AI로 읽는 글로벌 이슈: 한국 환율 예측을 위한 이벤트와 감성 융합

Reading Global Issues with AI: Event and Sentiment Fusion for Exchange Rate Prediction in Korea

Jaeheung Parka, Kyungwon Kima,*∗*

*aSchool of International Trade and Business, Incheon National University, Incheon, 22012, Republic of Korea*

# 요약

# *본 연구는 원/달러 환율의 단기 예측 성능을 향상시키기 위해 거시·금융시장 지표와 더불어 뉴스 기반 이벤트(GDELT) 및 감성 요약 지표를 통합한 딥러닝 모델을 제안하였다. LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 아키텍처를 대상으로 다양한 입력 조합과 시차 구간(lookback=5~90일)을 실험한 결과, 이벤트·감성 통합 입력과 CNN-LSTM(lookback=20)의 조합이 최적 성능을 달성하였다(MSE=47.71, MAPE=0.39%). 변수 중요도 분석(Permutation Importance)을 통해 금·알루미늄·구리 등 원자재 가격, 미국 정책금리와 국채금리, KOSPI/KOSDAQ 지수 및 거래량 등이 환율 변동의 핵심 설명력을 가지는 것으로 확인되었으며, 뉴스 요약의 감성 변동성 역시 의미 있는 신호로 작용하였다. 시각화 결과, CNN-LSTM 기반 예측치는 실제 환율과 유사한 흐름을 포착하며 단기 방향성을 효과적으로 설명하였다. 본 연구는 사건·감성 정보의 활용이 환율 예측의 정확도를 높일 수 있음을 실증적으로 보여주며, 향후 금융시장 예측에 있어 텍스트 기반 신호의 중요성을 뒷받침한다.*

# *중심어 : 환율, 시계열 예측, 딥러닝, 글로벌 이벤트, 뉴스 감성 분석*

# Abstract

*This study proposes a deep learning framework for enhancing short-term forecasting of the KRW/USD exchange rate by integrating macro-financial indicators with news-based event (GDELT) and sentiment summary variables. Four architectures—LSTM, GRU, CNN-LSTM, and CNN-GRU—were evaluated across multiple input settings and lookback windows (5–90 days). The combination of event–sentiment integration and CNN-LSTM with a 20-day lookback achieved the best performance (MSE=47.71, MAPE=0.39%). Permutation importance analysis revealed that commodity prices (gold, aluminum, copper), U.S. policy and bond yields, and domestic equity indices (KOSPI/KOSDAQ) exerted the strongest explanatory power on exchange rate fluctuations, while sentiment volatility in news summaries also provided significant predictive signals. Visualization of the best-performing model demonstrated that CNN-LSTM predictions closely tracked actual exchange rate dynamics, effectively capturing short-term market direction. These findings provide empirical evidence that incorporating event and sentiment information can improve forecasting accuracy, underscoring the role of text-based signals in financial market prediction.*

*Keywords:* Exchange Rate, Time Series Forecasting, Deep Learning, Global Events, News Sentiment

*∗*Corresponding author: thekimk.kr@gmail.com

*ORCID(s):* 0000-0001-6530-8426 (K. Kim)

# Ⅰ. 서론

환율(Exchange rate)은 외국 통화 한 단위를 획득하기 위해 자국 통화를 얼마나 지불해야 하는지를 나타내는, 즉 자국 통화와 외국 통화 간의 교환 비율을 의미한다[1]. 이는 두 나라 통화의 상대적 가치를 보여주는 지표로, 오늘날 거의 모든 경제 활동이 국제 거래와 밀접하게 연계되어 있기 때문에 개인의 일상생활 뿐 아니라 기업 경영, 국가 경제 운영에 이르기까지 필수적인 정보로 기능한다[2]. 특히 원/달러 환율은 한국 대외경제에서 핵심적인 역할을 한다. 환율 변동은 수출입 가격, 외국인 투자, 물가 수준, 중앙은행의 통화정책 등 다양한 거시경제 변수와 밀접하게 연관되어 있으며, 환율의 향후 변화를 예측하는 것은 국가 경제의 안정성과 기업의 재무 전략 수립에 있어 중요한 과제로 인식된다. 그러나 환율은 경제 펀더멘털뿐만 아니라 지정학적 리스크, 글로벌 자본 흐름, 시장 심리 등 복합적인 요인의 영향을 받는다. 이로 인해 환율 변동은 종종 ‘랜덤워크(Random Walk)’로 설명되며, 이는 술에 취한 사람이 어디로 걸어갈지 예측하기 어렵듯 환율의 다음 시점을 정확히 전망하기 힘들다는 비유적 표현이다[3].

환율은 외환시장의 수요와 공급에 의해 결정되며, 장기적으로는 물가 수준이나 생산성 변화와 같은 경제 여건이 통화 가치에 영향을 미친다. 중기적으로는 대외거래와 거시경제정책이 주요 요인으로 작용하고, 단기적으로는 시장 참가자들의 기대, 주변국 환율 변동, 각종 뉴스에 의해 크게 좌우된다. 특히 시장 참가자들의 기대가 환율 상승 혹은 하락 쪽으로 쏠리면 자기실현적(Self-fulfilling) 거래를 통해 실제 환율 변동을 초래하기도 한다. 예컨대 다수가 환율 상승을 예상할 경우, 환율이 오르기 전에 외환을 선매입하려는 수요가 증가하여 실제로 환율 상승을 유발하게 된다. 또한 뉴스와 같은 정보 역시 시장 참가자들의 기대를 변화시켜 단기 환율 변동에 중요한 영향을 미친다. 실제로 2010년 5월 천안함 침몰조사 결과 발표로 지정학적 리스크가 부각되자 원/달러 환율이 단기간 급등한 사례가 있다[4].

이와 같은 환경에서 환율의 경로를 정밀하게 예측하는 것은 정책당국의 통화·재정정책 수립, 기업의 수출입 가격전략 조정, 금융기관의 리스크 관리에 필수적이다. 기존 국내 연구들은 이러한 필요성을 강조해왔다.

김우석·한규식(2021)은 팬데믹 이후 환율의 구조적 변화를 분석하며 변동성 확대에 따른 예측 필요성을 제시하였고[14], 오인정·김우주(2022)는 코로나19 구간을 대상으로 SARIMA와 ARDL 모형을 비교하여 국면별로 원/달러 예측모형의 성능이 달라진다는 점을 보여주었다[15].

그러나 이러한 연구들은 대부분 종가 기반의 시계열 모형 비교 또는 구조변화 탐색에 초점을 맞추었으며, 이벤트 데이터와 감성 분석 지표를 결합하여 환율을 예측한 사례는 드물다. 이에 본 연구는 기존 선행연구의 거시경제적·구조적 요인을 포괄하면서, 나아가 뉴스 감성 점수와 GDELT 기반 이벤트 지표를 통합한 융합적 접근을 시도한다. 이를 통해 환율 변동성의 구조적 요인과 심리적 요인을 동시에 반영하는 정교한 예측 모델을 제시하고자 한다.

이러한 맥락에서 시장 참여자들의 심리와 뉴스에 담긴 감정을 정량화해 반영하는 감성 분석(Sentiment Analysis)은 환율 예측의 새로운 단서를 제공할 수 있다. 따라서 본 연구는 감성 지표를 분석하여 이를 딥러닝 기반 시계열 예측 모델에 결합함으로써, 원/달러 환율 예측의 정확도를 향상시킬 수 있는 가능성을 탐색하고자 한다.

# Ⅱ. 문헌연구

환율 예측 연구들은 전통적으로 거시경제 변수와 금융시장 변수를 주요 독립변수로 설정하였다.  
Plakandaras et al.(2015)은 1999년부터 2011년까지의 데이터를 바탕으로 원자재(19개), 금속(10개), 주가지수(7개), 무역지수(3개), 금리(11개), 거시경제 지표(11개), 환율(8개) 등 광범위한 변수를 활용하였다. 다양한 모형을 비교한 결과, EEMD 기반 비선형 모형이 일간 및 월간 환율 예측에서 우수한 성능을 보였다[5].  
국내 연구로는 임현욱 외(2021)가 연합인포맥스와 Bloomberg 데이터를 사용하여 원/달러 환율을 예측하였다. 이들은 1년·10년 만기의 원화 IRS 금리, KTB 수익률, 본드 스왑 스프레드 등 금리 및 채권시장 변수를 중심으로 구성하였으며, ANN 모델이 평균 Hit Ratio 50.96%로 가장 우수한 성과를 보였다[6]. Cao et al.(2020)은 유가와 금값 같은 실물자산 변수 뿐 아니라 통화공급(M1, M2), CPI, PPI, 산업생산, 기준금리, 인플레이션율, 무역수지, 정책 불확실성 지수 등 거시경제 지표를 결합하였다. 이들은 제안한 DC-LSTM 모형이 CNN, LSTM, SVR, ARIMA보다 낮은 예측 오차(MAE 0.0145)를 기록했다고 보고하였다[7].  
또한 Qureshi(2025)는 USD/INR 환율을 대상으로 금리, 인플레이션율, GDP 성장률, 외환보유액 등 거시경제 펀더멘털 변수를 활용하였다. 다양한 머신러닝 모형을 비교한 결과, Random Forest가 RMSE 0.029로 가장 우수한 성과를 나타냈다[8]. 마지막으로 Wang et al.(2021)은 USD/CNY 환율 뿐만 아니라 나스닥, 다우존스, 상하이 종합지수, 항셍지수 등 글로벌 주가지수를 독립변수로 결합하였다. CNN, RNN, LSTM 등 여러 모델을 비교한 결과, CNN-TLSTM 모형이 MAPE 0.18945로 가장 높은 예측력을 보였다[9]. 종합하면, 기존 환율 예측 연구들은 공통적으로 금리, 물가, 무역수지, 주가지수, 원자재 가격 등 거시경제 및 금융 변수를 중심으로 환율 변동을 설명하고자 했음을 확인할 수 있다.

기존 선행연구 대비 본 연구의 가장 큰 차별성은 전통적 변수군에 외생적 사건을 반영하는 이벤트 기반 지표를 통합한 점에 있다. 이를 위해 GDELT(Global Database of Events, Language, and Tone) 데이터베이스를 활용하였으며, 이는 전 세계 언론(인쇄·방송·웹 기반)을 실시간으로 모니터링하여 사건 유형, 위치, 감정 톤, 인물 및 조직, 문서 수 등 다양한 속성으로 사건을 정량화하는 데이터베이스이다[16]. GDELT의 데이터는 전 세계 사회적 사건을 정량적 시계열 데이터로 전환할 수 있다는 점에서 기존 거시·금융 예측 변수의 한계를 보완할 수 있는 강력한 도구로 주목받고 있다.

경제 및 금융 분야에서도 GDELT 기반 이벤트 데이터를 예측 변수로 활용한 사례가 있다. 예컨대 ARIMA 모형에 GDELT 데이터를 추가하여 환율 예측 정확도를 약 1% 향상시키고, Gradient Boosting 기반 분류 모델에서는 정확도를 6% 상승시킨 연구[17]와, 유럽 채권시장 분석에서 GDELT 지표가 nowcasting 모델의 성능을 향상시킨 사례가 있다[18]. 또한 미국 주식시장 예측에서, GDELT 지표는 전통적 거시 경제 모형을 능가하는 설명력 및 예측 성능을 제공했다는 실증 결과도 보고되었다[19].

본 연구는 뉴스 기사와 온라인 댓글이라는 두 가지 비정형 텍스트 데이터를 대상으로 감성 분석을 수행하였다. 감성 분석은 자연어 처리(NLP)의 하위 분야로, 텍스트에 내재된 정서적 톤을 식별하고 분류하는 작업을 의미한다[20].

기존 연구에서는 규칙 기반 접근이나 전통적 머신러닝 기법이 주로 활용되었으나, 도메인 의존성, 언어 확장성, 은유·풍자와 같은 복잡한 뉘앙스 처리의 한계가 지속적으로 지적되었다[21]. 이러한 제약을 극복하기 위해 최근에는 Transformer 기반 딥러닝 모델이 적극 도입되었다. 특히 BERT 계열 모델은 양방향 문맥 정보를 활용하여 단어 의미를 정교하게 파악하고, 적은 데이터로도 파인튜닝이 가능하다는 장점이 있다.

