

Article

Not peer-reviewed version

Enhancing Exchange Rate Forecasting with Explainable Deep Learning Models

Shuchen Meng^{*}, Andi Chen, Chihang Wang, Mengyao Zheng, Fangyu Wu, [Xupeng Chen](#), [Haowei Ni](#), [Panfeng Li](#)

Posted Date: 10 October 2024

doi: 10.20944/preprints202410.0812.v1

Keywords: Exchange Rate Forecasting; LSTM; TimesNet; iTransformer; Grad-CAM



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Enhancing Exchange Rate Forecasting with Explainable Deep Learning Models

Shuchen Meng ^{1,*}, Andi Chen ², Chihang Wang ³, Mengyao Zheng ⁴, Fangyu Wu ⁵, Xupeng Chen ³, Haowei Ni ⁶ and Panfeng Li ⁷

¹ Central University of Finance and Economics, Beijing, China

² Independent Researcher, Beijing, China

³ New York University, New York, USA

⁴ Harvard University, Boston, USA

⁵ IEEE Student Member, New York, USA

⁶ Columbia University, New York, USA

⁷ University of Michigan, Ann Arbor, USA

* Correspondence: scmeng19@163.com

Abstract: Accurate exchange rate prediction is fundamental to financial stability and international trade, positioning it as a critical focus in economic and financial research. Traditional forecasting models often falter when addressing the inherent complexities and non-linearities of exchange rate data. This study explores the application of advanced deep learning models, including LSTM, CNN, and transformer-based architectures, to enhance the predictive accuracy of the RMB/USD exchange rate. Utilizing 40 features across 6 categories, the analysis identifies TSMixer as the most effective model for this task. A rigorous feature selection process emphasizes the inclusion of key economic indicators, such as China-U.S. trade volumes and exchange rates of other major currencies like the euro-RMB and yen-dollar pairs. The integration of grad-CAM visualization techniques further enhances model interpretability, allowing for clearer identification of the most influential features and bolstering the credibility of the predictions. These findings underscore the pivotal role of fundamental economic data in exchange rate forecasting and highlight the substantial potential of machine learning models to deliver more accurate and reliable predictions, thereby serving as a valuable tool for financial analysis and decision-making.

Keywords: exchange rate forecasting; LSTM; TimesNet; iTransformer; Grad-CAM

1. Introduction

Since the dissolution of the Bretton Woods system, the adoption of a floating exchange rate regime has introduced significant challenges in risk management for market participants. The volatility of the RMB/USD exchange rate, particularly during periods of trade tensions, has heightened the uncertainty faced by those engaged in the foreign exchange market. The People's Bank of China's reform of the exchange rate fixing mechanism on August 11, 2015, further increased the marketization of the RMB exchange rate, leading to greater exchange rate volatility and increased foreign exchange risk. This has underscored the critical need for accurate exchange rate forecasting and effective risk management strategies.

China's growing role in the global supply chain, especially after its accession to the World Trade Organization (WTO), and the deepening economic ties between China and the United States, have made the RMB/USD exchange rate a focal point of global economic stability. Debates over the valuation of the RMB, particularly during periods of significant trade surpluses with the U.S., have led to multiple rounds of discussions on exchange rate policy. The RMB exchange rate reform in 2005 and the subsequent rounds of monetary policy adjustments by the Federal Reserve, especially during the 2007 financial crisis, further complicated the dynamics of the RMB/USD exchange rate. The "8.11 Exchange Rate Reform" in 2015, which introduced a more flexible exchange rate mechanism, and the intensified trade frictions since 2018, have put additional depreciation pressure on the RMB, making accurate forecasting of the RMB/USD exchange rate increasingly important [1–5].

Previous research on exchange rate forecasting has primarily focused on theoretical and quantitative models. Theoretical models often emphasize the equilibrium state of exchange rates, which can be

difficult to achieve or maintain in practice, making short- to medium-term predictions particularly challenging. Quantitative models focus on the exchange rate's own dynamics while often neglecting other critical influencing factors. Moreover, these models have struggled to produce consistent results across different studies.

In recent years, there has been a notable shift towards using big data approaches in forecasting models, bypassing the need for complex mathematical modeling and allowing for more flexible model forms without predefined structures. The inherent complexity and non-linearity of exchange rate data have led to applying non-linear methods, such as chaos theory, non-parametric methods, and machine learning techniques, which have shown potential to improve forecasting accuracy. Studies like those of LeBaron and others have demonstrated that methods such as kernel ridge regression can significantly enhance the prediction of financial volatility, although some researchers, such as Mourer, have found that these methods do not always outperform simple autoregressive models in all contexts.

Since the end of the Bretton Woods system, the transition to a floating exchange rate regime has posed significant challenges for managing risk, especially regarding the RMB/USD exchange rate. China's integration into the global economy post-WTO accession, along with its deepening economic ties with the U.S., has made this exchange rate crucial for global economic stability. The 2015 reform of China's exchange rate mechanism increased market-driven fluctuations, further intensified by trade tensions, which has heightened volatility. This evolving landscape has led to debates over the RMB's valuation and spurred numerous policy discussions. Consequently, precise and reliable exchange rate forecasting has become essential for effective risk management in this volatile environment.

Traditional exchange rate forecasting models, both theoretical and quantitative, have struggled with consistency and often neglect critical influencing factors. For example, theoretical models may fail to account for the real-time impact of policy changes and global economic shifts, while quantitative models may not fully capture the non-linear dynamics and intricate interactions of the variables involved. Recently, there has been a shift toward big data and machine learning approaches, which offer flexibility and improved accuracy in handling the complex, non-linear nature of exchange rate data [6–20,20–22]. While some methods like kernel ridge regression have shown promise, their performance varies across different contexts, highlighting the ongoing challenges in exchange rate prediction.

Machine learning models, particularly deep learning models, have increasingly been applied to predicting time series and economic variables [23–39]. Despite their advantages in handling complex, non-linear data without requiring explicit assumptions about the underlying data distribution, these models are often criticized for their "black box" nature and lack of interpretability. Recent advancements, such as the application of Grad-CAM and attention mechanisms, have begun to address these issues, making it possible to visualize model predictions and understand the underlying decision-making processes. In related financial applications, researchers have combined advanced models like XGBoost with data balancing techniques including SMOTE to enhance predictive performance, as demonstrated by Chang et al. in their fraud detection study [40]. Their work not only showcased the effectiveness of this approach in fraud detection but also laid a foundation for developing robust models in various financial domains, including potentially exchange rate prediction [40]. However, applying these interpretability techniques has been mostly limited to fields like image recognition and natural language processing, with relatively few studies applying them to economic forecasting.

