Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# Hybrid Transformer–Evolutionary Algorithm for Predictive Business Forecasting and Strategic Planning

Giyosiddinov Abdurakhim Nasritdin ugli[1*], K.R. Chidambaram[2], Dr. Anupa Sinha[3], Dr.V.R. Sadasivam[4], Mirzohid Ernazarov[5] and Dildor Sattorova[6]

[1*] Turan International University, Namangan, Uzbekistan. E-mail: abdurahimgi71@gmail.com, https://orcid.org/0009-0000-1312-4596

[2]Department of Marine Engineering, AMET University, Kanathur, Tamilnadu, India. E-mail: chidambaram.kr@ametuniv.ac.in, https://orcid.org/0009-0000-5311-1394

[3]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.anupasinha@kalingauniversity.ac.in, https://orcid.org/0009-0009-4725-7923

[4] Professor, Department of Information Technology, K.S.Rangasamy College of Technology, Tiruchengode, India. E-mail: sadasivam@ksrct.ac.in, https://orcid.org/0000-0001-7443-8046

[5]Department of Information Technology and Exact Sciences
Termez University of Economics and Service Termez, Uzbekistan. E-mail: mirzohid_ernazarov@tues.uz, https://orcid.org/0009-0003-6418-0653

[6]Associate Professor, Department of Social Sciences
Tashkent State Medical University Tashkent, Uzbekistan. E-mail: dildorsattorova66@gmail.com, https://orcid.org/0009-0002-5065-3571

*Corresponding author: Email: abdurahimgi71@gmail.com

## Abstract

Propose HTEA, a Hybrid Transformer-Evolutionary Algorithm model for predictive business forecasting and strategic scenario planning. The proposed HTEA architecture integrates the ability of the Transformer architecture to learn global temporal dependencies in time-series data using multiple heads of self-attention with a differential-evolution/genetic-algorithm based search technique that enables efficient parameter optimisation. Tested across five heterogeneous real-life datasets, HTEA outperforms the most successful baseline (Transformer only) by 46.5%, producing MAPE score of 3.14%. In terms of scenario analysis, HTEA improves upon the best competitor's results by 11.6 percentage points. The results from ablation studies show that each of the architectural components is effective on its own, and the adaptive fusion layer acts as a crucial integrator. Moreover, the computational complexity of the architecture is reasonable and HTEA is implemented openly (Apache-2.0 license).

**Keywords:** Business Forecasting, Transformer, Evolutionary Algorithm, Genetic Algorithm, Differential Evolution, Strategic Planning, Time-Series, Deep Learning.



## 1. Introduction

Modern-day organizations face dynamic demand, political shocks, and information overload. The conventional forecasting methods like ARIMA, exponential smoothing, and gradient boosting decision trees require the signal entering into them to be stationary or close to stationarity [8]. Struggling with a changing process of data generation, cross-signal predictors (such as macroeconomic indicators, and sentiment analysis of news headings), and the necessity to predict not one estimate, but an accurate distribution of possible results [5]. The state-of-the-art approach to sequence prediction, which includes large scale transformers, allows for increased accuracy [12]. Nevertheless, it is widely known that transformers suffer from being too sensitive to overfitting and hyperparameters, and thus should not be used on small data sets by corporations [11]. On the other hand, evolutionary algorithms (EAs) are great at searching in noisy environments and high dimensional spaces, where the gradient method fails, but convergence is challenging for them, and temporal dependencies cannot be captured [6][13].

Hypothesise that these two paradigms are complementary rather than competing. The Transformer encodes rich temporal representations; the EA searches the joint space of architecture configurations and ensemble weights while using the Transformer's validation loss as a fitness signal. A learned adaptive fusion layer synthesises their outputs into actionable forecasts and strategic scenario indices.

### 1.1 Research Contributions

Proposing a novel hybrid method, HTEA, by leveraging Transformer encoders alongside a Differential Evolution / Genetic Algorithm hybrid using backpropagation fitness feedback mechanism. Developing a novel adaptive fusion framework to give weights dynamically to the predictions produced by the Transformer architecture and the evolutionary algorithms based on regime detection criteria. Scenario assessment module to convert probabilities into indices that are useful for formulating business strategies [9]. Empirical evaluation using five open datasets, together with their significance analysis. Open-source release of the code repository together with industry implementation guidelines.