금융 분야에서는 FinBERT와 같이 금융 특화 언어모델이 등장하여 기존 머신러닝 접근 대비 성능이 크게 향상되었으며[22], FiQA·Financial PhraseBank와 같은 벤치마크에서도 Transformer 기반 모델들이 전통적 방법을 압도하는 성능을 기록하였다[23]. 더 나아가 최근에는 GPT-4o와 같은 대규모 언어모델(LLM)이 프롬프트 튜닝이나 few-shot 학습만으로도 FinBERT와 동등하거나 그 이상의 성능을 보이는 것으로 보고되고 있다[24][25]. 이는 금융 뉴스·리포트 분석과 같은 복잡한 텍스트 처리에서 LLM이 높은 적응력과 확장성을 제공함을 보여준다.

이러한 흐름을 반영하여, 본 연구는 전통적인 규칙 기반·머신러닝 접근 대신 GPT-4o-mini 기반 LLM 감성 분석 절차를 설계하였다.

최근 금융시장 예측에서는 전통적인 거시·금융 변수 뿐 아니라 시장 참여자의 심리 변수를 반영하기 위해 감성 분석이 적극적으로 활용되고 있다.  
Mohan et al.(2019)은 Apple, Amazon, Microsoft 등 S&P500 주요 기업의 종가 정보와 함께 국제 뉴스 기사에서 추출한 긍·부정 감정 점수를 독립변수로 사용하였다. 분석 결과, 감정 변수를 포함한 RNN 모델이 MAPE 2.03으로 가장 우수한 성능을 보였다[10]. Jing et al.(2021)은 Shanghai Stock Exchange 데이터와 Eastmoney.com 투자자 게시글의 감정 점수를 CNN으로 분류하고, 이를 주가 기술적 지표와 함께 LSTM 모델에 입력하였다. 그 결과 CNN-LSTM 모델이 평균 MAPE 0.0449로 가장 뛰어난 예측력을 보여, 투자자 심리 반영의 효과성을 입증하였다[11]. 또한 정가연 외(2024)는 네이버 증권 뉴스, 한국은행 API, Naver DATALAB 데이터를 활용하여 코스피, S&P500, 유가, 금값, 원/엔 환율, 금리 등 거시·금융 변수와 함께 KLUE-BERT 기반 뉴스 감정 점수, 코스피 검색 빈도수 등 시장 심리 변수를 결합하였다. 실험 결과, CNN-GRU 모델이 MAPE 1.38%로 가장 낮아 감성 분석의 유용성을 보여주었다[12]. 이러한 연구들은 전통적인 금융 변수에 뉴스 및 게시글 감정 점수를 결합하는 융합적 독립변수 구성이 주가 예측 정확도를 높일 수 있음을 시사한다.

환율 예측에도 최근에는 감성 분석을 접목한 연구가 시도되고 있다. Ding et al.(2024)은 Investing.com과 ForexEmpire의 뉴스 및 댓글 데이터를 기반으로 감정 점수를 산출하였다. 이들은 해당 감성 지표를 USD/CAD, USD/MXN, USD/CNY, USD/JPY, USD/KRW 등 주요 교차환율, 원자재 가격(유가, 금, 구리, 옥수수 등), 글로벌 지수(S&P500, EuroStoxx50, VIX), 채권수익률(미국·유로존 10년물)과 같은 거시·금융 변수와 결합하였다. 다양한 모형을 비교한 결과, PSO-LSTM 모델이 RMSE 0.0958로 가장 높은 성능을 기록하였다[13]. 이러한 결과는 환율 예측에서도 단순한 거시경제 지표만으로 설명하기 어려운 단기 변동성을 뉴스 및 댓글에 담긴 시장 심리 변수가 효과적으로 보완할 수 있음을 보여준다.

<표 1>에 정리된 선행연구를 종합하면, 다음과 같은 시사점을 도출할 수 있다. 첫째, 환율 예측에서는 여전히 금리, 물가, 무역수지, 원자재 가격, 주가지수 등 거시경제 및 금융 변수가 핵심적 설명 변수로 기능하고 있다. 둘째, 최근 연구들은 뉴스, 투자자 게시글, 댓글 등 비정형 텍스트 데이터에서 추출한 감정 점수를 독립변수로 추가하여 예측 성능을 향상시키고 있으며, 이는 금융시장 예측에서 시장 심리 변수의 중요성을 뒷받침한다. 셋째, 감성 분석을 결합한 환율 예측 연구는 아직 초기 단계에 있으나, 뉴스·댓글 감성과 같은 심리 변수의 확장적 활용 가능성이 크다는 점에서 향후 연구 방향성을 제시한다. 따라서 본 연구는 이러한 시사점을 바탕으로, 거시경제 변수와 함께 GDELT 이벤트 톤, 뉴스 감성 분석을 통합한 딥러닝 기반 원/달러 환율 예측 모델을 제안한다. 이를 통해 환율 변동성을 설명할 수 있는 새로운 변수를 발굴하고, 기존 연구 대비 예측 정확도를 향상시키고자 한다.

**<표 1> 환율 예측 및 주가 예측과 감성 분석을 활용한 선행연구 분석 방향과 성능 요약**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 연구 목적 | 연구자(연도) | 데이터 소스 | 데이터 기간 | 주요 독립변수 | 방법론 | 성능 |
| **환율 예측** | Plakandaras et al.(2015)  [5] | 미기재 | 1999~2011 | 원자재(19), 금속(10), 주가지수(7), 금리(11), 무역·거시지표 등 | ARIMA, GARCH, AR-NN, MARS-NN, SVR, EEMD 결합 | 일간: EEMD-MARS-SVR, MARS-NN 우수 / 월간: EEMD-AR-SVR 우수 |
| 임현욱 외(2021)  [6] | 연합인포맥스, Bloomberg | 2008~2020 | IRS 금리, KTB 수익률, 본드 스왑 스프레드 | ANN, LR, DT | ANN 평균 Hit Ratio 50.96% |
| Cao et al.(2020)  [7] | 미기재 | 2016.6~2019.4 | 유가, 금값, CPI, PPI, 산업생산, 기준금리, 무역수지, 정책 불확실성 지수 등 | ARIMA, SVR, CNN, LSTM, DC-LSTM | DC-LSTM MAE 0.0145 |
| Qureshi(2025)  [8] | FRED, IMF, World Bank 등 | 2015~2023 | 금리, 인플레이션율, GDP 성장률, 외환보유액 | ARIMA, RW, RF, SVM | RF RMSE 0.029 |
| Wang et al.(2021)  [9] | Wind DB | 2006~2020 | USD/CNY, 나스닥, 다우, 상하이, 항셍 등 지수 | MLP, CNN, RNN, LSTM, CNN-LSTM, CNN-TLSTM | CNN-TLSTM MAPE 0.18945 |
| **주가 예측 + 감성분석** | Mohan et al.(2019)  [10] | 해외 뉴스 웹사이트 | 2013~2017 | S&P500 기업 종가, 뉴스 감정 | ARIMA, Prophet, RNN | RNN-pp MAPE 2.03 |
| Jing et al.(2021)  [11] | SSE, Eastmoney.com | 2017~2019 | 기술적 지표 , 게시글 감정 | CNN+LSTM | MAPE 0.0449 |
| 정가연 외(2024)  [12] | 네이버뉴스, KOSPI, DATALAB, 한은 API | 2021~2023 | 코스피·S&P500, 유가, 금값, 원/엔 환율, 금리, 검색 빈도, 감정 점수 | LSTM, GRU, CNN-LSTM, CNN-GRU | CNN-GRU + KLUE-BERT 감정 MAPE 1.38% |
| **환율 예측 + 감성분석** | Ding et al.(2025)  [13] | Investing.com, ForexEmpire | 2016~2024 | 뉴스·댓글 감정, 교차환율, 원자재, 글로벌 지수, 채권수익률 | PSO-LSTM, SVM, GRU, VAR, ARIMA/GARCH | PSO-LSTM RMSE 0.0958 |

# Ⅱ. 데이터 및 방법론

1. 종속변수 데이터

본 연구는 2020년 1월부터 2024년 12월까지의 일별 원/달러 환율 종가를 예측 대상으로 설정하였다. 선택된 시기는 코로나19 팬데믹의 발생과 그로 인한 경제적 충격, 글로벌 통화정책의 급격한 전환, 그리고 여러 지정학적 리스크가 동시다발적으로 나타난 시기이기 때문에, 환율 변동성이 경제에 미친 영향을 분석하기에 적합하다. 2020년 이후 원/달러 환율은 팬데믹 초기에 급락한 후, 2021년에는 저점 국면을 경험했고, 2022년에는 미국 연방준비제도의 초고속 금리 인상과 한·미 금리 차 확대, 국내 정치·정책 변화, 러시아·우크라이나 전쟁 등의 외적 요인들이 복합적으로 작용하며 큰 폭으로 상승하였다.

특히 2020년과 2021년, 2022년 사이의 환율 추이는 다양한 경제적 요인들이 결합된 결과로, 매우 높은 변동성을 보였다. <그림 1>에서 확인할 수 있듯이, 2021년 초 약 1,070원 수준까지 환율이 하락한 뒤, 2022년 5월에는 1,300원 이상으로 급등하였다. 이는 단순히 환율 수준만이 아니라, 변동성 또한 급격히 확대된 시점을 의미한다. 예를 들어, 2021년 상반기 일평균 환율 등락폭은 ±5원에 불과했으나, 2022년 상반기에는 ±15원으로 크게 확대되었으며, 일간 표준편차 기준 환율 변동성 지표도 같은 기간 동안 0.4%에서 1.2%로 세 배 이상 증가하였다. 환율 변동성의 급증은 한국 경제에 중대한 영향을 미쳤다. 특히 2022년에는 사상 최대의 무역수지 적자가 발생하고, 원화 약세가 지속되며 수입물가 상승, 내수 위축, 소비심리 악화 등의 부정적인 경제적 효과를 불러왔다. 반면, 수출 기업에는 단기적으로 경쟁력이 제공되었으나, 환 헤지 비용이 전년 대비 35% 증가하며 중소기업을 중심으로 금융비용 부담이 커졌다.

라인, 그래프, 스크린샷, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 1> 원/달러 환율 추이**

2. 독립변수 데이터

원/달러 환율 예측을 위한 독립변수는 기존 환율 예측 연구에서 사용된 주요 경제적 요인을 바탕으로 설정하되, 환율 변동성의 구조적 요인과 시장 심리적 요인도 동시에 반영하도록 설계되었다.

기존 연구에서는 거시경제 및 금융시장 지표가 주로 독립변수로 활용되었다. Plakandaras et al. (2015)은 원자재 가격, 금속, 주가지수, 금리, 거시경제 변수 등 60여 개의 지표를 활용하여 환율 예측을 시도했으며 [5], 임현욱 외(2021)는 금리 및 채권시장 변수를 중심으로 한 연구를 진행하였다 [6]. 또한, Cao et al. (2020)은 유가, 금 가격, M1, M2, CPI, PPI 등의 거시지표를 활용했으며 [7], Ali(2025)는 금리, 인플레이션율, GDP 성장률, 외환보유액 등을 핵심 변수로 설정하였다 [8]. Wang et al. (2021)은 USD/CNY 환율과 함께 주요 글로벌 주가지수를 포함하여 국제적 요인을 반영하였다 [9].

최근에는 감성 분석을 통한 시장 심리 변수화가 활발히 이루어지고 있다. Mohan et al. (2019)은 국제 뉴스 기사에서 추출한 감정 점수를 S&P500 주가 예측에 반영하였고 [10], Jing et al. (2021)은 투자자 게시글의 감정을 CNN 기반 분류기로 분석하여 주가 예측에 활용하였다 [11]. 정가연 외(2024)는 KLUE-BERT 기반 뉴스 감정 점수를 거시경제 변수와 결합하여 예측 정확도를 높였고 [12], Ding et al. (2025)은 환율 관련 뉴스와 댓글 감성을 결합하여 환율 예측 성능을 향상시켰다 [13].

본 연구는 기존 연구의 방법론을 바탕으로, 주요 독립변수들을 확장하여 설정하였다. 종속변수로 원/달러 환율을 설정하고, 경쟁적인 환율 변동성을 반영하기 위해 USD/JPY와 USD/CNY를 교차환율 변수로 추가하였다. 또한, 국내외 주요 주식시장 지표인 KOSPI, KOSDAQ, 다우존스, S&P500을 자본 흐름과 위험 선호도에 대한 영향을 고려하여 포함하였다. 원자재 변수로는 WTI 유가, 금, 구리, 니켈, 알루미늄을 채택하여 글로벌 경제 사이클과 인플레이션 기대를 반영하였으며, 거시경제 변수로는 한국과 미국의 정책금리, 소비자물가지수(CPI), 생산자물가지수(PPI), 경상수지, 통화량(M1, M2)을 포함하여 환율 결정에 중요한 금리 차, 물가 수준, 대외 거래, 유동성 등을 반영하였다. 또한 리스크 지표로는 VIX, OVX, KSVKOSPI를 포함하여 금융시장의 불확실성과 투자자의 위험회피 성향을 측정하였다.