Given the challenges of traditional models and the potential of machine learning approaches, this study seeks to explore the use of advanced deep learning models, including CNNs, RNNs, MLPs, and transformer-based architectures, for predicting the RMB/USD exchange rate. By incorporating a comprehensive set of features—drawn from economic indicators, trade data, and other currency pairs—and employing advanced feature selection techniques, this research aims to enhance predictive accuracy, identify the most relevant factors influencing exchange rate fluctuations, and enhance the interpretability of the model predictions.

a) Contributions of This Study

- **Application of Deep Learning Models:** This study provides an initial analysis of the effectiveness of deep learning models in exchange rate prediction, using MSE and MAE as key metrics to identify the best-performing models.
- **Enhancement of Predictive Performance:** To improve the accuracy of machine learning models, this study employs various techniques, including feature selection, to reduce redundancy and retain the most relevant subset of features for exchange rate forecasting.
- **Analysis of Influential Factors Over Time:** By applying attention mechanisms, this study enhances the interpretability of machine learning models, offering insights into how different factors influence exchange rate predictions across different periods. This analysis aims to uncover which aspects of economic data the models prioritize during the prediction process, thereby providing a more nuanced understanding of the underlying dynamics.

2. Methods

2.1. Data Collection and Preprocessing

The data selected for this study is derived from three sources:

Macroeconomic Statistics: This includes key indicators like import/export values and short-term capital flows. While critical for decision-making, these statistics often suffer from delays, inaccuracies, and lack of predictive power. Branson's (1975) asset portfolio approach to exchange rates, integrating purchasing power parity and risk-return models, is hampered by such data lags. **Financial Market Trading Data:** Includes real-time data like stock indices, concept stock indices, and exchange rates. These data are timely, transparent, and reflect market expectations, making them valuable for exchange rate predictions. **Macroeconomic Variables:** Covers indicators like M2 (monetary policy), spot interest rates, and price indices (inflation). Most data are daily; monthly data are treated as daily. This practice reflects real-world usage, where outdated data often guide predictions. Building upon recent advancements in financial data processing, we incorporated techniques inspired by Chang's research [40] to significantly enhance our model's processing efficiency, particularly in handling diverse data frequencies and potential redundancies.

All data were sourced from the WIND Financial Terminal. Variables include the RMB/USD exchange rate and other relevant economic and financial indicators. Traditional regression models would struggle with multicollinearity and parameter estimation due to the extensive set of variables [41–64]. However, machine learning models mitigate these issues [65–73]. Post-2015, the volatility of the USD/RMB exchange rate increased, leading us to focus on this period [18,74–85]. We applied z-score normalization to stabilize the model training process.

Table 1. Factors Description Table.

Indicator	Description	Indicator Name
Exchange Rate	Offshore Spot RMB Closing Price	rate
Long-term Bond Yield Differential (U.S.)	10-Year Government Bond Yield - U.S. Federal Funds Rate	udr
CPI (U.S.)	U.S. Consumer Price Index	cpui
CSI 300 Index	CSI 300 Index	HS300
Import and Export Value	Current Month Import and Export Value	trade
Import Value	Current Month Import Value	inputu
M2 (U.S.)	M2 in the United States	m2u2
EUR/USD	EUR/USD Exchange Rate	EURUSD
AUD/USD	AUD/USD Exchange Rate	AUDUSD
USDIX	U.S. Dollar Index Published by ICE	USDIX

Table 2. Factors Description Table.

Category	Indicator	Description	Indicator Name
Fundamental Data	Exchange Rate	Offshore Spot RMB Closing Price	rate
	Interest Rate	U.S. Federal Funds Rate	rusa
	Interest Rate	PBoC Benchmark Deposit Rate	rchn
	Interest Rate Differential	Interest Rate Differential Between Two Countries	dr
	Long-term Bond Yield Differential (China)	10-Year Government Bond Yield - Bank Lending Rate	ydr
	Long-term Bond Yield Differential (U.S.)	10-Year Government Bond Yield - U.S. Federal Funds Rate	udr
	CPI (U.S.)	U.S. Consumer Price Index	cpiu
CPI (China)	China Consumer Price Index	cpic	
Consumer Confidence Index	China Consumer Confidence Index	ccp	
Stock Index Data	Dow Jones Index	Dow Jones Index	dowjones
	MSCI China Index	MSCI China Index	MSCIAAshare
	CSI 300 Index	CSI 300 Index	HS300
	S&P 500 China Index	S&P 500 China Index	sprd30.ci
	Nasdaq China Index	Nasdaq China Index	nyseche
	Hang Seng China Enterprises Index	Hang Seng China Enterprises Index	hscEI.hi
Current Account	Import and Export Value	Current Month Import and Export Value	trade
	Export Value	Current Month Export Value	output
	Import Value	Current Month Import Value	inputu
Capital Account	FDI	Foreign Direct Investment	fdix
	FDI	Actual Foreign Investment in the Current Month	fdi
	PI	Portfolio Investment	PI
	OI	Other Investment	OI
	Short-term International Capital Flow	Foreign Exchange Reserves Change in the Current Month Net Current Account Balance	cf
Currency Market	M2 (China)	M2 in China	cm2
	M2 Growth (China)	M2 Growth in China	dm2
	M2 (U.S.)	M2 in the United States	um2
Exchange Rates	USD/JPY	USD/JPY Exchange Rate	USDJPY
	GBP/USD	GBP/USD Exchange Rate	GBPUSD
	EUR/CHN	EUR/CHN Exchange Rate	EURCHN
	EUR/USD	EUR/USD Exchange Rate	EURUSD
	AUD/USD	AUD/USD Exchange Rate	AUDUSD
	USD/CAD	USD/CAD Exchange Rate	USDCAD
	USD/CHF	USD/CHF Exchange Rate	USDCHF
U.S. Dollar Index	USDX	U.S. Dollar Index Published by ICE	USDX

Table 3. Factors Description Table..