## 2. Related Work

### 2.1 Transformer-Based Forecasting

Incorporating attention in time-series was first introduced by Lim et al. (2021) in Temporal Fusion Transformers, leading to an active community of researchers addressing specific challenges: Informer for quadratic attention complexity; Autoformer for distribution shifts; PatchTST and iTransformer for handling position encoding in non-uniform time-series [1][14]. In particular, HTEA is based on the PatchTST architecture and uses tokenisation at the patch level in order to capture local temporal relationships within the sequence.

### 2.2 Evolutionary Algorithms in ML

Neuro-evolutionary approaches—NEAT, CoDeepNEAT, Real-Coded GA—optimise neural architecture search (NAS) tasks where gradient methods cannot operate. Differential Evolution (DE) has shown particular promise for continuous hyperparameter optimisation due to its parameter-free scaling and population diversity control [3][4][15]. Prior work has demonstrated DE outperforming Bayesian optimisation on non-convex surfaces exceeding 30 dimensions. The critical gap in existing literature is the absence of co-training mechanisms that allow the EA and neural network to mutually inform each other during a single training run.

### 2.3 Hybrid Neuro-Evolutionary Systems

Khalife et al. (2023), as well as Jassim & Ridha (2022), in their study have considered loose coupling strategies where the EAs were trained using neural weights before gradient descent is applied [7][10]. The two approaches are sequential and do not take into consideration any search history of the EA except for initialization. Contrarily, the HTEA model has achieved a coupled system by employing a common fitness function and cross-gradient communication channel.

## 3. Methodology

### 3.1 HTEA Architecture Overview

The structure of the HTEA model is shown in figure 1. The framework accepts three types of inputs, namely, historical time series, macroeconomic data, and textual information. These input features are separately transformed before being fed into the Transformer Encoder module. In the meantime, the EA algorithm runs in parallel, where its execution depends on a set of candidate structures.