3. 파생변수 데이터

(1) 이벤트 데이터

본 연구에서는 GDELT(Global Database of Events, Language, and Tone)에서 추출한 다섯 가지 핵심 이벤트 기반 변수를 활용하였다. GDELT는 전 세계적으로 발생하는 다양한 사건 및 뉴스의 텍스트 데이터를 기반으로 사건들의 속성(예: 장소, 시간, 관련된 국가 등)을 기록하는 데이터베이스로, 이를 통해 다음과 같은 변수들을 정의하였다: (1) 전체 이벤트 수, (2) 평균 감성 톤, (3) 한국 관련 이벤트 톤, (4) 전체 문서 수, (5) 한국 관련 문서 수이다. 이러한 이벤트 지표는 기존의 거시적 구조적 요인(예: 금리, 주가, 원자재 등) 외에 “사건 기반 리스크 및 심리적 충격”을 반영할 수 있는 변수로서, 본 연구의 독창성을 강조하는 중요한 요소이다.

이벤트 수와 관련된 변수들은 주로 뉴스와 같은 외부 사건이 경제 및 금융시장에 미치는 영향을 반영하는 데 사용된다. 예를 들어, 경제 위기나 정치적 불안정성 등 사건 기반 리스크가 환율 변동성에 미치는 영향을 포착할 수 있다. 평균 감성 톤과 한국 관련 이벤트 톤은 각 사건이나 뉴스의 감성을 수치화하여, 특정 사건이나 뉴스가 환율에 미치는 영향을 심리적 측면에서 분석할 수 있는 기반을 제공한다. 또한, 문서 수 변수는 뉴스의 빈도나 그 영향을 반영하여 환율 예측에 중요한 역할을 할 수 있다.

이러한 이벤트 기반 변수는 본 연구가 다루고 있는 환율 변동성 예측에 있어 “사건 발생에 따른 경제적 반응”을 통합적으로 반영할 수 있는 기회를 제공하며, 기존 연구에서 주로 다루어진 금리나 원자재 가격 같은 전통적인 지표와는 다른 시각에서 변동성의 원인을 분석하는 데 기여한다.

(2) 감성분석 데이터

본 연구의 뉴스 감성 분석 데이터는 네이버 금융 뉴스의 “환율” 섹션에서 2020년 1월 1일부터 2024년 12월 31일까지 총 35,633건의 기사를 수집하여 활용하였다. 수집된 원문은 정규표현식을 사용하여 대괄호, 소괄호 내 불필요한 문구, 이메일 주소, 광고성 표현, 그리고 “=” 기호 이전의 문장 등을 제거하여 정제하였다. 또한, 제목과 본문이 동일한 중복 기사나 내용이 비어 있거나 공백만 포함된 기사는 분석 목적에 맞지 않기 때문에 제외하였다. 이를 통해 데이터의 품질을 높이고, 불필요한 정보 노이즈를 최소화하였다.

정제된 뉴스는 GPT-4o-mini를 활용하여 원/달러 환율과의 관련성을 기준으로 세부 분류를 진행하였다. 그 결과, 뉴스는 크게 세 가지 범주로 구분되었다: 원/달러 환율과 직접적으로 관련된 뉴스 22,756건, 간접적으로 관련된 뉴스 7,320건, 그리고 관련이 없는 뉴스 5,557건으로 나누어졌다. 이 과정에서 “직접 및 간접 관련 뉴스”에 대해 추가적인 전처리를 수행하여 감성 분석의 정확성을 더욱 높였다.

뉴스 본문은 길이가 다양하고, 종종 불필요한 수식어, 문맥적 중복, 장황한 서술을 포함하고 있기 때문에 이를 그대로 감성 분석에 활용할 경우 모델의 입력 효율성과 분석 정확도가 떨어질 수 있다. 따라서 본 연구에서는 감성 분석 전처리 과정의 일환으로 뉴스 본문 요약을 수행하였다. 요약 모델은 HuggingFace Hub에서 제공되는 여러 한국어 요약 모델 중에서 mT5-multilingual-XLSum, t5-base-korean-summarization, KoBART-summary-v3 세 가지 모델을 비교하였다. 성능 평가는 AI Hub 한국어 문서 요약(신문기사) 데이터셋을 사용하였으며, ROUGE-1, ROUGE-2, ROUGE-L 지표를 통해 평가하였다. ROUGE는 "Recall-Oriented Understudy of Gisting Evaluation"의 약자로, 텍스트 요약의 품질을 평가하는 데 널리 사용되는 지표 집합이다[22].

본 연구에서는 KoBART-summary-v3 모델이 모든 ROUGE 지표에서 가장 높은 성능을 기록하여 최종적으로 요약 모델로 채택하였다. 구체적으로 ROUGE-1, ROUGE-2, ROUGE-L F1 점수는 각각 0.4128, 0.1632, 0.3835로 나타났으며, 이는 다른 모델에 비해 상대적으로 높은 수준이었다. ROUGE-1은 단어 수준의 중복 정도를 나타내며 원문이 핵심 어휘를 얼마나 잘 포함했는지 평가한다. ROUGE-2는 연속된 2-gram의 중복을 기준으로 문맥적 연결성과 유창성을 간접적으로 반영한다. ROUGE-L은 가장 긴 공통 부분수열을 기준으로 요약이 원문의 전체 문장 구조와 맥락을 얼마나 잘 보존하는지에 대한 품질을 평가한다 [26]. Dalal 외(2024)에 따르면 일반적으로 ROUGE-1 F1 점수가 0.5 이상이면 높은 성능으로 판단할 수 있고, ROUGE-2 F1 는 0.4 이상인 경우, 그리고 ROUGE-L F1은 0.5 이상이 일반적으로 높게 판단될 수 있다 (<https://www.nature.com/articles/s41598-024-70618-w>). 다만 이런 기준은 언어, 도메인, 요약 길이, 참조 요약의 수 등에 따라 매우 달라지므로 절대적 기준이라기보다는 비교 맥락에서 참고하는 지침으로 활용된다. 이 지표들이 높은 점수를 기록한 모델은 문맥을 잘 보존하면서도 원문 내용을 간결하게 요약할 수 있음을 시사한다.

감성 분석은 GPT-4o-mini 기반의 대형 언어 모델(LLM)을 활용한 프롬프트 분류 방식으로 수행되었으며, 감성 라벨 체계는 환율 변동의 금융적 맥락을 반영하여 정의되었다. 긍정적인 감성은 원/달러 환율 하락(원화 강세), 위험 선호 확대, 완화적 환경을 의미하며, 부정적인 감성은 환율 상승(원화 약세), 위험 회피 확대, 긴축적 환경을 의미한다. 중립 감성은 환율 방향에 대한 명확한 판단이 어려운 경우를 나타낸다. GPT-4o-mini에는 few-shot 프롬프트 튜닝을 적용하여, 별도의 파인튜닝 없이 금융 문맥에 특화된 감성 분류가 가능하도록 하였다. 분류 결과는 긍정=+1, 중립=0, 부정=-1로 수치화하였으며, 이를 기반으로 날짜별 평균 감성 점수를 산출하여 일별 감성 지표(daily sentiment index)를 구축하였다.

분석의 세분화를 위해 뉴스의 환율 관련성을 기준으로 세 가지 일별 변수 세트를 생성하였다: (1) 직접적인 환율 뉴스의 일별 평균 감성 점수 및 기사 수, (2) 간접적인 환율 뉴스의 일별 평균 감성 점수 및 기사 수, (3) 두 범주를 모두 포함한 전체 뉴스의 일별 감성 점수 및 기사 수. 이러한 세분화된 감성 지표는 후속 환율 예측 모델의 중요한 입력 변수로 활용되었다.

4. 데이터 전처리 및 통계량

결과적으로 본 연구는 거시경제 지표, 금융시장 지표, 원자재 가격, 리스크 변수뿐만 아니라 이벤트 기반 변수와 뉴스 및 댓글 감성 지표까지 포괄하여, 환율 변동의 구조적 요인과 단기 시장 심리 요인을 동시에 고려하는 통합적 예측 프레임워크를 구축하고자 하였다. 이 연구는 기존의 전통적인 예측 변수들과 함께 사건 기반의 리스크 및 심리적 충격을 반영하는 새로운 변수를 도입함으로써 예측 성능을 향상시키는 것을 목표로 한다.

<표 3>은 본 연구에서 사용된 주요 데이터를 요약한 것이다. 목표변수인 원/달러 환율(KRW/USD)을 중심으로, 교차환율, 국내외 주식지수, 원자재, 거시지표, 리스크지표, 그리고 사건과 심리를 측정하는 이벤트와 감정(네이버 뉴스) 정보까지 총 8개 블록으로 구성하였다. 월 단위로 제공되는 거시지표는 발표 기준을 반영한 대표값으로 일 단위 프레임에 확장하여, 모든 변수가 일 단위 시점에서 일관되게 모델에 투입되도록 동기화하였다. 이와 같은 구성은 환율을 움직이는 구조적 요인과 단기 심리와 사건 요인을 동시에 포착하는 통합적 예측 프레임워크라는 점에서 의의가 있다.

한편, 표의 구성은 몇 가지 실무적 유의사항을 시사한다. 첫째, 교차환율·글로벌 지수·변동성 지표와 KRW/USD를 동일 거래일 종가로 결합할 경우 동행성에 따른 정보 누설 가능성이 존재하므로, 보수적으로는 설명변수를 t−1 등 시차형으로 투입하는 것이 바람직하다. 둘째, 월간 지표를 일 단위로 확장할 때는 forward-fill과 공시일(발표일) 정렬 중 어떤 기준을 채택하는지에 따라 추정치가 민감하게 달라질 수 있다. 셋째, 동일 계열 내 변수(예: KOSPI·KOSDAQ·S&P500·다우, 혹은 VIX·KSVKOSPI, 유가·OVX) 사이에는 높은 상관이 예상되므로, 로그차분·수익률 변환, 가변 선택(라쏘) 또는 SHAP·VIF 점검 등으로 다중공선성을 완화하는 절차가 필요하다. 마지막으로, 이벤트·감정 데이터는 수집 범위와 편향이 상이하므로 동일 스케일로의 일괄 표준화보다는 데이터군별 정규화 전략을 분리 적용하는 편이 안정적이다.

**<표 3> 환율 예측에 사용한 데이터**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 대분류 | 출처 | 데이터명 | 기간 | 단위 | 추출 변수 | 활용 변수 |
| **목표 변수** | Yahoo Finance | KRW/USD | 2020.1–2024.12 | 일 | 종가 | Target (원/달러 환율) |
| **교차 환율** | Yahoo Finance | USD/JPY, USD/CNY | 2020.1–2024.12 | 일 | 종가 | USD/JPY 종가, USD/CNY 종가 |
| **국내외 주식지수** | Investing.com | KOSPI, KOSDAQ, 다우존스, S&P500 | 2020.1–2024.12 | 일 | 종가, 거래량 | KOSPI 종가·거래량, KOSDAQ 종가·거래량, 다우존스 종가, S&P500 종가 |
| **원자재** | Yahoo Finance, Investing.com | WTI, 금(XAU), 구리, 니켈, 알루미늄 | 2020.1–2024.12 | 일 | 종가 | WTI 종가, 금, 구리, 알루미늄, 니켈 종가 |
| **거시 지표** | ECOS, FRED, 통계청, 한국은행 | 정책금리, CPI, PPI, 경상수지, 통화량(M1, M2) | 2020.1–2024.12 | 월 | 한국/미국 정책금리, 소비자물가지수, 생산자물가지수, 경상수지, M1, M2 | 동일 |
| **리스크 지표** | Yahoo Finance | VIX, MOVE, OVX, KSVKOSPI | 2020.1–2024.12 | 일 | 주식·채권·유가 변동성, 코스피 불안 지수 | 동일 |
| **이벤트 지표** | GDELT | events\_cnt, tone\_mean, tone\_kor\_mean, gkg\_doc\_cnt, gkg\_kr\_loc\_cnt | 2020.1–2024.12 | 일 | 이벤트 수, 평균 톤, 한국 관련 톤, 전체 GKG 문서 수, 한국 관련 문서 수 | 동일 |
| **감정 지표** | Naver News | 환율 관련 뉴스 | 2020.1–2024.12 | 일 | 뉴스 감정 점수 | 일일 감정 점수 평균, 뉴스 개수 |

<표 4>는 본 연구에서 사용된 데이터의 일 단위로 정렬된 최종 통합 패널의 기술통계량을 제시한다(N=1,186). 환율 종가의 평균과 중앙값은 각각 약 1,256원과 1,182원으로, 상단 꼬리가 두터운 분포를 보인다(최대 1,472원). 시장과 리스크 지표는 코로나19와 에너지 쇼크 국면을 반영해 분산이 크게 확대되었는데, VIX와 OVX의 최대치는 각각 82.69와 325.15로 스트레스 피크를 확인할 수 있다. 이벤트와 문서 수는 평균적으로 큰 규모를 보이지만 톤 지표는 음(-)의 편향을 나타내며, 뉴스 감정지표는 일평균이 0 부근의 약한 음수로 나타난다. 직접 관련 뉴스의 일평균 기사 수가 간접 관련보다 높아, 환율 관련 보도의 집중적 생산이 관측된다.