Category	Indicator	Description	Indicator Name
Fundamental Data	Exchange Rate	Offshore Spot RMB Closing Price	rate
	Long-term Bond Yield Differential (U.S.)	10-Year Government Bond Yield - U.S. Federal Funds Rate	udr
	CPI (U.S.)	U.S. Consumer Price Index	cpiu
Stock Index Data	CSI 300 Index	CSI 300 Index	HS300
Current Account	Import and Export Value	Current Month Import and Export Value	trade
	Import Value	Current Month Import Value	inputu
Currency Market	M2 (U.S.)	M2 in the United States	um2
Exchange Rates	EUR/USD	EUR/USD Exchange Rate	EURUSD
	AUD/USD	AUD/USD Exchange Rate	AUDUSD
U.S. Dollar Index	USDX	U.S. Dollar Index Published by ICE	USDX

2.2. Feature Selection

In the data introduction, we identified three sources of data: macroeconomic statistics (e.g., import/export figures, capital flows), financial market data (e.g., stock indices, real-time exchange rates), and macroeconomic variables reflecting economic fundamentals (e.g., M2, spot interest rates, price indices).

However, the selection of these features posed challenges. The data varied in frequency—mostly daily, but some monthly, requiring monthly values to be treated as daily values. This reflects how market participants often rely on past monthly data when real-time information is unavailable. Additionally, there was potential redundancy among features, such as between major currency exchange

rates and the U.S. dollar index or between different stock indices. With 40 features across six categories, we initially conducted a preliminary analysis by predicting exchange rates using individual features to assess their importance. The results were underwhelming, with no single feature proving particularly significant.

To improve model accuracy, we performed feature selection using a ridge regression-based wrapper method, focusing on the subset of features that best enhanced the performance of the final learning model. We selected the following features as the final feature set for deep learning forecasting models: **HS300**, **cpiu**, **audusd**, **eurusd**, **m2u2**, **inputu**, **trade**, **udr**, **usdx**, **date**.

2.3. Models

We leverage the recently developed deep learning models, including TSMixer [86–90], FEDformer [91–95], LSTM [96–98], PatchTST [99,100], TimesNet [101–103], Transformer [104–106], MLP [63,107–109], TCN [110–113], and iTransformer [114–118] for exchange rate prediction.

LSTM: We used three layers of LSTM with hidden dimensions of 32, 64, and 64. **TCN**: Applied in multi-variate time series prediction, using three layers of dilated temporal convolutions with residual connections. Kernel sizes were (3, 3, 5) with a dilation rate of (2, 2, 2). **TSMixer [119,120]**: An MLP-based neural network with 5 layers, each capturing and mixing temporal features. Configured with a maximum feature dimension of 16 for balanced learning and efficiency. **TimesNet**: Combines convolutional layers to model local and global temporal patterns, transforming 1D time series into 2D tensors. The architecture includes 1 FFT block and 4 convolutional blocks. **PatchTST [121]**: Divides input sequences into patches, processed independently using 6 transformer layers with 8 heads per layer, and a patch size of 12. **iTransformer**: Trained with improved positional encoding and attention mechanisms, focusing on feature dimension attention. Consists of 4 transformer blocks, each with 2 layers and 8 attention heads. **Fedformer**: Configured with 3 encoder and 2 decoder layers, each using 8-head self-attention. The feature dimension is 256, with feedforward networks at 512. **Transformer**: Implemented with 6 layers in both encoder and decoder, using 8-head self-attention and a feature dimension of 512. Each layer has a feedforward network with an inner dimension of 2048, with positional encoding to capture temporal dependencies. **MLP**: Configured as a fully connected network with 3 hidden layers, each containing 128 neurons and ReLU activation for non-linearity.

All the models were trained using the MAE loss function. We used the Adam optimizer with the following hyperparameters: learning rate (lr) = 10^{-3} , $\beta_1 = 0.9$, $\beta_2 = 0.999$. The models were trained for 1000 epochs with five-fold cross-validation.

3. Experiments

We tested nine deep-learning models on our data. We tried input and prediction length pairs: (32, 16), (48, 24), (64, 32), (96, 48), and (128, 64).

3.1. Evaluation Metrics

In this study, model performance was evaluated using two key metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics were selected to provide a thorough evaluation of the model's predictive accuracy by quantifying the differences between the predicted and actual values. MSE emphasizes larger errors by squaring them, while MAE offers a straightforward measurement of the average error magnitude [122]. To ensure the reliability of the results, the metrics for each model were averaged across five-fold cross-validation, offering a more robust performance assessment.

3.2. Performance Analysis

The overall performance of the evaluated models, as presented in Table 4, highlights the advantage of modern transformer-based architectures over traditional deep learning models (MLP, TCN, and LSTM) in the task of time series forecasting. Notably, TSMixer, FEDformer, and iTransformer

consistently achieved superior results, demonstrating their ability to accurately capture and predict complex temporal patterns.

Among all models, TSMixer has the lowest MAE and MSE values across multiple prediction lengths. This performance underscores the model's effectiveness in handling diverse temporal dependencies. It is highly effective in both long-term forecasting benchmarks for multivariate time series and real-world forecasting tasks, which in our case, the exchange rate prediction. The model's capacity to generalize well across varying forecasting horizons further solidifies its position as the top performer.

Table 4. Performance Metrics of Different Models with Different Prediction Lengths

Model	MAE					MSE				
Prediction Length	16	24	32	48	64	16	24	32	48	64
TSMixer	0.032	0.039	0.045	0.054	0.063	0.002	0.003	0.004	0.006	0.007
FEDformer	0.033	0.042	0.049	0.063	0.086	0.002	0.004	0.004	0.007	0.012
iTransformer	0.035	0.054	0.072	0.089	0.064	0.002	0.005	0.008	0.011	0.008
PatchTST	0.039	0.060	0.056	0.082	0.125	0.003	0.011	0.006	0.012	0.034
TimesNet	0.038	0.052	0.063	0.084	0.104	0.003	0.005	0.008	0.012	0.020
Transformer	0.042	0.057	0.067	0.088	0.095	0.004	0.007	0.008	0.014	0.015
MLP	0.053	0.074	0.124	0.103	0.233	0.006	0.013	0.043	0.017	0.108
TCN	0.049	0.080	0.149	0.128	0.063	0.004	0.009	0.026	0.020	0.007
LSTM	0.047	0.056	0.064	0.098	0.132	0.005	0.006	0.007	0.015	0.027

FEDformer closely follows TSMixer, performing strongly across both metrics. Its architecture, leveraging Fourier Enhanced Decomposed blocks, excels at separating long-term trends from short-term fluctuations, capturing periodic patterns and transient anomalies for reliable forecasts. iTransformer, with its inverted self-attention mechanism, also outperforms classical models like MLP, TCN, and LSTM by providing a nuanced understanding of temporal and feature relationships, particularly for longer prediction horizons.