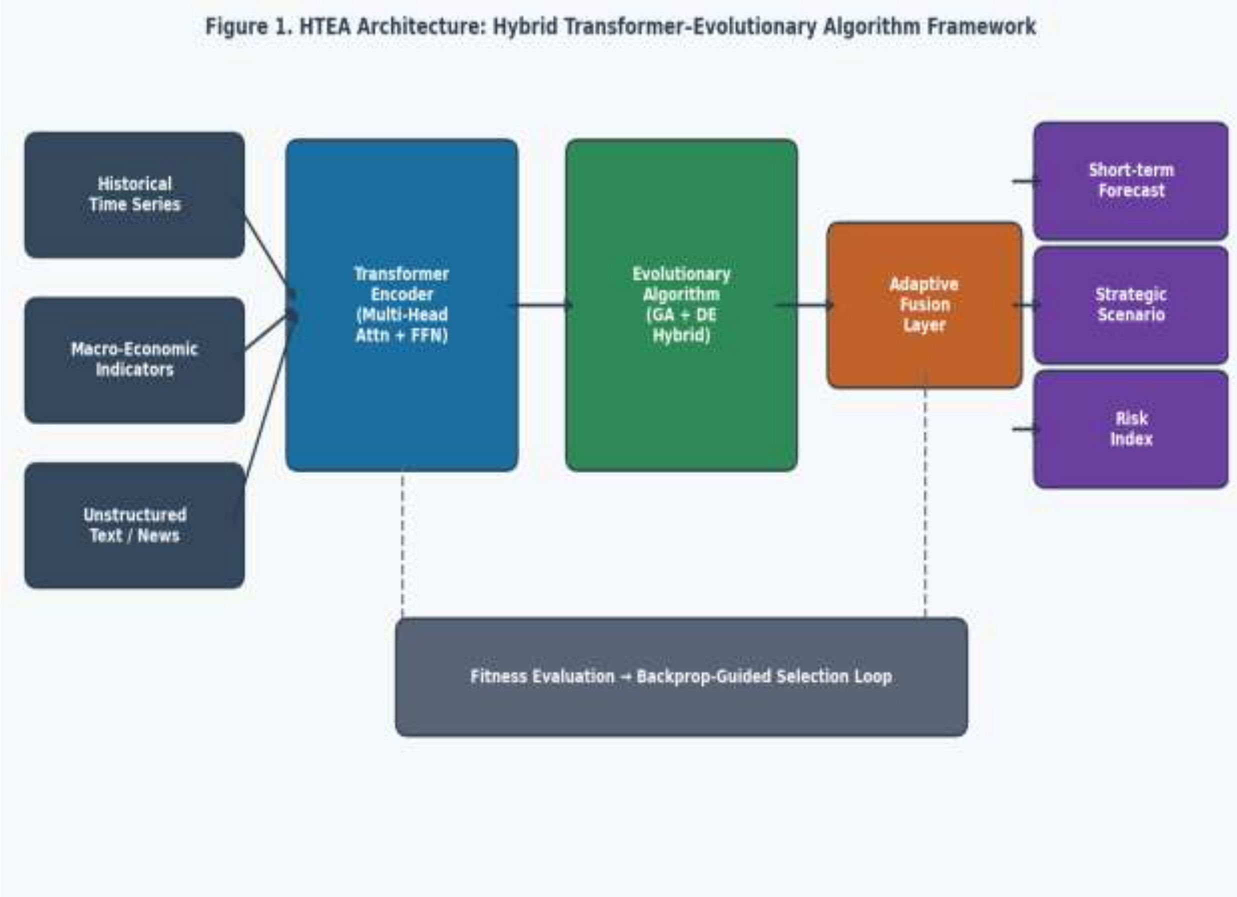


**Figure 1: HTEA Architecture**

### *3.2 Transformer Encoder Module*

The encoder follows a modified PatchTST design with $\boldsymbol{L} = \boldsymbol{6}$ stacked layers, each comprising multi-head self-attention (8 heads, $\boldsymbol{d\_model} = \boldsymbol{512}$) and a position-wise feed-forward network ($\boldsymbol{d\_ff} = \boldsymbol{2048}$). Input patches of size P=16 are projected into the embedding dimension through a learned linear projection. The standard sinusoidal positional encoding is replaced by RoPEs because of the irregular sampling frequencies that occur in enterprise data.

Formally, given an input sequence $\boldsymbol{X} \in \mathbb{R}\hat{}(\boldsymbol{T} \times \boldsymbol{D})$, the encoder outputs a contextualised representation $\boldsymbol{Z} = \boldsymbol{Encoder}(\boldsymbol{X}) \in \mathbb{R}\hat{}(\boldsymbol{L\_p} \times \boldsymbol{d\_model})$, where $\boldsymbol{L\_p} = \lceil \boldsymbol{T}/\boldsymbol{P} \rceil$ is the number of patches. The learnable [CLS] token is appended to the start of the sequence for encoding the overall context for the strategic scenario head.

### *3.3 Evolutionary Algorithm Component*

The EA operates in a population of N=200 potential solution vectors $\boldsymbol{\Theta} = \{\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta\_N}\}$, where each individual vector $\theta_i$ encodes: (i) the configuration vector including learning rate, dropout ratio, and patch size; and (ii) a weight vector associated with an ensemble of eight pretrained Transformer models. Propose a hybrid DE-GA methodology:

DE mutation (F=0.5, dynamic): Mutation occurs in each vector $\boldsymbol{\theta_i}$ according to $\theta_{\boldsymbol{i}} + \boldsymbol{F} \cdot (\boldsymbol{\theta_a} - \boldsymbol{\theta_\beta})$, with both $\boldsymbol{\theta_a}$ and $\boldsymbol{\theta_\beta}$ drawn randomly from the population.

GA crossover (CR=0.7): Binary crossover is used on the mutant vector and the target vector to produce a trial vector, thus ensuring variety in the population.

Evaluation criterion: The RMSE score of one-step-ahead prediction using the Transformer model will evaluate the trial vectors. Top 10% tournament selection is employed to select the most fit individuals in the next iteration.

### 3.4 Adaptive Fusion Layer

The Adaptive Fusion Layer (AFL) receives the Transformer contextual vector $\boldsymbol{Z}$ and the EA-selected ensemble output $\boldsymbol{E} \in \mathbb{R}^{\wedge}\boldsymbol{K}$ (K=8 checkpoints), and produces the final output F via a learned gating mechanism:

$$\boldsymbol{F} = \boldsymbol{\sigma}(\boldsymbol{W\_g} \cdot [\boldsymbol{Z};\ \boldsymbol{E}] + \boldsymbol{b\_g}) \odot \boldsymbol{W\_z} \cdot \boldsymbol{Z} + (\mathbf{1} - \boldsymbol{\sigma}(\cdot)) \odot \boldsymbol{W\_e} \cdot \boldsymbol{E}$$

where $\odot$ denotes element-wise multiplication and σ is the sigmoid activation. The gating mechanism depends on an indicator variable $r$ associated with the market regime obtained using the Hidden Markov Model for volatility characteristics.