기술통계는 동시에 데이터 품질 점검 포인트를 제공한다. 일부 항목은 단위 또는 스케일의 불일치를 시사한다. 예컨대 KRW/USD의 최솟값 93.74원은 현실과 괴리가 크므로 소스·환산 과정(역변환, 통화 스케일)의 재검증이 요구된다. 미국 10년물 금리의 수치 역시 퍼센트(%) 대비 소수점 표기 여부가 불명확하여, 모든 금리·물가지표의 단위를 통일하는 것이 바람직하다. 또한 GDELT 관련 지표에는 0값이 포함되어 수집 공백(크롤링 실패·API 지연 등) 가능성을 시사하므로, 해당 일자를 보간하기보다 “데이터 공백일” 플래그를 도입해 모델이 구조적 결측을 인지하도록 하는 편이 안전하다. 컬럼명 오탈자(예: gkg\_kr\_loc\_춧)는 파이프라인 단계에서 스키마 검증과 명명 규칙 통일을 통해 사전에 제거해야 한다. 끝으로, 변동성·거래량·이벤트 계열의 우측 장꼬리와 감정 점수의 음(-)쪽 치우침은 모델 학습에서 외란에 민감할 수 있으므로, RobustScaler·Huber 손실 등 이상치 강건화 설정을 병행하는 것이 유의미한 성능 개선으로 이어질 수 있다.

**<표 4>** 변수들의 통계량 (N, Mean, Median, Std. Q1, Q3)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **변수명** | **Count** | **Mean** | **Min** | **25%** | **50%** | **75%** | **Max** | **Std** |
| **날짜** | 1186 | 2022-06-25 | 2020-01-02 | 2021-03-29 | 2022-06-25 | 2023-09-18 | 2024-12-30 | Nan |
| **원/달러 환율 종가** | 1186 | 1256.283 | 93.74322 | 1083.53 | 1181.822 | 1265.61 | 1333.02 | 1472.5 |
| **원/달러 환율 시가** | 1186 | 1256.633 | 93.7873 | 1083.87 | 1182.46 | 1266.325 | 1333.395 | 1471.1 |
| **코스피 지수** | 1186 | 2596.073 | 339.2823 | 1457.64 | 2399.31 | 2568.48 | 2758.637 | 3305.21 |
| **코스피 거래량** | 1186 | 7.13E+08 | 3.55E+08 | 824510 | 4.63E+08 | 6.14E+08 | 8.66E+08 | 3.46E+09 |
| **코스닥 지수** | 1186 | 841.1601 | 114.4289 | 428.35 | 762.185 | 847.505 | 919.4125 | 1060 |
| **코스닥 거래량** | 1186 | 1.31E+09 | 5.21E+08 | 6.2E+08 | 9.52E+08 | 1.17E+09 | 1.5E+09 | 5.07E+09 |
| **WTI 유가** | 1186 | 69.53344 | 23.29195 | 7.79 | 62.09 | 73.735 | 82.03 | 126.47 |
| **금(XAU) 가격** | 1186 | 1937.514 | 263.5906 | 1471 | 1778.307 | 1870.39 | 1985.455 | 2786.19 |
| **다우존스 지수** | 1186 | 33603.14 | 4657.789 | 18591.93 | 31022.09 | 33913.62 | 35627.82 | 45014.04 |
| **다우존스 거래량** | 1186 | 3.69E+08 | 1.13E+08 | 1.17E+08 | 3.02E+08 | 3.43E+08 | 4.02E+08 | 1.22E+09 |
| **S&P500 지수** | 1186 | 4250.18 | 764.0219 | 2237.4 | 3810.15 | 4200.01 | 4591.015 | 6090.27 |
| **통화량(M1)** | 1186 | 1219.395 | 103.6807 | 944.3064 | 1184.088 | 1221.11 | 1287.187 | 1371.921 |
| **통화량(M2)** | 1186 | 3616.388 | 347.3367 | 2935.64 | 3309.121 | 3718.333 | 3843.787 | 4183.535 |
| **한국 정책금리** | 1186 | 2.037789 | 1.303587 | 0.5 | 0.548387 | 1.75 | 3.5 | 3.5 |
| **미국 정책금리** | 1186 | 2.60763 | 2.31595 | 0.25 | 0.25 | 1.75 | 5.25 | 5.5 |
| **소비자물가지수(CPI)** | 1186 | 2.811298 | 1.658261 | -0.2 | 1.425 | 2.7 | 3.8 | 6.3 |
| **생산자물가지수(PPI)** | 1186 | 111.5014 | 7.401748 | 98.93 | 103.89 | 116.27 | 117.5 | 119.56 |
| **경상수지** | 1186 | 5333.791 | 4255.906 | -4205 | 1862 | 6365 | 7438 | 13096 |
| **코스피 변동성(KSVKOSPI)** | 1186 | 20.30508 | 6.604767 | 11.73 | 16.5625 | 18.38 | 22.415 | 69.24 |
| **달러/엔 환율** | 1186 | 127.8556 | 18.42469 | 102.678 | 109.3763 | 130.8025 | 145.341 | 161.607 |
| **달러/위안 환율** | 1186 | 6.867263 | 0.318528 | 6.3115 | 6.522125 | 6.94415 | 7.1455 | 7.3278 |
| **미국 10년물 국채금리** | 1186 | 0.267769 | 0.139671 | 0.0499 | 0.1415 | 0.2915 | 0.401 | 0.4988 |
| **구리 가격** | 1186 | 3.820936 | 0.631159 | 2.1195 | 3.553 | 3.8995 | 4.28875 | 5.119 |
| **알루미늄 가격** | 1186 | 2313.547 | 420.7198 | 1452 | 2126 | 2302.5 | 2534.125 | 3873 |
| **VIX(주식 변동성)** | 1186 | 21.43804 | 8.368224 | 11.86 | 15.855 | 19.63 | 25.0225 | 82.69 |
| **OVX(유가 변동성)** | 1186 | 44.72318 | 24.29387 | 23.73 | 34.28 | 39.005 | 47.485 | 325.15 |
| **이벤트 수(events\_cnt)** | 1186 | 133581.2 | 24788.61 | 0 | 117551 | 133423.5 | 150155 | 216375 |
| **이벤트 평균 톤(events\_tone\_mean)** | 1186 | -1.979816 | 0.258274 | -3.333169 | -2.133663 | -1.963814 | -1.832024 | 0 |
| **한국 관련 톤(events\_tone\_kor\_mean)** | 1186 | -0.832832 | 0.832322 | -3.924516 | -1.336452 | -0.710838 | -0.234973 | 1.06649 |
| **GKG 문서 수(gkg\_doc\_cnt)** | 1186 | 88501.87 | 24450.6 | 0 | 68474.75 | 85789 | 107725 | 157565 |
| **한국 관련 문서 수(gkg\_kr\_loc\_춧)** | 1186 | 617.3356 | 276.5282 | 0 | 435 | 547 | 750 | 3016 |
| **직접 뉴스 평균 감정 점수** | 1186 | -0.070743 | 0.429306 | -1 | -0.3636 | -0.07415 | 0.207875 | 1 |
| **직접 뉴스 기사 수** | 1186 | 12.69646 | 4.440581 | 0 | 10 | 13 | 16 | 20 |
| **간접 뉴스 평균 감정 점수** | 1186 | -0.040528 | 0.36739 | -1 | -0.1667 | 0 | 0 | 1 |
| **간접 뉴스 기사 수** | 1186 | 3.875211 | 3.31109 | 0 | 1 | 3 | 6 | 18 |
| **전체 뉴스 평균 감정 점수** | 1186 | -0.068784 | 0.351629 | -1 | -0.298525 | -0.0588 | 0.15 | 1 |
| **전체 뉴스 기사 수** | 1186 | 16.57167 | 3.709767 | 0 | 15 | 18 | 19 | 20 |

본 연구의 데이터 전처리는 다음과 같은 절차를 거쳐 수행되었다. 우선, 기준 데이터 설정이다. 전체 분석의 기준이 되는 날짜(Date) 변수와 원/달러 환율의 종가 및 시가 데이터를 구축하였다. 이때 주말과 공휴일 등 비영업일은 제거하여 실제 금융시장에서의 거래일만을 반영하였다.

두 번째 단계는 일 단위 변수 병합이다. 날짜를 기준으로 유가, 금 가격, 주요 주가지수, 변동성 지표와 같은 일 단위 변수들을 순차적으로 병합하였다. 각 월 단위의 거시경제 변수(예: 정책금리, 소비자물가지수(CPI), 생산자물가지수(PPI), 경상수지, 통화량(M1, M2))는 해당 월 발표 기준으로 월간 대표값을 설정하여, 해당 월의 모든 거래일에 동일하게 반영되도록 병합하였다.

세 번째는 GDELT 이벤트 데이터 및 뉴스 감성 결합이다. GDELT에서 추출한 이벤트 데이터와 뉴스 감성 데이터를 동일한 일 단위로 맞추어 결합하였다. 이벤트 데이터에는 글로벌 및 한국 관련 이벤트 수, 평균 톤 등이 포함되며, 뉴스 감성 변수는 직/간접적으로 관련된 뉴스의 일별 평균 감성 점수 및 기사 수로 구성된다.

네 번째는 결측치 처리이다. 뉴스 감성 데이터에서만 결측치가 발생했으며, 2020–2024년 기간 동안 총 13일치의 뉴스 감정 값이 결측되었다. 해당 결측치는 분석의 일관성을 유지하기 위해 0으로 대체하였다. 다른 변수들에서는 결측치가 없었고, 별도의 보정 과정은 필요하지 않았다. 이와 같은 절차를 통해 모든 데이터는 일 단위로 정렬된 통합형 시계열 데이터프레임으로 구성하였다.

마지막은 데이터 분리 및 정규화이다. 최종 데이터는 1,186개의 관측치를 기반으로 학습(80%)과 테스트(20%) 데이터로 분할하였으며, 모든 수치형 변수는 RobustScaler로 정규화하였다. RobustScaler는 중앙값과 사분위 범위를 활용하여 극단값(outlier)의 영향을 최소화하는 장점이 있어, 다양한 범위의 데이터에 대해 더욱 효과적인 정규화 방법을 제공한다.

4. 분석 방법론: LSTM, GRU, CNN-LSTM, CNN-GRU

본 연구는 환율의 단기 예측력을 비교 및 평가하기 위해 순환신경망(RNN) 계열의 단일 모형(LSTM, GRU)과 합성곱–순환 하이브리드 모형(CNN-LSTM, CNN-GRU)을 병렬적으로 구축하였다. 최근 외환시장은 거시, 금융, 원자재, 이벤트, 감정 등 이질적 신호가 동시적으로 상호작용하는 복합계로서, 저차원의 선형적 상호작용을 가정하는 통계 모형만으로는 비정형적 동학을 포착하기 어렵다. 이에 따라, 장기 의존성과 국소 패턴을 동시에 학습하는 딥러닝 아키텍처가 시계열 예측의 주된 대안으로 부상하고 있다.