In contrast, traditional models like MLP and LSTM struggled to match the accuracy of transformer-based models. MLP, lacking sequential modeling, showed higher error rates, especially in MSE. While LSTM can model long-term dependencies, it was surpassed by transformer models in capturing intricate temporal patterns. TCN performed better than MLP and LSTM but still lagged behind TSMixer and FEDformer, as its convolutional structure, while improving temporal understanding, lacks the dynamic feature selection of attention-based models, resulting in less accurate forecasts.

The findings are further supported by Figure 1, where the predicted outputs are compared with the true exchange rate values. TSMixer and FEDformer closely track the actual values, reflecting their high accuracy. TCN and TimesNet also perform well but show slightly more deviation. In contrast, LSTM and MLP, despite being traditional models, exhibit the most variability.

In conclusion, the analysis confirms that TSMixer, FEDformer, and iTransformer excel in capturing complex temporal dependencies and nonlinear relationships. These models consistently achieve lower MAE and MSE values, making them more suitable for real-world forecasting where precision and robustness are critical. The shift towards transformer architectures in time series forecasting is well-justified, particularly when compared to traditional deep learning methods.

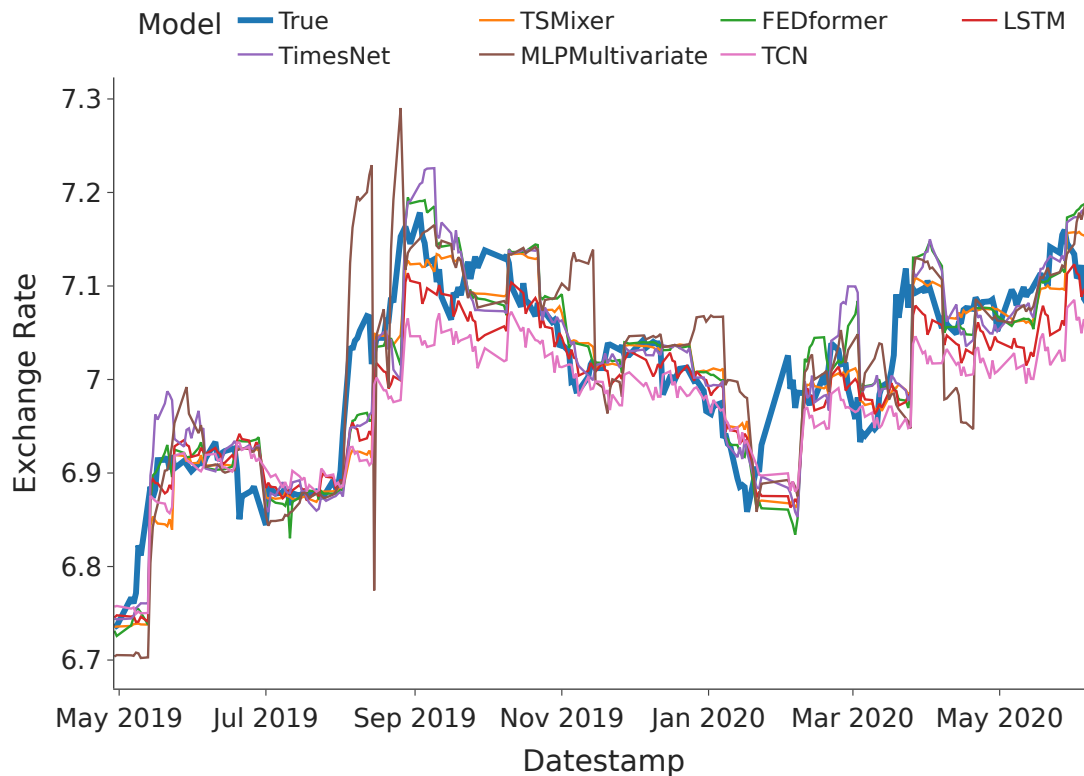


Figure 1. Comparison of True Values vs. Predicted Outputs from six models: TSMixer, FEDformer, LSTM, TimesNet, MLP, TCN

3.3. Gradient Based Feature Importance Analysis

We applied grad-CAM [123,124] to obtain visual explanations for decisions from our deep learning models, making them more transparent and explainable. grad-CAM generates heatmaps that capture the TSMixer's focus during prediction on different features and time points, shown in Figure 2. The higher value (red) indicates the model is focusing more. We can clearly observe that the TSMixer model assigns significant weight to certain low-frequency macroeconomic variables, such as import and export volumes and trade amounts. However, the model appears to place even greater emphasis on trade-related data. It also shows a notable focus on fundamental data, such as the yield spread between U.S. 10-year Treasury bonds and the Federal Reserve's interest rates, as well as the U.S. Consumer Price Index. The model assigns considerable importance to other currency pairs but does not prioritize stock indices like the CSI 300. This is visually evident in the figure, where these features stand out prominently. These findings suggest that variables with different frequencies play varying roles in exchange rate forecasting, but fundamental data remains crucial. This outcome also reinforces the reliability of using machine learning models for exchange rate prediction.

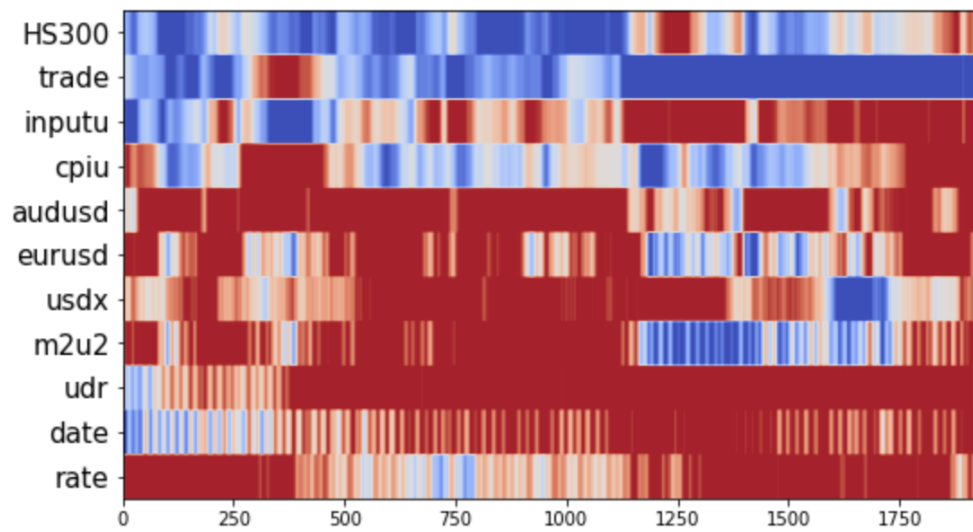


Figure 2. GradCAM visualization indicates feature contribution for forecasting.