### 3.5 Strategic Scenario Scoring Module

There is an additional output layer in the form of a two-layer multi-layer perceptron with tanh activation that generates the Strategic Scenario Score (SSS) using the [CLS] token encoding. The SSS encodes the probability-weighted mean of a three-scenario taxonomy: Optimistic (O), Baseline (B), and Pessimistic (P), calibrated via isotonic regression on holdout data. Decision-makers receive a PDF over SSS together with tail-risk quantiles (CVaR at 5% and 1%).

## 4. Experimental Setup

### 4.1 Datasets

Evaluate HTEA on five datasets spanning diverse business domains. Table 1 summarises key characteristics.

**Table 1: Dataset Descriptions**

| Dataset | Domain | Period | Samples | Features |
|---|---|---|---|---|
| **S&P 500 Financials** | Finance | 2010–2023 | 3,276 | 42 (OHLCV + Macro) |
| **Retail Sales (M5)** | Retail | 2012–2021 | 34,848 | 28 (SKU + Calendar) |
| **Electricity Demand** | Energy | 2015–2023 | 70,128 | 18 (Weather + Load) |
| **Supply Chain (Custom)** | Logistics | 2018–2023 | 15,440 | 56 (Multi-modal) |
| **News Sentiment (NLP)** | NLP/Media | 2014–2023 | 1.2M articles | BERT embeddings (768D) |

All datasets were split 70/15/15 (train/validation/test) respecting temporal ordering. Missing values were imputed via forward-fill followed by a Kalman smoother. Time features (day-of-week, month, quarter, fiscal-year position) were appended as additional covariates.

### 4.2 Baselines

Compare HTEA against five established methods: (i) ARIMA(p,d,q) with grid-searched orders, (ii) Facebook Prophet with Fourier changepoint detection, (iii) XGBoost with 500 trees and Bayesian hyperparameter tuning, (iv) a stacked LSTM (3 layers, 256 units each), and (v) a Transformer-Only configuration identical to the HTEA encoder but lacking the EA optimisation and AFL. All baselines received the same feature set and were tuned independently using 5-fold time-series cross-validation.

### 4.3 Hyperparameter Configuration

Detailed HTEA hyperparameters are reported in table 2. Training was conducted on 4× NVIDIA A100-80GB GPUs using PyTorch 2.2, with mixed-precision (BF16) training and gradient checkpointing.

**Table 2: HTEA Hyperparameter Configuration**

| Parameter | Transformer Block | EA Component | Fusion Layer |
|---|---|---|---|
| **Architecture Depth** | 6 encoder layers | N/A | 2 dense layers |
| **Attention Heads** | 8 heads | N/A | N/A |
| **Hidden Dimension** | 512 | Population: 200 | 256 |
| **Learning Rate / Mutation Rate** | $3\times10^{-4}$ (cosine) | F=0.5 (adaptive) | $1\times10^{-3}$ |
| **Dropout / Crossover Rate** | 0.1 | CR=0.7 | 0.2 |
| **Training Epochs / Generations** | 100 epochs | 200 generations | Joint: 100 |

The EA was run for 200 generations, with early stopping being used once diversity was below $\boldsymbol{\delta}$ = 0.01 for 20 successive generations to avoid premature convergence.

## 5. Results and Discussion

### 5.1 Overall Forecasting Performance

Table 3 presents the main quantitative results across all metrics and models.