1. 단일 RNN 모형: LSTM, GRU

LSTM은 Hochreiter와 Schmidhuber(1997)가 제안한 메모리 셀 구조로, 입력·망각·출력 게이트를 통해 vanishing/exploding gradient 문제를 완화하고 장기 의존성을 안정적으로 학습한다 [28]. 반면 GRU는 update/reset의 두 게이트로 구조를 단순화하여 매개변수 수를 절감하면서도 LSTM에 근접한 성능을 보이며, 동일한 데이터·에폭 하에서 수렴 속도와 연산 효율이 우수하다는 장점이 있다 [29]. 외환 예측 맥락에서 LSTM은 전통 RNN 대비 RMSE과 MAE를 유의하게 낮춘다는 보고가 다수 존재하며(예: EUR/USD 사례) [30], GRU는 비선형·비정상 구간에서 패턴 전이를 민감하게 추적해 CNN 대비 우수한 결과를 보이기도 한다(ARS/USD) [31]. 또한 두 모형의 보완적 강점을 결합할 경우, 단일 구조 대비 예측정확도가 향상된다는 하이브리드 근거도 축적되어 왔다 [32]. 종합하면, 선행연구들은 LSTM과 GRU가 외환 시장 예측에서 높은 잠재력을 지니며, 상황에 따라 하이브리드 접근이 예측 성능 개선에 유의미하게 작동할 수 있음을 보여준다.

입력은 전처리된 일 단위 통합 패널을 슬라이딩 윈도우로 절단하여 구성한다(윈도우 길이 𝐿, 예측 시차 𝐻). 각 시점의 피처는 환율 자체, 교차환율, 주가, 거래량, 원자재, 거시, 리스크, 이벤트 및 감정 지표로 이루어진다. LSTM/GRU 스택 뒤에 드롭아웃과 완전연결층을 배치하고, 손실함수로 MAE 또는 Huber(극단값 강건화), 최적화는 Adam을 사용한다. 조기종료와 러닝레이트 감쇠를 적용하고, 시계열 Walk Forward 검증으로 일반화 성능을 평가한다.

따라서 본 연구는 뉴스 요약 및 댓글 감정 점수와 같은 비정형 텍스트 기반 심리 변수를 포함한 시계열 데이터를 CNN-LSTM과 CNN-GRU 모델에 적용하였다. 이로써 CNN이 시장 심리의 국소적 변동을 효과적으로 포착하고, RNN 계열이 이를 시간 축에서 장기적 흐름으로 연결함으로써, 원/달러 환율 예측의 정밀도를 한층 높일 수 있을 것으로 기대된다.

본 연구는 환율의 단기 예측력을 비교 및 평가하기 위해 순환신경망(RNN) 계열의 단일 모형(LSTM, GRU)과 합성곱–순환 하이브리드 모형(CNN-LSTM, CNN-GRU)을 병렬적으로 구축하였다. 최근 외환시장은 거시, 금융, 원자재, 이벤트, 감정 등 이질적 신호가 동시적으로 상호작용하는 복합계로서, 저차원의 선형적 상호작용을 가정하는 통계 모형만으로는 비정형적 동학을 포착하기 어렵다. 이에 따라, 장기 의존성과 국소 패턴을 동시에 학습하는 딥러닝 아키텍처가 시계열 예측의 주된 대안으로 부상하고 있다.

1. 하이브리드 모형: CNN-LSTM, CNN-GRU

딥러닝 기반 시계열은 최근 국소 필터링(CNN)과 장기 의존 학습(RNN)을 결합하는 하이브리드로 진화하고 있다. CNN-LSTM/GRU는 1D 합성곱을 통해 시계열의 단주기 및 국소적 급등락, 스파이크, 군집 변동 등 미시 패턴을 전처리 단계에서 추출하고, 그 고수준 특징 맵을 LSTM 혹은 GRU에 전달하여 중장기 흐름을 학습한다. 이 구조는 잡음이 많은 고차원 입력에서 표현 효율성을 높이고, 장주기 신호의 단절 없는 전파를 돕는다. 금융 예측에서 CNN-LSTM이 단일 구조를 능가한다는 증거가 보고되어 있으며 [33], 투자자 심리를 포함할 때 추가적 개선이 관찰된다 [11].

합성곱 블록(1D Conv–ReLU–BatchNorm–MaxPool)을 1–2회 적용해 다중 스케일 커널(예: 3·5·7)로 이상치·고주파 잡음을 완화한 특징을 구성하고, 이후 LSTM/GRU로 시계열 의존을 학습한다. 하이브리드의 출력층과 훈련 체계(손실·최적화·정규화·검증)는 단일 RNN과 동일 원칙을 따른다. 특히 본 연구의 텍스트 기반 심리 변수(뉴스 요약 감정, 기사 수)와 이벤트 톤은 CNN에서 국소적으로 요약·정규화되어 RNN에 전달됨으로써, \*\*감정 급변(예: 지정학 이벤트 직후)\*\*이 환율에 미치는 단·중기 파급을 더 정밀하게 반영한다.

1. 모형 선택의 근거와 기대효과

딥러닝 기반 시계열은 최근 \*\*국소 필터링(CNN)\*\*과 \*\*장기 의존 학습(RNN)\*\*을 결합하는 하이브리드로 진화하고 있다. CNN-LSTM/GRU는 1D 합성곱을 통해 시계열의 단주기·국소적 급등락, 스파이크, 군집 변동 등 미시 패턴을 전처리 단계에서 추출하고, 그 고수준 특징 맵을 LSTM 혹은 GRU에 전달하여 중·장기 흐름을 학습한다. 이 구조는 잡음이 많은 고차원 입력에서 표현 효율성을 높이고, 장주기 신호의 단절 없는 전파를 돕는다. 금융 예측에서 CNN-LSTM이 단일 구조를 능가한다는 증거가 보고되어 있으며 [33], 투자자 심리를 포함할 때 추가적 개선이 관찰된다 [11].

LSTM/GRU는 금리와 물가 등 저주파 구조 신호의 누적 효과를, CNN은 VIX/OVX 급등, 이벤트 카운트 급증, 감정 급변 등 고주파 충격을 요약한다. 결합 시 다중 주기성(Multi-horizon Dynamics)을 공시적으로 반영한다. 또한, 원자재와 리스크 지표의 우측 장꼬리와 이벤트와 감정의 이질적 스케일은 합성곱, 풀링, 배치정규화로 1차 완충되고, RNN의 게이트로 2차 선별된다. 그리고 GRU 기반 하이브리드는 매개변수가 상대적으로 적어 적은 에폭에서도 수렴이 빠르며, 과적합 위험에 더 강건하다.

5. 예측 성능평가 지표

본 연구에서는 원/달러 환율 예측 모델의 성능을 다각적 관점에서 평가하기 위해, 총 여섯 가지 회귀 지표를 활용하였다: RMSE (Root Mean Squared Error), MSPE (Mean Squared Percentage Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MedAE (Median Absolute Error), MedAPE (Median Absolute Percentage Error). 이들 지표는 예측 모델이 제공하는 성능을 정확도, 강건성(robustness), 그리고 이상치에 대한 민감도까지 종합적으로 평가할 수 있게 해준다.

RMSE는 예측 오차의 제곱을 평균한 후 그 제곱근을 취한 값으로, 오차의 크기를 더 강하게 반영하는 특징이 있다. 예측 값과 실제 값의 차이가 클수록 제곱이 되어 RMSE 값이 더 크게 증가하므로, 모델의 오차 제어 능력을 평가하는 데 매우 유용하다. RMSE는 데이터의 단위를 그대로 유지하므로 해석이 직관적이고, 모델의 예측 성능을 정량적으로 평가하는 데 널리 사용된다. 특히 큰 오차가 중요한 예측 상황에서는 매우 중요한 지표로, 예측 정확도를 극대화하는 모델을 찾는 데 도움을 준다.

MSPE는 예측 오차를 실제값으로 정규화한 후 제곱하여 평균한 값이다. 이는 상대적 오차 크기에 주목하여 예측의 정확도를 평가하며, 모델이 예측 오차의 상대적 크기에 얼마나 민감한지를 평가할 수 있다. MSPE는 예측 오차가 모델의 편향(bias)과 분산(variance)을 동시에 반영하기 때문에, 모델의 정확성뿐만 아니라 분포의 불균형에도 민감하게 반응할 수 있다. 또한, 상대적인 오차를 강조하여, 편향이나 고른 분산을 평가하는 데 중요한 역할을 한다.

MAE는 예측값과 실제값 간의 절대적 차이를 평균한 값으로, 각 오차를 동일한 가중치로 반영하여 평균적인 예측 정확도를 평가하는 데 유용하다. RMSE와 비교할 때, 이상치에 대한 민감도가 적어 이상치가 존재하는 데이터에서 모델의 성능을 안정적으로 평가할 수 있다. MAE는 직관적으로 이해할 수 있는 지표로, 예측이 잘 맞았는지 아닌지를 빠르게 파악하는 데 유용하다. 또한, 오차 크기에 대해 균등한 평가를 제공하여, 모델의 일관된 예측 정확도를 확인하는 데 적합하다.

MAPE는 절대 오차를 실제값으로 나눈 뒤 백분율로 환산한 값이다. 이를 통해 모델의 오차를 직관적으로 이해할 수 있으며, 비율 오차를 기반으로 예측 성능을 평가할 수 있다. 그러나 실제값이 0이거나 매우 작은 경우에는 무한대의 값이 발생할 수 있어 이 지표의 사용에 주의가 필요하다. 이런 문제로 인해 일부 연구에서는 MAPE의 대안 지표를 제시하기도 한다. 그럼에도 불구하고, MAPE는 예측이 잘 이루어졌는지 빠르게 파악할 수 있는 유용한 지표로 여전히 많은 경우에 사용된다.

MedAE는 절대 오차의 중앙값을 사용하여 이상치에 대한 영향을 최소화한 지표이다. 평균(Mean)이 아닌 중앙값(Median)을 사용함으로써, 데이터 분포의 왜곡에 덜 민감하며 모델이 일관되게 성능을 발휘하는지 평가할 수 있다. MedAE는 이상치(outlier)가 존재하는 경우에도 모델의 예측 성능을 더 안정적으로 평가할 수 있기 때문에, 특히 비정상적인 분포나 급격한 변동이 있는 데이터셋에서 유용하게 활용될 수 있다.

MedAPE는 절대 백분율 오차의 중앙값을 사용하여 MAPE의 직관성을 유지하면서도 극단적인 오차의 영향을 완화하는 지표이다. MAPE와 유사하게 백분율 기준으로 예측 성능을 평가하지만, 중앙값(Median)을 사용하여 이상치에 강건한 특성을 가짐으로써, 금융 시계열과 같이 이상치가 자주 발생하는 데이터셋에서 더 안정적이고 정확한 평가가 가능하다.

이처럼 다양한 예측 성능 지표를 함께 활용함으로써, 본 연구는 모델의 평균적 예측 정확도뿐만 아니라 이상치에 대한 강건성(robustness)과 실제 금융 예측 환경에서의 해석 가능성을 종합적으로 검증하였다. 여러 지표를 활용하는 것은 각 지표가 모델의 다른 성능 측면을 강조하기 때문이다. 예를 들어, RMSE는 일관된 환율예측의 정확성을, MSPE는 높은 변동성에도 정확한 예측을, MAE와 MAPE는 전반적인 정확도 이해를 강조한다. 또한, MedAE와 MedAPE는 이상치가 많은 비정상적인 환율 변동에도 강건하고 안정적인 예측을 확인하는 데 중요한 역할을 한다. 따라서, 이들 지표를 종합적으로 고려하여 본 연구에서는 모델 성능의 균형을 잡고, 여러 가지 측면에서의 평가를 통해 예측 모델이 금융 시장에서의 실제 예측 환경에 적합한지를 다각도로 검토하였다.

Ⅲ. 연구결과

3.1 예측 및 성능평가 과정

본 연구는 원/달러 환율의 다음 영업일 종가(1-step ahead) 예측을 목표로, LSTM, GRU, CNN-LSTM, CNN-GRU의 네 가지 딥러닝 아키텍처를 동일한 실험 설계 하에 비교하였다. 전체 표본은 시계열 순서를 보존한 상태로 학습 80% / 테스트 20%로 분할하였으며, 시간의 전진 정보 누설을 차단하기 위해 모든 전처리와 모형 적합은 학습 세트 기준으로 학습된 매개변수만을 사용하여 테스트 세트에 적용하였다. 예측력의 과거 정보 활용 범위를 점검하기 위해 입력 시퀀스 길이(lookback window)를 5, 10, 20, 30, 60, 90일로 그리드화하여, 단/중/장기 메모리 길이 변화가 성능에 미치는 영향을 체계적으로 분석하였다.