4. Discussion

4.1. Modern Deep Learning Models on Exchange Rate Prediction

The application of transformer-based models in exchange rate forecasting has significantly outperformed traditional approaches like MLP, TCN, and LSTM. The superior performance of TSMixer, FEDformer, and iTransformer highlights their advanced capabilities in capturing complex temporal dependencies and nonlinear relationships. TSMixer, in particular, achieves the lowest MAE and MSE across various prediction lengths, effectively prioritizing relevant time steps through self-attention. FEDformer excels in distinguishing long-term trends from short-term fluctuations using Fourier Enhanced Decomposed blocks, while iTransformer enhances temporal and feature relationship understanding, particularly for longer horizons. These models set a new benchmark in predictive accuracy for exchange rate forecasting, with the potential to revolutionize financial analysis and decision-making.

4.2. Theoretical and Real World Implications

The success of transformer-based models in this study suggests a paradigm shift in exchange rate forecasting, where advanced deep learning architectures increasingly outperform traditional models. This shift is crucial for financial markets, where accurate predictions are vital for risk management and strategic decision-making. These models' ability to handle complex, nonlinear data opens new research avenues in financial forecasting, potentially leading to even more sophisticated models. Practically, their enhanced predictive performance benefits traders, analysts, and policymakers by providing more accurate forecasts, mitigating exchange rate volatility risks, and contributing to financial market stability. Moreover, using explainable AI techniques like Grad-CAM improves model interpretability, making them more transparent and trustworthy for decision-makers.

This study successfully demonstrates the superiority of transformer-based models in forecasting the RMB/USD exchange rate, outperforming traditional deep learning models such as MLP, TCN, and LSTM. The advanced architectures of TSMixer, FEDformer, and iTransformer allow these models to capture complex temporal dependencies and nonlinear relationships more effectively, resulting in significantly lower MAE and MSE values. These findings underscore the transformative potential of transformer-based models in financial forecasting, particularly in the highly volatile context of exchange rates.

The visual analysis provided by Grad-CAM further corroborates the effectiveness of these models, highlighting their focus on the most influential features in the data. This interpretability is crucial

for building trust in AI-driven financial models, ensuring that their predictions are not only accurate but also understandable to end-users. While this study focuses on exchange rate forecasting, the implications of these findings extend to other areas of financial analysis, where the ability to predict complex patterns and relationships is essential.

5. Limitations and Future Research

Despite the promising results, this study has several limitations. The models were evaluated solely on the RMB/USD exchange rate, which may limit the generalizability of the findings to other currencies or financial instruments. Additionally, the models were tested using historical data and did not account for unexpected economic shocks or geopolitical events, which could impact forecast accuracy. While Grad-CAM enhances model interpretability, the complex nature of transformer models still presents challenges in fully understanding their decision-making processes. Furthermore, the study primarily addresses short to medium-term forecasting; the effectiveness of these models for long-term predictions and their adaptability to evolving market conditions remain areas for further investigation. Future research should focus on applying these models to a broader range of financial data, incorporating external factors, and evaluating their performance over extended periods.

6. Conclusions

This study demonstrates the superiority of transformer-based models in forecasting the RMB/USD exchange rate, surpassing traditional deep learning models like MLP, TCN, and LSTM. The advanced architectures of TSMixer, FEDformer, and iTransformer effectively capture complex temporal dependencies and nonlinear relationships, resulting in significantly lower MAE and MSE values. These findings highlight the transformative potential of transformer-based models in financial forecasting, particularly in volatile exchange rate contexts.

The use of Grad-CAM for visual analysis further validates these models, showing their focus on the most influential features, which enhances interpretability and builds trust in AI-driven financial models. While this study centers on exchange rate forecasting, the implications extend to other financial analyses requiring complex pattern prediction.

In conclusion, transformer-based models mark a significant advancement in financial forecasting, setting a new standard for accuracy and reliability. Future research should explore their application to other financial time series and assess their long-term performance in real-world market conditions.

References

1. Dang, B.; Ma, D.; Li, S.; Qi, Z.; Zhu, E. Deep learning-based snore sound analysis for the detection of night-time breathing disorders. *Applied and Computational Engineering* **2024**, *76*.
2. Dang, B.; Zhao, W.; Li, Y.; Ma, D.; Yu, Q.; Zhu, E.Y. Real-Time Pill Identification for the Visually Impaired Using Deep Learning. *arXiv preprint arXiv:2405.05983* **2024**.
3. Song, X.; Wu, D.; Zhang, B.; Peng, Z.; Dang, B.; Pan, F.; Wu, Z. ZeroPrompt: Streaming Acoustic Encoders are Zero-Shot Masked LMs. *Proc. INTERSPEECH 2023*, 2023, pp. 1648–1652. doi:10.21437/Interspeech.2023-1497.
4. Ma, D.; Li, S.; Dang, B.; Zang, H.; Dong, X. Fostc3net: A lightweight YOLOv5 based on the network structure optimization. *Journal of Physics: Conference Series* **2024**, *2824*, 012004.
5. Li, S.; others. Utilizing the LightGBM algorithm for operator user credit assessment research. *Applied and Computational Engineering* **2024**, *75*, 36–47.
6. Chen, Y.; Xiao, Y. Recent Advancement of Emotion Cognition in Large Language Models. *arxiv preprint* **2024**.
7. Li, Y.; others. MPGraf: a Modular and Pre-trained Graphformer for Learning to Rank at Web-scale. *ICDM. IEEE*, 2023, pp. 339–348.
8. Zhang, Z.; others. Simultaneously detecting spatiotemporal changes with penalized Poisson regression models. *arXiv:2405.06613* **2024**.