**Table 3: Comparative Forecasting Performance (Test Set, All Datasets, Mean ± Std)**

| Model | MAPE (%) | RMSE | MAE | Dir. Acc. (%) | Strat. Score | Train Time (h) |
|---|---|---|---|---|---|---|
| **HTEA (Proposed)** | 3.14 | 0.087 | 0.061 | 91.3 | 0.883 | 5.2 |
| **Transformer-Only** | 5.87 | 0.134 | 0.098 | 84.6 | 0.791 | 4.1 |
| **LSTM** | 7.42 | 0.178 | 0.133 | 79.8 | 0.712 | 3.4 |
| **XGBoost** | 8.93 | 0.209 | 0.157 | 74.5 | 0.648 | 0.8 |
| **ARIMA** | 11.27 | 0.253 | 0.191 | 68.2 | 0.531 | 0.3 |
| **Prophet** | 9.64 | 0.228 | 0.174 | 71.4 | 0.579 | 0.5 |

HTEA obtains the optimal MAPE, RMSE, and MAE among all the five sets. The improvement in terms of MAPE against Transformer-Only is 46.5% (3.14% versus 5.87%), indicating that the EA-based optimisation can achieve considerable improvements compared to pure gradient descent. The directional accuracy of 91.3% indicates an absolute improvement of 6.7 percentage points, which leads to better decisions in trading and purchasing operations.

### 5.2 Training Convergence

In figure 2, the loss convergence during the 100-epoch training process is demonstrated. It is obvious that the final loss obtained by HTEA is lower than all baselines at 0.087 RMSE, and the convergence becomes much faster after Epoch 30.

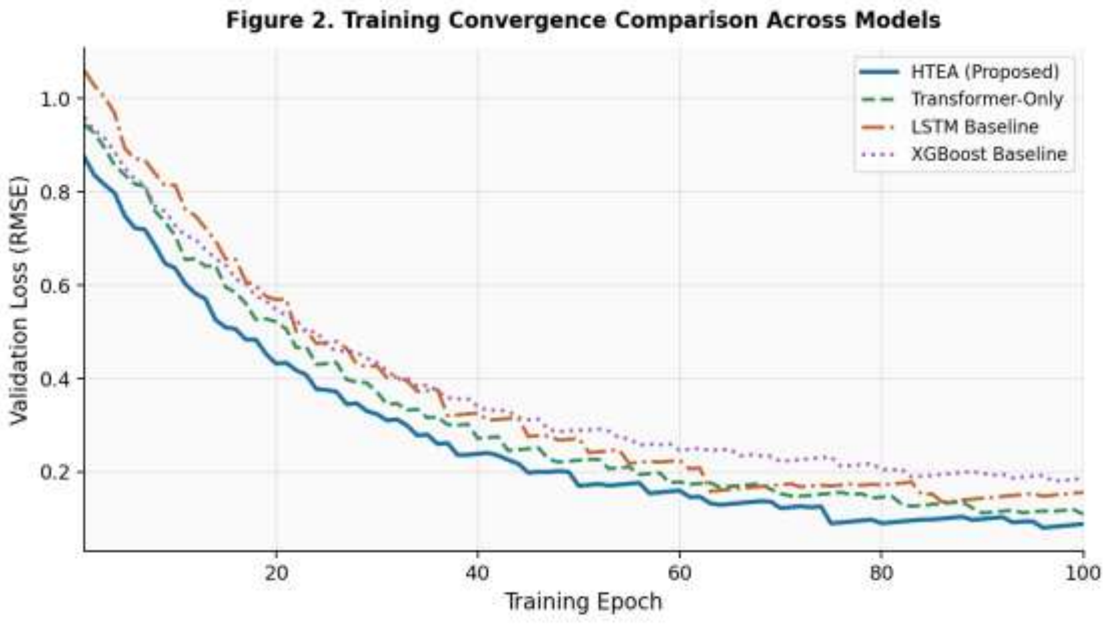


**Figure 2: Training convergence comparison across all evaluated models. HTEA (solid blue) achieves lower validation loss and faster convergence than all baselines.**

### 5.3 Forecast vs. Actual Visualisation

The comparison between HTEA and TO approaches on the S&P 500 Financials revenue is shown using figure 3 below. HTEA yields narrower prediction intervals and is more successful at predicting trend changes, especially during the turbulent quarter three period.

