X. D. Zhang, A. Li, and R. Pan, "Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine," *Applied Soft Computing*, vol. 49, pp. 385–398, 2016.
[32] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
[33] T. Loughran and B. McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
[34] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.
[35] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178, 115019, 2021.
[36] M. Sardelich and S. Manandhar, "Multimodal deep learning for short-term stock volatility prediction," arXiv preprint, arXiv: 1812.10479, 2018.
[37] U. Naseem, I. Razzak, K. Musial *et al.*, "Transformer based deep intelligent contextual embedding for twitter sentiment analysis," *Future Generation Computer Systems*, vol. 113, pp. 58–69, 2020.
[38] J. Jang, T. Kim, and S. H. Park. "Stock index forecasting using an explainable TAFT model with online data-driven social sentiment index," in *Proc. 5th ACM International Conf. on AI in Finance*, 2024, pp. 787–794.
[39] R. Hu, J. Yi, A. Chen *et al.*, "Multichannel cross-modal fusion network for multimodal sentiment analysis considering language information enhancement," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 7, pp. 9814–9824, 2024.
[40] S. Gite, H. Khatavkar, K. Kotecha *et al.*, "Explainable stock prices prediction from financial news articles using sentiment analysis," *PeerJ Computer Science*, vol. 7, e340, 2021.
[41] Ardisurya and M. Rizkinia, "Implementation of diffusion variational autoencoder for stock price prediction with the integration of historical and market sentiment data," *International Journal of Electrical, Computer, and Biomedical Engineering*, vol. 2, no. 2, pp. 229–242, 2024.
[42] M. C. Lee, J. W. Chang, S. C. Yeh *et al.*, "Applying attention-based BiLSTM and technical indicators in the design and performance analysis of stock trading strategies," *Neural Computing and Applications*, vol. 34, pp. 13267–13279, 2022.
[43] H. Liu, "Leveraging financial news for stock trend prediction with attention-based recurrent neural network," arXiv preprint, arXiv:1811.06173, 2018.
[44] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction," *Journal of Big data*, vol. 8, 17, 2021.
[45] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
[46] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Conf. on Neural Information Processing Systems* (*NIPS 2017*), 2017, pp. 1–10.
[47] F. D. Velez and Been Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint, arXiv:1702.08608, 2017.
[48] S. Kumar, M. Vishal, and V. Ravi, "Explainable reinforcement learning on financial stock trading using SHAP," arXiv preprint, arXiv:2208.08790, 2022.
[49] S. Park and J. S. Yang, "Interpretable deep learning LSTM model for intelligent economic decision-making," *Knowledge-Based Systems*, vol. 248, 108907, 2022.
[50] R. Liang, W. Zhang, and H. Ye, "Interpretable deep learning based text regression for financial prediction," *Expert Systems*, vol. 40, no. 9, e13368, 2023.
[51] R. Jain and R. Vanzara, "Emerging trends in AI-based stock market prediction: A comprehensive and systematic review," *Engineering Proceedings*, vol. 56, no. 1, 254, 2023.
[52] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proc. 34th International Conf. on Machine Learning*, 2017, pp. 1885–1894.
[53] T. Hospedales, A. Antoniou, P. Micaelli *et al.*, "Meta-learning in neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5149–5169, 2022.
[54] D. Ari and B. B. Alagoz, "DEHypGpOls: A genetic programming with evolutionary hyperparameter optimization and its application for stock market trend prediction," *Soft Computing*, vol. 27, pp. 2553–2574, 2022.

[55] J. Černevičienė and A. Kabašinskas, "Explainable Artificial Intelligence (XAI) in finance: A systematic literature review," *Artificial Intelligence Review*, vol. 57, 216, 2024
[56] C. Wong and M. Versace, "CARTMAP: A neural network method for automated feature selection in financial time series forecasting," *Neural Computing and Applications*, vol. 21, pp. 969–977, 2012.
[57] M. Elhoseny, N. Metawa, G. Sztano *et al.*, "Deep learning-based model for financial distress prediction," *Annals of Operations Research*, vol. 345, pp. 885–907, 2025.
[58] K. Lei, B. Zhang, Y. Li *et al.*, "Time-driven feature-aware jointly deep reinforcement learning for financial signal representation and algorithmic trading," *Expert Systems with Applications*, vol. 140, 112872, 2020.
[59] V. C. S. S. and A. H. S., "Nature inspired meta heuristic algorithms for optimization problems," *Computing*, vol. 104, pp. 251–269, 2022.
[60] H. I. Fawaz, G. Forestier, J. Weber *et al.*, "Deep learning for time series classification: A review," *Data Mining and Knowledge Discovery*, vol. 33, pp. 917–963, 2019.
[61] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, 106181, 2020.
[62] M. Ozcalci and M. Kilic, "GA-LDA approach for topic modeling in Turkish accounting and finance articles: Performance optimization in text classification," *Spectrum of Operational Research*, vol. 2, no. 1, pp. 305–322, 2025.
[63] U. Singh, K. Saurabh, N. Trehan *et al.*, "GA-LSTM: Performance optimization of LSTM driven time series forecasting," *Computational Economics*, vol. 66, pp. 2873–2908, 2025.