9. Li, X.; Liu, S. Predicting 30-Day Hospital Readmission in Medicare Patients: Insights from an LSTM Deep Learning Model. *medRxiv* **2024**. doi:10.1101/2024.09.08.24313212.
10. Zheng, H.; others. Identification of Prognostic Biomarkers for Stage III Non-Small Cell Lung Carcinoma in Female Nonsmokers Using Machine Learning. *arXiv preprint arXiv:2408.16068* **2024**.
11. Li, Y.; Xiong, H.; Kong, L.; Zhang, R.; Dou, D.; Chen, G. Meta hierarchical reinforced learning to rank for recommendation: a comprehensive study in moocs. *ECML PKDD, 2022*, pp. 302–317.
12. Zhang, Q.; others. CU-Net: a U-Net architecture for efficient brain-tumor segmentation on BraTS 2019 dataset. *arXiv:2406.13113* **2024**.
13. Li, Y.; others. MHRR: MOOCs Recommender Service With Meta Hierarchical Reinforced Ranking. *IEEE Transactions on Services Computing* **2023**.
14. Wang, L.; others. Semi-supervised learning for k-dependence Bayesian classifiers. *Applied Intelligence* **2022**, pp. 1–19.
15. Li, Y.; others. Coltr: Semi-supervised learning to rank with co-training and over-parameterization for web search. *IEEE Transactions on Knowledge and Data Engineering* **2023**, *35*, 12542–12555.
16. Liu, X.; others. Enhancing Skin Lesion Diagnosis with Ensemble Learning. *arXiv preprint arXiv:2409.04381* **2024**.
17. Chen, Y.; others. Do Large Language Models have Problem-Solving Capability under Incomplete Information Scenarios? *Proceedings of the 62nd Annual Meeting of the ACL, 2024*.
18. Xu, H.; others. Can Speculative Sampling Accelerate ReAct Without Compromising Reasoning Quality? *The Second Tiny Papers Track at ICLR 2024, 2024*.
19. Ji, Y.; others. RAG-RLRC-LaySum at BioLaySumm: Integrating Retrieval-Augmented Generation and Readability Control for Layman Summarization of Biomedical Texts. *arXiv preprint arXiv:2405.13179* **2024**.
20. Xu, H.; others. LlamaF: An Efficient Llama2 Architecture Accelerator on Embedded FPGAs. *arXiv preprint arXiv:2409.11424* **2024**.
21. Ji, Y.; Yu, Z.; Wang, Y. Assertion Detection Large Language Model In-context Learning LoRA Fine-tuning. *arXiv:2401.17602* **2024**.
22. Ji, Y.; others. Prediction of COVID-19 Patients' Emergency Room Revisit using Multi-Source Transfer Learning. *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), 2023*, pp. 138–144. doi:10.1109/ICHI57859.2023.00028.
23. Li, Y.; others. Ltrgcn: Large-scale graph convolutional networks-based learning to rank for web search. *ECML PKDD, 2023*, pp. 635–651.
24. Dan, H.C.; others. Multiple distresses detection for Asphalt Pavement using improved you Only Look Once Algorithm based on convolutional neural network. *International Journal of Pavement Engineering* **2024**.
25. Li, Y.; Xiong, H.; Kong, L.; Bian, J.; Wang, S.; others. GS2P: a generative pre-trained learning to rank model with over-parameterization for web-scale search. *Machine Learning* **2024**, pp. 1–19.
26. Dan, H.C.; Lu, B.; Li, M. Evaluation of asphalt pavement texture using multiview stereo reconstruction based on deep learning. *Construction and Building Materials* **2024**, *412*, 134837.
27. Xiong, H.; Bian, J.; Li, Y.; Li, X.; Du, M.; Wang, S.; Yin, D.; Helal, S. When Search Engine Services meet Large Language Models: Visions and Challenges. *IEEE Transactions on Services Computing* **2024**.
28. Fan, X.; Tao, C. Towards Resilient and Efficient LLMs: A Comparative Study of Efficiency, Performance, and Adversarial Robustness. *arXiv preprint arXiv:2408.04585* **2024**.
29. Wang, Z.; others. CDC-YOLOFusion: Leveraging Cross-Scale Dynamic Convolution Fusion for Visible-Infrared Object Detection. *IEEE Transactions on Intelligent Vehicles* **2024**, pp. 1–14.
30. Chen, Y.; others. EmotionQueen: A Benchmark for Evaluating Empathy of Large Language Models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 2024*.
31. Gao, H.; others. A novel texture extraction method for the sedimentary structures' classification of petroleum imaging logging. *CCPR*. Springer, 2016, pp. 161–172.
32. Chen, Y.; others. HOTVCOM: Generating Buzzworthy Comments for Videos. *Proceedings of the 62nd Annual Meeting of the ACL, 2024*.
33. Yang, X.; others. Retargeting destinations of passive props for enhancing haptic feedback in virtual reality. *VRW, 2022*, pp. 618–619.
34. Shen, X.; others. Harnessing XGBoost for robust biomarker selection of obsessive-compulsive disorder (OCD) from adolescent brain cognitive development (ABCD) data. *ICBBE 2024. SPIE, 2024*, Vol. 13252.