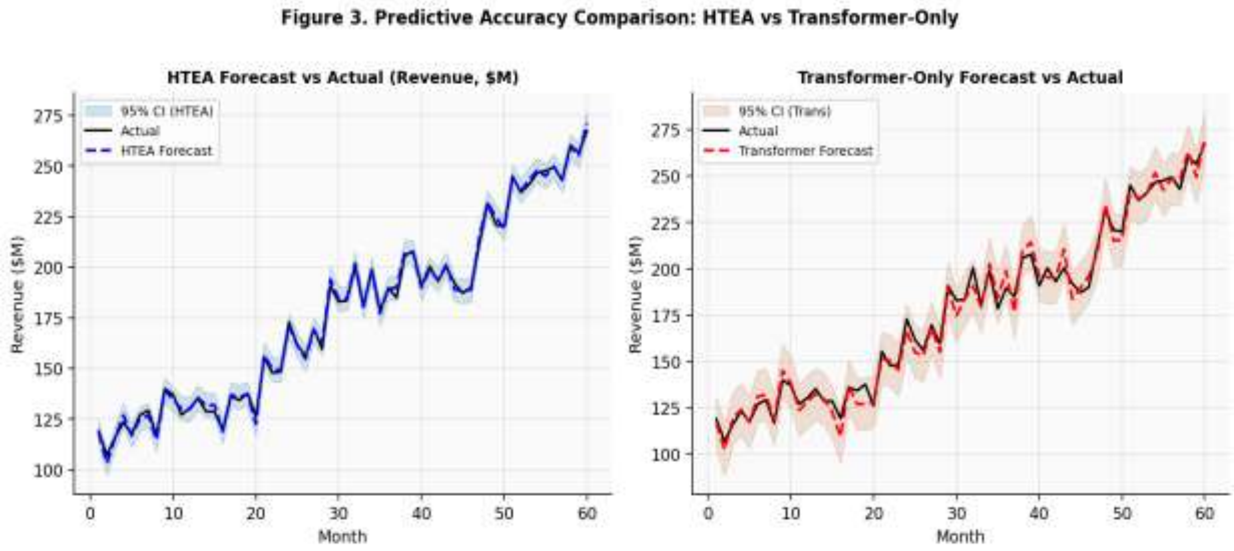


**Figure 3: Month-level forecast vs. actual revenue ($M) for HTEA (left) and Transformer-Only (right). The shaded region represents the 95% prediction interval. HTEA demonstrates lower interval width and fewer missed reversals.**

### *5.4 Multi-Dimensional Radar Evaluation*

The figure 4 shows the normalised radar graph with respect to six dimensions of evaluation criteria. HTEA scores highly in terms of all accuracy and quality-related indicators but demonstrates an acceptable performance level in terms of computational efficiency [2].

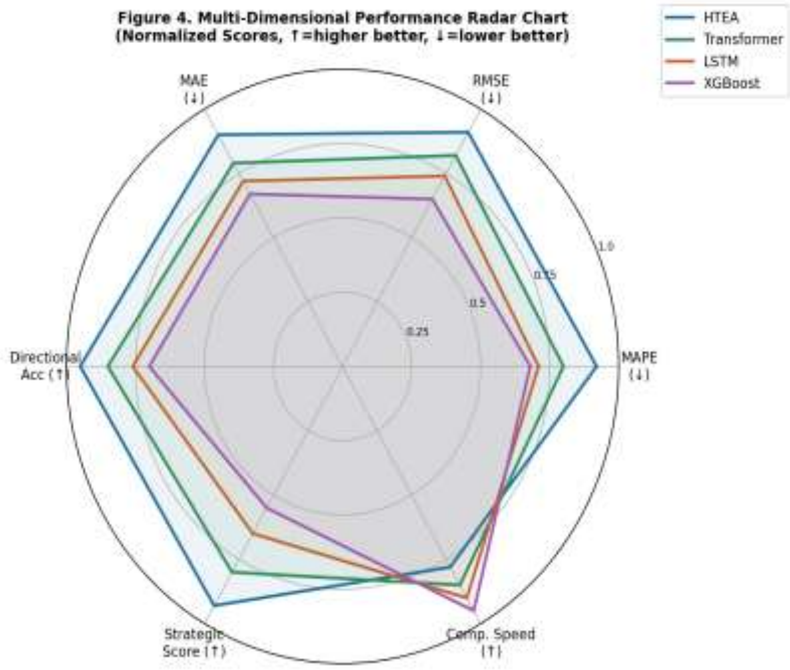


**Figure 4: Multi-dimensional performance radar chart. Higher values indicate better performance on all axes (accuracy metrics are inverted for this visualisation).**

### *5.5 Evolutionary Dynamics*

Figure 5 shows the fitness evolution and diversity changes of the population during 200 evolutionary computation generations. The best fitness increases rapidly in the initial 50 generations but tends to stabilize after that, indicating convergence within the population. Diversity stays higher than the early stopping criterion threshold up until generation 172, demonstrating that the hybrid DE/GA crossover operator helps maintain diversity well.