모형의 추정 불확실성과 초기화 민감도를 완화하기 위해, 각 윈도우와 입력변수 조합마다 난수 시드 42, 55, 68을 적용해 3회 반복 학습을 수행하였고, 보고치는 이들의 산술평균으로 제시하였다(표준편차는 부록 표에 병기). 입력 특성은 모든 실험에서 RobustScaler로 정규화하였다. 본 스케일러는 중앙값과 사분위 범위를 사용하여 극단값(outlier)에 대한 민감도를 낮추므로, 고변동과 장꼬리 분포를 보이는 금융 시계열의 스케일 불안정성을 완충하는 데 적합하다.

입력 변수는 다음의 네 가지 케이스로 구성하여 구조적 요인과 사건 및 심리 요인의 한계 기여를 분리 확인하였다. (1) Baseline: 전통적인 거시 및 시장 변수만을 포함하여 예측, (2) Baseline+이벤트 데이터: Baseline에 GDELT 이벤트 변수를 포함하여 예측, (3) Baseline+감성 데이터: Baseline에 뉴스 요약 정보와 감정 변수 포함하여 예측, (4) Baseline에 이벤트와 감성 데이터를 모두 반영하여 예측한 조합이다.

모형 아키텍처는 다음과 같이 동일한 설계 원칙을 유지하였다. LSTM/GRU는 각각 64 → 32 유닛의 2층 순환 스택으로 구성하고, 층 사이에 Dropout=0.1을 적용하여 과적합을 억제하였다. CNN-LSTM/CNN-GRU는 입력 시계열에 1D 합성곱(Conv1D, kernel\_size=3, causal padding, ReLU)을 적용해 국소적인 급등락이나 스파이크와 같은 패턴을 추출한 뒤, Batch Normalization과 Dropout을 거쳐 각각 LSTM(64→32) 또는 GRU(64→32) 블록으로 중장기 의존성을 학습하도록 구성하였다. 모든 모델의 출력층은 Dense(1)이며, 손실함수는 MSE로 고정하였다. 최적화, 배치 크기, 에폭, 조기 종료 등 세부 학습 하이퍼파라미터는 <표 5>에 요약하였다.

**<표 5> 모델별 세부 하이퍼파라미터**

|  |  |  |
| --- | --- | --- |
| 알고리즘 | 주요 파라미터 | 값 |
| **LSTM** | Unit Numbers | 64 → 32 (2-스택) |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **GRU** | Unit Numbers | 64 → 32 (2-스택) |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **CNN-LSTM** | Filters | 32 |
| Kernel Size | 3 (causal padding) |
| Activation | ReLU |
| Batch Normalization | 적용 |
| LSTM Units | 64 → 32 |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **CNN-GRU** | Filters | 32 |
| Kernel Size | 3 (causal padding) |
| Activation | ReLU |
| Batch Normalization | 적용 |
| GRU Units | 64 → 32 |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |

모델 성능 평가는 MSE, RMSE, MAE, MedAE, MAPE, MSPE, MedAPE의 6가지 지표를 사용하였다. MSE, RMSE, MAE, MedAE는 원 단위로, MAPE, MSPE, MedAPE는 백분율 단위로 산출되었다.

본 연구는 평균 정확도, 상대 오차, 강건성의 세 축을 동시에 점검하기 위해 RMSE, MSPE, MAE, MAPE, MedAE, MedAPE의 여섯 가지의 평가지표를 사용하였다. 또한 모든 결과는 테스트 구간 전체에 대한 지표를 기본으로 보고하고, 윈도우 길이 및 입력 변수 케이스별로 상대적 우위 패턴을 비교하였다. 이와 같은 통일된 실험은 데이터 분할 일관성, 입력 스케일 처리의 재현성, 초기화 불확실성의 평균화라는 세 측면에서 공정한 모형 간 비교 가능성을 확보한다.

요약하면, (모형) LSTM과 GRU 단일 구조와 하이브리드 구조, (메모리 길이) 5~90일 lookback, (입력 집합) 구조적 요인과 사건 및 심리 요인의 단계적 추가, (평가축) 절대, 상대 오차 및 강건성 등 네 축을 직교적으로 결합함으로써, 환율 예측에서 심리와 이벤트 변수의 한계 기여와 하이브리드 구조의 효용을 동시에 검증하도록 설계되었다.

3.2 실험 결과

단기(lookback 5일)에서는 합성곱–순환 하이브리드(CNN-LSTM, CNN-GRU)가 순수 RNN(LSTM, GRU) 대비 전반적으로 우수했다. CNN-LSTM(Macro+Event)이 최저 RMSE=7.162로 절대오차 최소화에서 1위, CNN-GRU(Macro+Event)가 나머지 모든 지표에서 1위를 기록했다. 이벤트 지표 추가는 단기 창구에서 일관된 성능 개선을 제공했으나, 감성 단독 추가의 개선폭은 제한적이었다. 이벤트와 감성 동시 반영은 오히려 미세한 성능 저하 또는 정체를 보였다. GRU는 모든 케이스에서 RMSE ≥ 7.376으로 관측되어, 5일 창구의 고주파 변동 포착에서 상대적 열위를 보였다. 다만 CNN 전처리를 결합하면 성능이 회복되었다. 단기 절대오차를 최소화하려면 CNN-LSTM(Macro+Event), 상대 및 중앙값 지표 중심의 강건성을 중시하면 CNN-GRU(Macro+Event)가 바람직하다. 그리고 이벤트 기반 핵심 신호 중심의 경량 피처링이 효과적이다.

더욱 상세한 수치적 해석을 통한 결과를 요약하면 아래와 같다. 첫째, 합성곱–순환 하이브리드의 구조적 이점이 단기 구간에서 분명하게 드러났다. 전반적인 RMSE가 7.1~7.6 범위로 수렴하는 가운데, 최저치는 CNN-LSTM(Macro+Event)의 7.162로 확인되며, 동일 창구의 최저 성능 대비 약 6~7% 수준의 절대오차 축소 효과를 보였다. 이는 합성곱(Conv1D)이 5거래일 수준에서 빈번히 나타나는 국소적 급등락이나 스파이크를 사전에 요약 및 정규화하고, 그 고수준 특징을 LSTM이 시간 축으로 연계 학습하는 분업 구조가 유효했음을 시사한다. 둘째, 이벤트 지표의 기여가 감성 지표를 상회한다. 같은 모델 내에서 Macro Only에서 Macro+Event 전환 시, CNN-LSTM은 RMSE가 7.403에서 7.162으로, CNN-GRU는 MAPE가 0.411에서 0.398로 개선되었다. 반면 Macro+Sentiment는 개선폭이 작거나 혼재적이다. 이는 사건 발생량과 톤 변화가 단기 환율 변동의 즉시적 충격을 더 명료하게 전달하며, 합성곱 블록이 그 신호를 효과적으로 증폭 및 정제함을 의미한다. 셋째, 강건성 지표에서도 하이브리드와 이벤트 조합이 우위다. CNN-GRU(Macro+Event)는 MedAE와 MedAPE가 각각 4.448과 0.328로, 같은 GRU 계열의 모든 조합 대비 각각 약 7.0%와 6.8% 개선된다. 이는 지정학 이벤트 직후나 VIX 급등일 등 꼬리 위험 구간에서 오차 분포의 두터운 꼬리(tail)를 효과적으로 억제함을 보여준다. 넷째, 모든 변수를 반영한 조합의 효율 저하가 관찰된다. 네 모델 공통으로 최적 조합을 능가하지 못하거나 유사 수준에 그쳤으며, 이는 다중공선성 또는 중복으로 인한 일부 과적합 가능성을 시사한다. 단기 예측에서는 핵심 신호(이벤트) 위주의 절제된 피처링이 더 효율적이다.

<표 6> LookBack : 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.229 | 5.557 | 0.407 | 4.609 | 0.341 |
| LSTM | Macro + Event | 7.316 | 5.619 | 0.412 | 4.606 | 0.34 |
| LSTM | Macro + Sentiment | 7.183 | 5.521 | 0.405 | 4.623 | 0.343 |
| LSTM | ALL | 7.317 | 5.645 | 0.414 | 4.66 | 0.337 |
| GRU | Macro Only | 7.376 | 5.676 | 0.416 | 4.58 | 0.337 |
| GRU | Macro + Event | 7.552 | 5.816 | 0.426 | 4.749 | 0.346 |
| GRU | Macro + Sentiment | 7.566 | 5.852 | 0.428 | 4.755 | 0.35 |
| GRU | ALL | 7.67 | 5.929 | 0.434 | 4.782 | 0.352 |
| CNN\_LSTM | Macro Only | 7.403 | 5.625 | 0.413 | 4.609 | 0.339 |
| CNN\_LSTM | Macro + Event | 7.162 | 5.44 | 0.399 | 4.549 | 0.336 |
| CNN\_LSTM | Macro + Sentiment | 7.168 | 5.468 | 0.401 | 4.551 | 0.338 |
| CNN\_LSTM | ALL | 7.243 | 5.553 | 0.407 | 4.616 | 0.341 |
| CNN\_GRU | Macro Only | 7.41 | 5.607 | 0.411 | 4.546 | 0.332 |
| CNN\_GRU | Macro + Event | 7.169 | 5.428 | 0.398 | 4.448 | 0.328 |
| CNN\_GRU | Macro + Sentiment | 7.24 | 5.55 | 0.407 | 4.55 | 0.339 |
| CNN\_GRU | ALL | 7.288 | 5.593 | 0.41 | 4.61 | 0.338 |

Lookback 10일 환경에서의 비교 결과, 절대오차 기준의 최저 성능은 LSTM(Macro Only)가 RMSE=7.089로 기록하여, 하이브리드보다 순수 RNN이 우세한 드문 사례가 확인되었다. 반면 비율이나 중앙값 지표에서는 CNN-GRU 계열이 강세로, CNN-GRU(Macro Only)가 MAE=5.420, MAPE=0.397로 최저치를 보였고, CNN-GRU(Macro+Event)는 MedAE=4.436, MedAPE=0.327로 강건성(robustness) 면에서 최적 조합이었다. GRU 단일 구조는 여전히 열위였으나, CNN 전처리를 결합하면 비율과 중앙값 지표에서 1위를 차지해 구조적 한계를 상당 부분 상쇄하였다. 특징 변수의 효과는 모델 구조에 따라 상이했는데, CNN-LSTM에서는 감성 추가가 RMSE, MAE, MedAPE를 고르게 개선한 반면, 이벤트 추가는 오히려 악화되었다. LSTM은 이벤트 추가 시 중앙값 지표 개선이 관찰되었고, CNN-GRU는 이벤트 추가 시 MedAE, MedAPE가 최저로 떨어져 스트레스 국면에서의 안정성이 돋보였다. 전반적으로, 절대오차 최소화는 LSTM(Macro Only), 상대오차와 중앙값 기반 강건성은 CNN-GRU(Macro+Event)가 각각 최적 선택으로 요약된다.

더욱 상세한 수치적 해석을 통한 결과를 요약하면 아래와 같다. 첫째, 절대오차 최소화(RMSE) 관점에서 LSTM(Macro Only)가 최저 RMSE=7.089로 확인되었다. 이는 Lookback 10일 구간에서 거시 및 시장 변수만으로도 저주파 신호를 안정적으로 학습하는 데 LSTM의 게이트 메모리가 충분히 작동했음을 시사하며, 이때 합성곱 전처리는 필수 요건이 아니었다. 둘째, MAPE와 MAE는 CNN-GRU(Macro Only)가 각각 0.397와 5.420으로 최저를 기록했다. GRU 단독 최악 조합 대비 약 10% 내외의 개선으로, 합성곱이 단주기 변동과 잡음을 요약해 GRU의 간결한 게이트가 과도한 파라미터 증가 없이 효율적으로 흡수한 결과로 해석된다. 셋째, 이상치와 꼬리 안정성은 CNN-GRU(Macro+Event)에서 최적화되었다. 이는 같은 GRU계의 ALL 대비 각각 약 8.9%과 7.9% 개선으로, 이벤트 기반 충격이 국소 급변의 꼬리 위험을 설명하는 데 기여하고, CNN–GRU 결합이 그 신호를 왜곡 없이 정규화함을 시사한다. 넷째, 이벤트와 감성의 기여는 구조 의존적이다. CNN-LSTM에서는 감성 추가가 RMSE, MAE, MedAPE를 동시에 개선했으나, 이벤트 추가는 오히려 RMSE 악화로 이어졌다. 반면 CNN-GRU는 이벤트 추가가 중앙값 지표를 최저로 끌어내려 스트레스 구간 대응력을 강화했다. 즉, 감성은 서서히 축적되는 서술적 신호로서 LSTM과의 결합에서 평균 오차를 줄이는 데 유효하고, 이벤트는 돌발과 국소 충격으로서 GRU와의 결합에서 tail 억제에 더 효과적이라는 역할 분화가 드러난다.