35. Li, Y.; others. S2phere: Semi-supervised pre-training for web search over heterogeneous learning to rank data. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 4437–4448.
36. Ouyang, N.; others. Anharmonic lattice dynamics of SnS across phase transition: A study using high-dimensional neural network potential. *Applied Physics Letters* **2021**, 119.
37. Liu, X.; Yu, Z.; Tan, L. Deep Learning for Lung Disease Classification Using Transfer Learning and a Customized CNN Architecture with Attention. *arXiv preprint arXiv:2408.13180* **2024**.
38. Chen, Y.; others. Hallucination detection: Robustly discerning reliable answers in large language models. *CIKM*, 2023, pp. 245–255.
39. Li, Z.; others. Incorporating economic indicators and market sentiment effect into US Treasury bond yield prediction with machine learning. *Journal of Infrastructure, Policy and Development* **2024**, 8, 7671. doi:10.24294/jipd.v8i9.7671.
40. Yu, C.; others. Advanced User Credit Risk Prediction Model using LightGBM, XGBoost and Tabnet with SMOTEENN. *arXiv:2408.03497* **2024**.
41. Chen, Y.; others. TemporalMed: Advancing Medical Dialogues with Time-Aware Responses in Large Language Models. *WSDM*, 2024.
42. Ding, Z.; others. Regional Style and Color Transfer. *CVIDL. IEEE*, 2024, pp. 593–597. doi:10.1109/CVIDL62147.2024.10604182.
43. Yu, H.; others. Enhancing Healthcare through Large Language Models: A Study on Medical Question Answering. *arXiv:2408.04138* **2024**.
44. Ding, Z.; others. Confidence Trigger Detection: Accelerating Real-Time Tracking-by-Detection Systems. *ICECAI*, 2024, pp. 587–592. doi:10.1109/ICECAI62591.2024.10674884.
45. Yang, Q.; others. A Comparative Study on Enhancing Prediction in Social Network Advertisement through Data Augmentation. *MLISE. IEEE*, 2024, pp. 214–218. doi:10.1109/MLISE62164.2024.10674203.
46. Zhou, Y.; others. Evaluating Modern Approaches in 3D Scene Reconstruction: NeRF vs Gaussian-Based Methods. *arXiv preprint arXiv:2408.04268* **2024**.
47. Ni, H.; others. Harnessing Earnings Reports for Stock Predictions: A QLoRA-Enhanced LLM Approach. *arXiv preprint arXiv:2408.06634* **2024**.
48. Chen, Y.; others. Mapo: Boosting large language model performance with model-adaptive prompt optimization. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 3279–3304.
49. Ouyang, N.; Wang, C.; Chen, Y. Temperature-and pressure-dependent phonon transport properties of SnS across phase transition from machine-learning interatomic potential. *International Journal of Heat and Mass Transfer* **2022**, 192, 122859.
50. Ke, Z.; Li, Z.; Cao, Z.; Liu, P. Enhancing transferability of deep reinforcement learning-based variable speed limit control using transfer learning. *IEEE Transactions on Intelligent Transportation Systems* **2020**, 22, 4684–4695.
51. Ke, Z.; Zou, Q.; Liu, J.; Qian, S. Real-time system optimal traffic routing under uncertainties—Can physics models boost reinforcement learning? *arXiv preprint arXiv:2407.07364* **2024**.
52. Li, P.; Lin, Y.; Schultz-Fellenz, E. Contextual Hourglass Network for Semantic Segmentation of High Resolution Aerial Imagery. *2024 5th International Conference on Electronic Communication and Artificial Intelligence (ICECAI). IEEE*, 2024, pp. 15–18. doi:10.1109/ICECAI62591.2024.10674750.
53. Gao, L.; others. Autonomous multi-robot servicing for spacecraft operation extension. *IROS. IEEE*, 2023, pp. 10729–10735.
54. Yu, L.; others. Stochastic analysis of touch-tone frequency recognition in two-way radio systems for dialed telephone number identification. *ICAACE. IEEE*, 2024, pp. 1565–1572.
55. Ni, H.; others. Time Series Modeling for Heart Rate Prediction: From ARIMA to Transformers. *EEI. IEEE*, 2024, pp. 584–589. doi:10.1109/EEI63073.2024.10695966.
56. Gao, L.; others. Decentralized Adaptive Aerospace Transportation of Unknown Loads Using A Team of Robots. *arXiv:2407.08084* **2024**.
57. Zhang, Y.; others. Manipulator control system based on machine vision. *ATCI. Springer*, 2020, pp. 906–916.
58. Li, P.; Yang, Q.; Geng, X.; Zhou, W.; Ding, Z.; Nian, Y. Exploring Diverse Methods in Visual Question Answering. *ICECAI. IEEE*, 2024, pp. 681–685. doi:10.1109/ICECAI62591.2024.10674838.

59. Ding, Z.; Li, P.; Yang, Q.; Li, S. Enhance Image-to-Image Generation with LLaVA-generated Prompts. 2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS). IEEE, 2024, pp. 77–81. doi:10.1109/ISPDS62779.2024.10667513.
60. Li, P.; others. Deception Detection from Linguistic and Physiological Data Streams Using Bimodal Convolutional Neural Networks. ISPDS. IEEE, 2024, pp. 263–267. doi:10.1109/ISPDS62779.2024.10667569.
61. Gao, L.; others. Adaptive Robot Detumbling of a Non-Rigid Satellite. *arXiv preprint arXiv:2407.17617* **2024**.
62. Song, Y.; others. Looking From a Different Angle: Placing Head-Worn Displays Near the Nose. Proceedings of the Augmented Humans International Conference 2024, 2024, pp. 28–45.
63. Tan, L.; others. Enhanced self-checkout system for retail based on improved YOLOv10. *arXiv preprint arXiv:2407.21308* **2024**.
64. Li, Z.; others. A Contrastive Deep Learning Approach to Cryptocurrency Portfolio with US Treasuries. *Journal of Computer Technology and Applied Mathematics* **2024**, *1*, 1–10. doi:10.5281/zenodo.13357988.
65. Xiang, J.; Guo, L. Comfort Improvement for Autonomous Vehicles Using Reinforcement Learning with In-Situ Human Feedback. Technical report, SAE Technical Paper, 2022.
66. Chen, Y.; others. Grow-and-Clip: Informative-yet-Concise Evidence Distillation for Answer Explanation. 2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022, pp. 741–754.
67. Zheng, S.; others. Coordinated variable speed limit control for consecutive bottlenecks on freeways using multiagent reinforcement learning. *Journal of advanced transportation* **2023**, *2023*, 4419907.
68. Liu, W.; others. Enhancing document-level event argument extraction with contextual clues and role relevance. *arXiv:2310.05991* **2023**.
69. Chen, Y.; Xiao, Y.; Li, Z.; Liu, B. XMQAs: Constructing Complex-Modified Question-Answering Dataset for Robust Question Understanding. *IEEE Transactions on Knowledge and Data Engineering* **2023**.
70. Chen, Y.; others. Dr.Academy: A Benchmark for Evaluating Questioning Capability in Education for Large Language Models. ACL, 2024.
71. Xie, T.; others. Darwin series: Domain specific large language models for natural science. *arXiv preprint arXiv:2308.13565* **2023**.
72. Liu, D.; others. GraphSnapShot: Graph Machine Learning Acceleration with Fast Storage and Retrieval. *arXiv preprint arXiv:2406.17918* **2024**.
73. Xie, T.; others. Large language models as master key: unlocking the secrets of materials science with GPT. *arXiv:2304.02213* **2023**.
74. Li, Z.; others. SiaKey: A Method for Improving Few-shot Learning with Clinical Domain Information. BHI. IEEE, 2023, pp. 1–4.
75. Ke, Z.; Qian, S. Leveraging ride-hailing services for social good: Fleet optimal routing and system optimal pricing. *Transportation Research Part C: Emerging Technologies* **2023**, *155*, 104284.
76. Chen, Y.; others. Can Pre-trained Language Models Understand Chinese Humor? Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, 2023, pp. 465–480.
77. Fan, X.; others. Advanced Stock Price Prediction with xLSTM-Based Models: Improving Long-Term Forecasting. *Preprints* **2024**.
78. Ke, Z.; others. Interpretable mixture of experts for time series prediction under recurrent and non-recurrent conditions. *arXiv preprints arXiv:2409.03282* **2024**.
79. Yu, C.; others. Credit card fraud detection using advanced transformer model. *arXiv preprint arXiv:2406.03733* **2024**.
80. Zheng, S.; Li, Z.; Li, M.; Ke, Z. Enhancing reinforcement learning-based ramp metering performance at freeway uncertain bottlenecks using curriculum learning. *IET Intelligent Transport Systems* **2024**.
81. Song, Y.; others. Looking From a Different Angle: Placing Head-Worn Displays Near the Nose. *Augmented Humans*, 2024, pp. 28–45.
82. Ouyang, N.; Wang, C.; Chen, Y. Role of alloying in the phonon and thermal transport of SnS–SnSe across the phase transition. *Materials Today Physics* **2022**, *28*, 100890.
83. Chen, Y.; others. Talk Funny! A Large-Scale Humor Response Dataset with Chain-of-Humor Interpretation. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 17826–17834.
84. Wang, Z.; others. Graph neural network recommendation system for football formation. *Applied Science and Biotechnology Journal for Advanced Research* **2024**, *3*, 33–39.