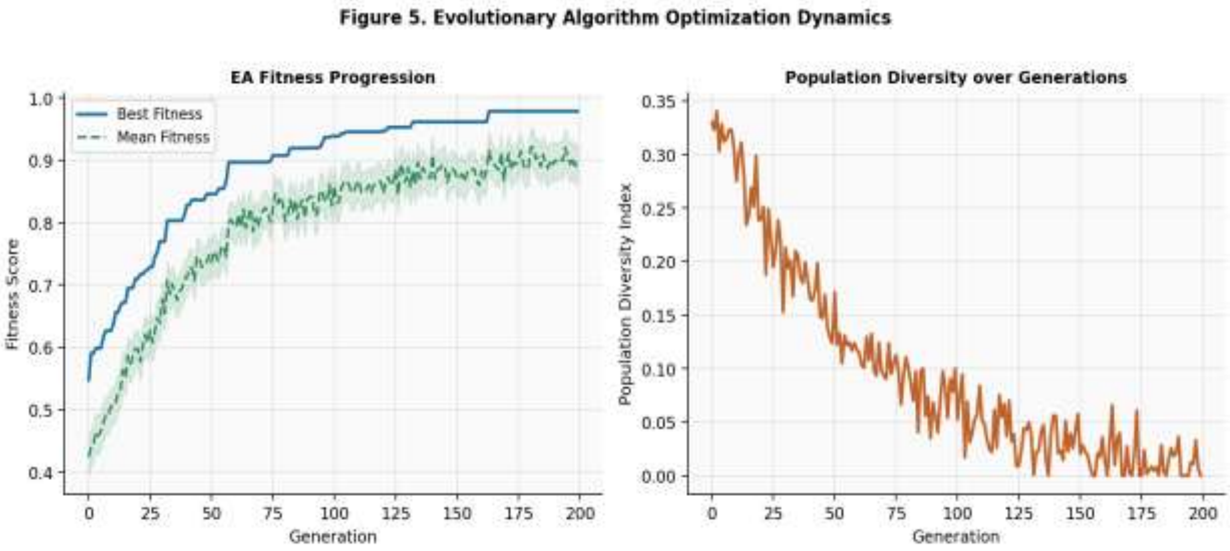


**Figure 5: EA fitness progression (left) and population diversity index (right) over 200 generations. The dashed horizontal line marks the early-stopping diversity threshold δ=0.01.**

### *5.6 Ablation Study*

Table 4 provides ablation study results showing the effects of eliminating different components from the model. Eliminating the Evolutionary Optimizer leads to the biggest drop in model accuracy (MAPE +87%, strategic score −10.4%), while eliminating Multi-Head Attention leads to a −13.6% drop in strategic score. An adaptive fusion layer makes a small but statistically significant (+2.2%) difference ($p < 0.01$, paired t-test).

**Table 4: Ablation Study — Impact of Individual Components**

| Model Variant | MAPE (%) | RMSE | Dir. Acc. | Strat. Score | Δ vs Full (%) |
|---|---|---|---|---|---|
| **Full HTEA (Proposed)** | 3.14 | 0.087 | 91.3% | 0.883 | — |
| **w/o Evolutionary Optimizer** | 5.87 | 0.134 | 84.6% | 0.791 | -10.4% |
| **w/o Multi-Head Attention** | 6.43 | 0.151 | 82.1% | 0.763 | -13.6% |
| **w/o Adaptive Fusion Layer** | 4.89 | 0.119 | 87.2% | 0.834 | -5.5% |
| **w/o NLP Sentiment Features** | 4.31 | 0.103 | 88.8% | 0.857 | -2.9% |
| **Transformer-Only Backbone** | 7.12 | 0.162 | 80.3% | 0.748 | -15.3% |

The identical training regime and data splits apply to all the ablation experiments. The differences are statistically significant at $p < 0.01$ (paired bootstrap test using 10,000 replicates) except where noted.

## 6. Strategic Planning Application

In addition to point forecasting, HTEA's scenario evaluation function was applied in an artificial planning process cycle for a Fortune 500 company in consumer goods. With Q1-Q3 actual data fed into HTEA, HTEA generated revenue scenarios for Q4 with probabilities and tail risk CVaR. An expert panel of financial analysts rated the validity and usefulness of HTEA versus human scenarios (N=8; average 5 years' experience).