<표 7> LookBack : 10

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.089 | 5.468 | 0.401 | 4.665 | 0.343 |
| LSTM | Macro + Event | 7.11 | 5.479 | 0.401 | 4.569 | 0.332 |
| LSTM | Macro + Sentiment | 7.124 | 5.511 | 0.404 | 4.684 | 0.347 |
| LSTM | ALL | 7.103 | 5.497 | 0.403 | 4.628 | 0.344 |
| GRU | Macro Only | 7.739 | 6.055 | 0.442 | 4.802 | 0.352 |
| GRU | Macro + Event | 7.596 | 5.917 | 0.432 | 4.707 | 0.348 |
| GRU | Macro + Sentiment | 7.595 | 5.924 | 0.433 | 4.87 | 0.355 |
| GRU | ALL | 7.56 | 5.869 | 0.429 | 4.812 | 0.354 |
| CNN\_LSTM | Macro Only | 7.254 | 5.586 | 0.409 | 4.716 | 0.345 |
| CNN\_LSTM | Macro + Event | 7.4 | 5.65 | 0.414 | 4.716 | 0.345 |
| CNN\_LSTM | Macro + Sentiment | 7.177 | 5.502 | 0.403 | 4.496 | 0.334 |
| CNN\_LSTM | ALL | 7.165 | 5.523 | 0.404 | 4.787 | 0.35 |
| CNN\_GRU | Macro Only | 7.116 | 5.42 | 0.397 | 4.519 | 0.331 |
| CNN\_GRU | Macro + Event | 7.19 | 5.428 | 0.398 | 4.436 | 0.327 |
| CNN\_GRU | Macro + Sentiment | 7.158 | 5.508 | 0.403 | 4.776 | 0.354 |
| CNN\_GRU | ALL | 7.126 | 5.442 | 0.399 | 4.731 | 0.346 |

LookBack 20일은 모든 모델에서 가장 우수한 성능을 나타낸 핵심 구간이었다. 절대오차(RMSE) 최저치는 LSTM(Macro+Event)에서 확인되었다(RMSE=6.946). MAE 최저는 CNN-GRU(Macro+Event)가 기록했다(MAE=5.383). MAPE 최저 역시 CNN-GRU(Macro+Event)가 0.394로 1위를 차지했고, LSTM(Macro+Event)와 CNN-LSTM(ALL)도 유사한 수치를 형성했다. 강건성 지표에서도 CNN-GRU(Macro+Event)가 MedAE=4.440, MedAPE=0.326으로 최저치를 보여, 가장 우수한 모델로 채택되었다.

반대로 GRU 단일 구조는 Macro Only일 때 RMSE는 비교적 양호하나, 이벤트/감성/ALL의 경우에선 일관되게 악화되는 패턴이 관찰되었다. CNN-LSTM은 이벤트만 추가 시 성능이 다소 저하되지만, 감성만 추가 또는 ALL에서는 개선되어 특징 간 상호작용 효과가 확인되었다. 요약하면, 절대오차 최소화는 LSTM(Macro+Event), 나머지 모든 지표에서는 CNN-GRU(Macro+Event)가 최적 선택으로 정리된다.

<표 8> LookBack : 20

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.094 | 5.56 | 0.407 | 4.654 | 0.341 |
| LSTM | Macro + Event | 6.946 | 5.384 | 0.394 | 4.562 | 0.333 |
| LSTM | Macro + Sentiment | 7.053 | 5.533 | 0.405 | 4.663 | 0.341 |
| LSTM | ALL | 7.086 | 5.567 | 0.407 | 4.652 | 0.341 |
| GRU | Macro Only | 7 | 5.424 | 0.397 | 4.585 | 0.337 |
| GRU | Macro + Event | 7.303 | 5.692 | 0.416 | 4.819 | 0.354 |
| GRU | Macro + Sentiment | 7.375 | 5.767 | 0.421 | 4.761 | 0.346 |
| GRU | ALL | 7.376 | 5.717 | 0.418 | 4.799 | 0.351 |
| CNN\_LSTM | Macro Only | 7.101 | 5.483 | 0.401 | 4.484 | 0.329 |
| CNN\_LSTM | Macro + Event | 7.173 | 5.557 | 0.407 | 4.611 | 0.338 |
| CNN\_LSTM | Macro + Sentiment | 7.115 | 5.501 | 0.402 | 4.527 | 0.332 |
| CNN\_LSTM | ALL | 6.976 | 5.39 | 0.394 | 4.519 | 0.335 |
| CNN\_GRU | Macro Only | 7.128 | 5.481 | 0.401 | 4.519 | 0.33 |
| CNN\_GRU | Macro + Event | 6.996 | 5.383 | 0.394 | 4.44 | 0.326 |
| CNN\_GRU | Macro + Sentiment | 7.289 | 5.727 | 0.419 | 4.804 | 0.351 |
| CNN\_GRU | ALL | 7.094 | 5.5 | 0.402 | 4.629 | 0.34 |

Lookback 30일에서는 LSTM(Macro+Event)가 평균 오차를 최소화하고, CNN-LSTM(Macro+Event)이 나머지 지표들의 성능을 극대화 하는 구도가 뚜렷하게 나타난다. RMSE과 MAE에서는 최적 조합이 LSTM(Macro+Event)로 수렴한다. MAPE 역시 LSTM(Macro+Event)가 0.392로 최저이며, CNN-LSTM(+Event)가 근소하게 뒤따른다. 반면 MedAE, MedAPE은 CNN-LSTM(Macro+Event)가 최저치를 기록한다. GRU 단일 구조는 전반적으로 열위이며, 이벤트와 감성 추가로도 절대오차 측면의 반전은 제한적이었다. 다만 CNN 전처리를 결합한 GRU는 감성 또는 이벤트 추가 시 MAPE와 MedAPE에서 유의한 개선이 확인된다.

<표 9> LookBack : 30

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.122 | 5.523 | 0.404 | 4.536 | 0.333 |
| LSTM | Macro + Event | 7.002 | 5.363 | 0.392 | 4.441 | 0.321 |
| LSTM | Macro + Sentiment | 7.146 | 5.593 | 0.409 | 4.66 | 0.341 |
| LSTM | ALL | 7.116 | 5.565 | 0.406 | 4.666 | 0.344 |
| GRU | Macro Only | 7.598 | 5.985 | 0.436 | 4.801 | 0.352 |
| GRU | Macro + Event | 7.483 | 5.841 | 0.426 | 4.809 | 0.352 |
| GRU | Macro + Sentiment | 7.642 | 6.047 | 0.441 | 4.991 | 0.363 |
| GRU | ALL | 7.569 | 5.979 | 0.436 | 4.833 | 0.354 |
| CNN\_LSTM | Macro Only | 7.052 | 5.426 | 0.397 | 4.533 | 0.331 |
| CNN\_LSTM | Macro + Event | 7.098 | 5.411 | 0.396 | 4.366 | 0.317 |
| CNN\_LSTM | Macro + Sentiment | 7.205 | 5.605 | 0.409 | 4.562 | 0.336 |
| CNN\_LSTM | ALL | 7.155 | 5.539 | 0.405 | 4.688 | 0.343 |
| CNN\_GRU | Macro Only | 7.543 | 5.761 | 0.421 | 4.718 | 0.347 |
| CNN\_GRU | Macro + Event | 7.353 | 5.593 | 0.409 | 4.676 | 0.341 |
| CNN\_GRU | Macro + Sentiment | 7.178 | 5.533 | 0.404 | 4.488 | 0.325 |
| CNN\_GRU | ALL | 7.336 | 5.7 | 0.416 | 4.693 | 0.34 |

Lookback 60일과 90일 구간에서는 모든 모델에서 RMSE가 다시 증가하였다. 특히 LookBack 60에서는 LSTM 계열이 전반 우위를 보였다. RMSE, MAE, MAPE)는 LSTM(ALL)이 1위를 기록했다. 반면 MedAE, MedAPE는 LSTM(Macro+Event)가 최저로 확인되었다. GRU 단일 구조는 전체적으로 열위였고, 특히 GRU(ALL)이 최악(RMSE=8.172)을 기록하여 장기 윈도우에서의 과적합과 잡음 민감성이 부각되었다. CNN 결합 모델은 평균 지표에서 LSTM에 못 미쳤고, 일부 조합(CNN-LSTM+Event 등)은 오히려 RMSE가 악화되었다. 종합하면, 60일 창구에서는 LSTM의 장기 의존 학습이 가장 안정적이며, 이벤트 추가는 꼬리 안정성에, ALL은 평균 오차에 각각 유리했다.

마지막으로 LookBack 90에서는 RMSE 최저가 LSTM(ALL)로 나타났다. MAE, MAPE 최저는 CNN-GRU(Macro Only)가 각각 5.511와 0.401로 기록하여, 장기 창에서는 GRU도 합성곱 전처리만으로 오차를 낮출 수 있음을 보여줬다. 강건성 지표는 LSTM(Macro+Event)가 MedAE=4.358, MedAPE=0.316으로 최저를 기록해 tail 안정성 면에서 가장 우수했다. 한편 GRU 단일 구조는 여전히 평균/강건성 모두에서 열위였고, CNN-LSTM은 평균 지표에서 LSTM에 못 미치지만, Event를 추가하면 MedAE, MedAPE가 개선되어 꼬리 억제 효과가 확인되었다.

<표 10> LookBack : 60

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.342 | 5.623 | 0.41 | 4.476 | 0.326 |
| LSTM | Macro + Event | 7.339 | 5.574 | 0.406 | 4.404 | 0.317 |
| LSTM | Macro + Sentiment | 7.367 | 5.673 | 0.413 | 4.58 | 0.334 |
| LSTM | ALL | 7.244 | 5.555 | 0.405 | 4.491 | 0.325 |
| GRU | Macro Only | 7.927 | 6.097 | 0.444 | 5.039 | 0.368 |
| GRU | Macro + Event | 7.849 | 6.05 | 0.44 | 5.01 | 0.366 |
| GRU | Macro + Sentiment | 7.581 | 5.9 | 0.429 | 4.756 | 0.343 |
| GRU | ALL | 8.172 | 6.38 | 0.464 | 5.3 | 0.387 |
| CNN\_LSTM | Macro Only | 7.276 | 5.579 | 0.406 | 4.468 | 0.327 |
| CNN\_LSTM | Macro + Event | 7.429 | 5.64 | 0.411 | 4.743 | 0.343 |
| CNN\_LSTM | Macro + Sentiment | 7.474 | 5.757 | 0.419 | 4.7 | 0.343 |
| CNN\_LSTM | ALL | 7.395 | 5.695 | 0.415 | 4.609 | 0.337 |
| CNN\_GRU | Macro Only | 7.414 | 5.632 | 0.41 | 4.806 | 0.344 |
| CNN\_GRU | Macro + Event | 7.421 | 5.66 | 0.412 | 4.601 | 0.331 |
| CNN\_GRU | Macro + Sentiment | 7.527 | 5.835 | 0.424 | 4.731 | 0.343 |
| CNN\_GRU | ALL | 7.493 | 5.729 | 0.417 | 4.595 | 0.337 |

<표 11> LookBack : 90

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.261 | 5.535 | 0.403 | 4.522 | 0.329 |
| LSTM | Macro + Event | 7.281 | 5.544 | 0.404 | 4.358 | 0.316 |
| LSTM | Macro + Sentiment | 7.294 | 5.592 | 0.407 | 4.59 | 0.333 |
| LSTM | ALL | 7.238 | 5.548 | 0.404 | 4.451 | 0.326 |
| GRU | Macro Only | 7.665 | 5.868 | 0.426 | 4.768 | 0.345 |
| GRU | Macro + Event | 7.803 | 5.996 | 0.436 | 4.956 | 0.364 |
| GRU | Macro + Sentiment | 7.789 | 6.052 | 0.439 | 4.905 | 0.359 |
| GRU | ALL | 7.709 | 5.951 | 0.433 | 4.921 | 0.356 |
| CNN\_LSTM | Macro Only | 7.382 | 5.644 | 0.411 | 4.638 | 0.338 |
| CNN\_LSTM | Macro + Event | 7.392 | 5.64 | 0.41 | 4.427 | 0.323 |
| CNN\_LSTM | Macro + Sentiment | 7.403 | 5.677 | 0.413 | 4.601 | 0.337 |
| CNN\_LSTM | ALL | 7.458 | 5.739 | 0.417 | 4.612 | 0.339 |
| CNN\_GRU | Macro Only | 7.273 | 5.511 | 0.401 | 4.694 | 0.338 |
| CNN\_GRU | Macro + Event | 7.35 | 5.61 | 0.408 | 4.549 | 0.329 |
| CNN\_GRU | Macro + Sentiment | 7.513 | 5.777 | 0.42 | 4.758 | 0.353 |
| CNN\_GRU | ALL | 7.517 | 5.744 | 0.417 | 4.712 | 0.346 |

Lookback별 평균 성능을 종합하면, 전체 모델과 케이스의 평균 RMSE는 20일 구간에서 7.132로 가장 낮았고, 그 다음이 10일(7.281), 30일(7.287) 순이었다. 반면 60일(7.516)과 90일(7.458)은 가장 높은 값을 나타냈다. 이는 지나치게 짧거나 긴 구간보다는 중간 길이의 시계열 창이 예측 안정성과 변동 포착의 균형을 가장 잘 달성한다는 점을 시사한다.