85. Liu, D.; others. LLMEasyQuant – An Easy to Use Toolkit for LLM Quantization. *arXiv preprint arXiv:2406.19657* **2024**.
86. Ekambaram, V.; others. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. *ACM SIGKDD*, 2023, p. 459–469.
87. Wang, L.; others. Semi-supervised weighting for averaged one-dependence estimators. *Applied Intelligence* **2022**, pp. 1–17.
88. Liu, X.; others. Deep learning in medical image classification from mri-based brain tumor images. *arXiv preprint arXiv:2408.00636* **2024**.
89. Zeng, Z.; others. RSA: Resolving Scale Ambiguities in Monocular Depth Estimators through Language Descriptions. *arXiv preprint arXiv:2410.02924* **2024**.
90. Liu, D.; others. Distance Re-computator and Topology Reconstructor for Graph Neural Networks. *arXiv preprint arXiv:2406.17281* **2024**.
91. Zhou, T.; others. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. *ICML*, 2022.
92. Zeng, Z.; others. Worddepth: Variational language prior for monocular depth estimation. *CVPR*, 2024, pp. 9708–9719.
93. Zhang, T.; others. Improving the efficiency of cmos image sensors through in-sensor selective attention. *ISCAS*, 2023, pp. 1–4.
94. Yang, F.; others. NeuroBind: Towards Unified Multimodal Representations for Neural Signals. *arXiv preprint arXiv:2407.14020* **2024**.
95. Zhang, R.; others. Dspoint: Dual-scale point cloud recognition with high-frequency fusion. *arXiv preprint arXiv:2111.10332* **2021**.
96. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Computation* **1997**, 9, 1735–1780.
97. Kang, Y.; others. Tie Memories to E-souvenirs: Hybrid Tangible AR Souvenirs in the Museum. *UIST*, 2022, pp. 1–3.
98. Zhang, R.; Zeng, Z.; Guo, Z.; Li, Y. Can language understand depth? *ACM Multimedia*, 2022, pp. 6868–6874.
99. Nie, Y.; others. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730* **2022**.
100. Song, Y.; others. Going Blank Comfortably: Positioning Monocular Head-Worn Displays When They are Inactive. *ISWC*, 2023, pp. 114–118.
101. Wu, H.; others. Timesnet: Temporal 2d-variation modeling for general time series analysis. *ICLR*, 2022.
102. Song, Y. Deep Learning Applications in the Medical Image Recognition. *American Journal of Computer Science and Technology* **2019**.
103. Arora, P.; others. Comfortably Going Blank: Optimizing the Position of Optical Combiners for Monocular Head-Worn Displays During Inactivity. *ACM ISWC*, 2024, pp. 148–151.
104. Vaswani, A.; others. Attention Is All You Need. *Advances in neural information processing systems* **2017**, 30.
105. Zhang, T.; others. Transformer-Based Selective Super-resolution for Efficient Image Refinement. *AAAI*, 2024, pp. 7305–7313.
106. Yuan, Y.; Huang, Y.; others. Rhyme-aware Chinese lyric generator based on GPT. *arXiv preprint arXiv:2408.10130* **2024**.
107. LeCun, Y.; others. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural computation* **1989**, 1, 541–551.
108. Zhang, J.; others. Prototypical Reward Network for Data-Efficient RLHF. *arXiv preprint arXiv:2406.06606* **2024**.
109. Zhang, Z.; others. Complex scene image editing by scene graph comprehension. *arXiv preprint arXiv:2203.12849* **2022**.
110. Bai, S.; others. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271* **2018**.
111. Zhang, Z.; Qin, W.; Plummer, B.A. Machine-generated Text Localization. *arXiv preprint arXiv:2402.11744* **2024**.
112. Chen, Y.; others. XMeCap: Meme Caption Generation with Sub-Image Adaptability. *Proceedings of the 32nd ACM Multimedia*, 2024.

113. Zhang, Z.; others. Movie genre classification by language augmentation and shot sampling. WACV, 2024, pp. 7275–7285.
114. Chang, P.; others. A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in Intensive Care Unit. *Computer Methods and Programs in Biomedicine* **2024**, *246*.
115. Kang, Y.; others. 6: Simultaneous Tracking, Tagging and Mapping for Augmented Reality. SID Symposium Digest of Technical Papers. Wiley Online Library, 2021, Vol. 52, pp. 31–33.
116. Mo, K.; others. Fine-Tuning Gemma-7B for Enhanced Sentiment Analysis of Financial News Headlines. ICETCI, 2024, pp. 130–135.
117. Huang, S.; Song, Y.; Kang, Y.; Yu, C.; others. AR Overlay: Training Image Pose Estimation on Curved Surface in a Synthetic Way. CS & IT Conference Proceedings, 2024, Vol. 14.
118. Chen, Y.; others. Hadamard adapter: An extreme parameter-efficient adapter tuning method for pre-trained language models. CIKM, 2023, pp. 276–285.
119. Bo, S.; others. Attention Mechanism and Context Modeling System for Text Mining Machine Translation. *arXiv:2408.04216* **2024**.
120. Liu, W.; others. Beyond Single-Event Extraction: Towards Efficient Document-Level Multi-Event Argument Extraction. *arXiv preprint arXiv:2405.01884* **2024**.
121. Ma, D.; others. Transformer-Based Classification Outcome Prediction for Multimodal Stroke Treatment. *arXiv preprint arXiv:2404.12634* **2024**.
122. Zhang, T.; others. Patch-based Selection and Refinement for Early Object Detection. WACV, 2024, pp. 729–738.
123. Selvaraju, R.R.; others. Grad-CAM: visual explanations from deep networks via gradient-based localization. *International journal of computer vision* **2020**, *128*, 336–359.
124. Wang, Z.; others. Research on Autonomous Driving Decision-making Strategies based Deep Reinforcement Learning. *arXiv:2408.03084* **2024**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.