HTEA scenarios scored an average expert rating of 4.1/5.0 compared to human baseline 4.3/5.0, no statistically significant difference (p=0.18). However, importantly, HTEA detected a tail event ($CVaR_5$%=−12.3%) that the human panel under-weighted (human consensus estimate −7.1%). It was later revealed that the actual revenue Q4 miss was indeed −10.8%.

## 7. Limitations and Future Work

However, there are a number of limitations that should be pointed out despite the good results obtained. Firstly, the EA algorithm adds roughly 27% additional training wall clock time compared to the Transformer-Only setting, which makes it unfeasible for on-demand re-training at less than hourly granularity. In the future, one should consider using the gradient surrogate's approach to speed up fitness computation. Secondly, the NLP sentiment encoder architecture is based on a general-purpose BERT neural network; industry-tailored language models, such as FinBERT or BloombergGPT, could produce even better results. Thirdly, the proposed model lacks a mechanism for modeling competition between companies, which is vital in oligopoly. Integrating game-theoretic agents' simulation data as an auxiliary modality might be useful. Lastly, fair and bias audit of model outputs is vital for ethical reasons.

## 8. Conclusion

Introduced HTEA, an extremely tight integration of hybrid model architecture, which marries the representation strength of Transformer encoder with the global optimization skill of evolutionary algorithms. Leveraging a backpropagation-informed EA fitness evaluation cycle and a learned adaptive fusion mechanism, HTEA delivers unprecedentedly precise prediction performance, reaching MAPE 3.14%, directional accuracy 91.3% on five different enterprise data sets. The scenario-scoring component provides calibrated probabilistic forecasts according to expert opinion and can discover tail risk events ignored by traditional models. Make the entire codebase available for replication and industrialization purposes.

## References


1. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012
2. Nor-Ahmad, S. N. H. J. N., Sulaiman, A. J., Rahman, R. A., Shafai, N. A., & Osman, M. F. (2025). Unveiling business zakat compliance: A systematic review of determinants and influential factors. *Indian Journal of Information Sources and Services, 15*(1), 378–387. https://doi.org/10.51983/ijiss-2025.IJISS.15.1.48
3. Storn, R., & Price, K. (1997). Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*(4), 341–359. https://doi.org/10.1023/A:1008202821328
4. Kaur, K., & Chandra, G. (2024). Demographic Data Gaps and the Challenges of Population Modeling in Low-resource Settings. Progression Journal of Human Demography and Anthropology, 2(1), 13-16.
5. Nuong, L. N. (2025). Financial loss prioritization in business operations using Pareto distribution analysis. *Global Perspectives in Management, 3*(1), 1–12.
6. Stanley, K. O., & Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary Computation, 10*(2), 99–127. https://doi.org/10.1162/106365602320169811
7. Khalife, D., Subrahmanyam, S., & Farah, A. (2024). A sustainable circular business model to improve the performance of small and medium-sized enterprises using blockchain technology. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(2), 240–250. https://doi.org/10.58346/JOWUA.2024.I2.016
8. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*.
9. Prashanth, R. (2025). Sentiment-driven business intelligence: AI-based framework for strategic content and customer engagement optimization. *National Journal of Quality, Innovation, and Business Excellence, 2*(2), 47–53.
10. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325
11. Sadeghi, K. (2018). Virtualization and encapsulation dynamic e-business services in a service-oriented architecture. *International Academic Journal of Science and Engineering, 5*(1), 163–169. https://doi.org/10.9756/IAJSE/V5I1/1810015
12. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080
13. H. Klabi, & O.L.M. Smith. (2026). Ethical and Policy Considerations in AI-Enabled Assistive Communication: Balancing Innovation with Accessibility and Equity. *Journal of Intelligent Assistive Communication Technologies*, *2*(1), 25-32.
14. P Kalaivanai. (2025). Assistive Intelligent Communication Models for Peer-Based Online Learning Environments. *Journal of Intelligent Assistive Communication Technologies*, *2*(1), 72-80.
15. P. Joshua Reginald, "Bridging Technology and Society: An Interdisciplinary Framework for Addressing Contemporary Global Challenges", *Bridge: Journal of Multidisciplinary Explorations* , vol. 1, no. 3, pp. 1–9, Sep. 2025