<표 12> lookback별 평균 지표값

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **lookback** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** | **MSE** | **MSPE** |
| 20 | 7.132 | 5.542 | 0.405 | 4.627 | 0.339 | 50.93 | 0.003 |
| 10 | 7.281 | 5.611 | 0.411 | 4.682 | 0.344 | 53.084 | 0.003 |
| 30 | 7.287 | 5.654 | 0.413 | 4.654 | 0.34 | 53.21 | 0.003 |
| 5 | 7.331 | 5.617 | 0.412 | 4.615 | 0.34 | 53.788 | 0.003 |
| 90 | 7.458 | 5.714 | 0.416 | 4.654 | 0.339 | 55.716 | 0.003 |
| 60 | 7.516 | 5.774 | 0.42 | 4.707 | 0.342 | 56.623 | 0.003 |

모델별·케이스별 최적 조합을 살펴보면, LSTM 계열은 주로 20일 구간에서 최소 RMSE를 기록하였고, GRU 계열 또한 대부분의 케이스가 동일 구간에서 최적점을 보였다. CNN-LSTM은 ‘Macro Only’ 및 ‘Macro + Event’ 조합에서 30일, 나머지 케이스에서 20일이 최적이었으며, CNN-GRU는 단기(10일) 및 중기(20일) 구간에서 우수한 결과를 보였다. 특히 CNN-LSTM의 ‘ALL’ 조합은 RMSE 6.976, MAE 5.390, MAPE 0.394%로 전체 실험 중 가장 균형 잡힌 성능을 나타냈으며, 이는 이벤트 및 감성 정보가 거시변수의 한계를 보완하며 환율의 단기·중기 변동을 효과적으로 설명함을 보여준다.

<표 13> 모델별 케이스별 최적 조합

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **Best Lookback** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro + Event | 20 | 6.946 | 5.384 | 0.394 | 4.562 | 0.333 |
| CNN\_LSTM | ALL | 20 | 6.976 | 5.39 | 0.394 | 4.519 | 0.335 |
| CNN\_GRU | Macro + Event | 20 | 6.996 | 5.383 | 0.394 | 4.44 | 0.326 |
| GRU | Macro Only | 20 | 7 | 5.424 | 0.397 | 4.585 | 0.337 |
| CNN\_LSTM | Macro Only | 30 | 7.052 | 5.426 | 0.397 | 4.533 | 0.331 |
| LSTM | Macro + Sentiment | 20 | 7.053 | 5.533 | 0.405 | 4.663 | 0.341 |
| LSTM | ALL | 20 | 7.086 | 5.567 | 0.407 | 4.652 | 0.341 |

종합적으로 볼 때, Lookback 20일 구간에서 모델의 예측력이 가장 높았으며, CNN-LSTM과 CNN-GRU 등 합성곱 기반 순환 모델이 전통적 RNN 구조보다 전반적으로 우수한 결과를 보였다. 이러한 결과는 CNN 계층이 단기적 급변 패턴을 포착하고, LSTM/GRU 계층이 시간적 연속성을 학습함으로써 서로 보완적인 역할을 수행했기 때문으로 해석된다. 또한 이벤트(GDELT)와 뉴스 감성 변수를 함께 포함한 ‘ALL’ 조합의 성능이 모든 Lookback 구간에서 안정적으로 우수하게 나타나, 거시적 요인에 더해 정성적 요인을 통합적으로 반영하는 접근법이 환율 예측에 효과적임을 실증적으로 확인할 수 있었다.

3.2 변수 중요도 분석

모델이 어떤 정보를 통해 예측력을 확보했는지를 정량적으로 파악하기 위해 Permutation Importance 분석을 수행하였다. 이 기법은 개별 피처의 값을 무작위로 섞은 후, 그로 인한 모델 예측 성능(MSE) 저하 정도를 계산함으로써 해당 피처가 모델 성능에 기여하는 상대적 중요도를 측정한다. 따라서 회귀계수에 직접 의존하지 않고, 비선형적 구조와 변수 간 상호작용을 고려할 수 있다는 점에서 딥러닝 기반 시계열 모델의 해석 가능성을 보완하는 데 유용하다.

분석 결과, S&P500 종가(0.397)가 전체 변수 중 가장 높은 중요도를 보였으며, 그 뒤를 경상수지(0.348), 다우존스 지수 (0.281)가 차지하였다. 이는 미국 증시 및 대외수지와 같은 글로벌 금융·거시 변수들이 원/달러 환율 변동에 가장 직접적인 영향을 미친다는 점을 보여준다. 특히 S&P500과 다우존스 지수는 투자심리와 위험자산 선호도를 반영하는 대표적 글로벌 지표로서, 환율의 단기·중기 변동에 큰 설명력을 제공하는 것으로 나타났다.

흥미로운 점은 events\_tone\_mean(0.229)이 네 번째로 높은 중요도를 기록했다는 것이다. 이는 GDELT 이벤트 데이터에서 추출된 사건 톤(감성)의 평균값으로, 글로벌 뉴스 내 긍·부정 정서의 편향이 환율 움직임과 유의한 상관관계를 가진다는 사실을 의미한다. 특히 이벤트 감성 변수의 영향력이 CPI(0.121)나 WTI(0.081) 등 전통적 거시 변수보다도 높게 나타났다는 점에서, 이벤트 정보가 경제 펀더멘털 변수 이상으로 시장 심리에 작용함을 보여준다.

그 외에도 CPI, WTI Price, KOSDAQ 종가, M2, Copper 가격 등이 주요 설명 변수로 확인되었다. 이는 물가·유동성·원자재 가격이 복합적으로 작용해 환율에 영향을 미친다는 점을 뒷받침한다. 반면 한국 정책금리, M1, usd/jpy, us10y 등의 변수는 상대적으로 낮은 중요도를 보였는데, 이는 단기 환율 변동 예측에서는 국내 단독 요인보다 글로벌 및 심리적 요인의 설명력이 더 크다는 점을 시사한다.

텍스트, 스크린샷, 소프트웨어, 번호이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 2> Permutation Importance (Top 20)**

3.3 시각화

본 절에서는 최적 구성(Macro+Event, LSTM, lookback=20)을 이용하여 2020년부터 2024년까지의 원/달러 환율 예측 결과를 시각화하였다<그림 3>. 그래프는 실제 값(Actual)과 모델의 예측 값(Predicted)을 전체 구간에 걸쳐 비교하였으며, 학습·검증에 사용된 구간과 테스트 구간은 파란색 점선으로 구분하였다.

전체 시계열 비교 결과, 모델은 환율의 장기적 추세뿐만 아니라 중단기적 변동 패턴을 비교적 안정적으로 추적하는 모습을 보였다. 특히 코로나19 이후 글로벌 유동성 확대 국면(2020–2021), 미 연준의 급격한 긴축과 환율 상승(2022), 그리고 이후 조정 국면(2023–2024)에 걸쳐 방향성과 국면 전환을 대체로 잘 포착하였다. 이는 이벤트 정보가 단순 시계열 신호를 넘어, 국면 전환기에서 유의미한 보조 지표로 작용했음을 시사한다.

또한 <그림 4>는 테스트 구간을 확대하여 실제 값과 예측 값의 오차를 상세히 비교한 것이다. 전반적으로 실제 값의 국지적 급등락을 예측하기에는 한계가 있었으나, 평균 수준과 변동성의 큰 방향성은 안정적으로 재현하였다.

결론적으로, 제안된 최적 모델은 환율의 구조적 추세와 단기적 변동성을 모두 일정 수준 포착하며, 이벤트·뉴스 감성 정보를 추가할 경우 단순 거시변수만을 사용할 때보다 전반적 예측력이 향상됨을 시각적으로 확인할 수 있었다.

텍스트, 그래프, 스크린샷, 라인이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 3> 원/달러 환율 예측 결과**

그래프, 스크린샷, 라인, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 4> Test 구간 예측 결과**

Ⅳ. 결론

본 연구는 원/달러 환율 예측의 정확도 향상을 위해 거시·시장 변수에 이벤트(GDELT) 및 뉴스 감성 요약 지표를 결합하고, LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 딥러닝 아키텍처를 비교·분석하였다. 2020년부터 2024년까지의 일별 시계열 데이터를 기반으로, 세 가지 시드(42, 55, 68)를 적용한 반복 실험을 통해 평균 성능을 산출하였다. 모든 모델은 동일한 학습 조건(epoch=80, batch=64, Adam optimizer, MSE 손실함수) 하에 학습되었으며, 입력 시퀀스 길이(lookback)를 5, 10, 20, 30, 60, 90일로 변화시켜 시간창 길이에 따른 예측력 변화를 검증하였다.

실험 결과, LSTM 모델의 ‘Macro + Event’ 조합과 lookback=20일 설정이 전체 실험 중 가장 우수한 성능(RMSE=6.946, MAE=5.384, MAPE=0.394%)을 기록하였다. 이는 약 1개월 규모의 시계열 윈도우가 환율의 정보 효율성이 가장 높은 구간임을 시사하며, 과도하게 짧거나 긴 윈도우(5일 이하, 60일 이상)는 오히려 노이즈 축적으로 인해 성능이 저하됨을 보여준다. LSTM 모델은 순환 구조를 통해 시계열의 장기 의존성을 안정적으로 학습하면서도, 이벤트(GDELT) 변수의 시점별 변화를 효과적으로 반영하여 단기 변동성까지 포착한 것으로 해석된다.

한편, CNN-LSTM과 CNN-GRU 모델은 합성곱 계층을 통한 단기 패턴 인식에 강점을 보여 전체적으로 낮은 RMSE를 유지했으나, LSTM 단일 구조의 단순성과 안정성이 중기 예측 구간(20일)에서는 오히려 더 높은 예측 효율을 보였다. GRU 모델은 전반적으로 LSTM 대비 오차가 크고 분산이 높아, 시계열 길이가 짧을수록 예측 변동성이 확대되는 경향을 보였다.

변수 중요도 분석(Permutation Importance) 결과에서는 S&P500 종가(0.397), 경상수지(0.348), 다우존스 종가(0.281)가 최상위에 위치하여, 글로벌 주식시장과 대외수지 요인이 환율 변동에 가장 강한 영향을 미치는 것으로 나타났다. 또한 events\_tone\_mean(0.229)이 네 번째로 높은 중요도를 기록하며, 뉴스 기반 사건의 정서적 톤이 전통적 거시 변수(CPI, WTI 등)보다 높은 설명력을 보였다. 이는 시장의 ‘심리적 방향성’이 환율의 단기 변동을 결정하는 주요 신호로 작용할 수 있음을 의미한다. 반면, 개별 이벤트(GDELT) 변수들은 일부 영향력을 보였으나 그룹 단위의 평균 중요도는 낮게 나타나, 데이터의 노이즈 및 국면 의존성이 작용했을 가능성이 제기된다.

시각화 결과에서도 제안된 CNN-LSTM 모델이 2020–2024년 전 기간 동안 환율의 구조적 추세와 단기 변동성을 안정적으로 추적하는 것으로 확인되었다. 특히 코로나19 팬데믹, 미 연준의 금리 인상, 글로벌 공급망 충격 등 주요 거시 이벤트 구간에서 환율의 방향성과 변동 폭을 실질적으로 반영하였다. 이는 사건 및 감성 정보가 단기 방향성 및 시장 불확실성 예측에 있어 보조지표로서 유용함을 실증적으로 보여준다.

종합하면, 본 연구는 (1) 거시·시장 변수에 사건·감성 요인을 결합한 복합 입력 구조의 유효성, (2) 약 20영업일 규모의 최적 시간창, (3) CNN-LSTM 기반 하이브리드 구조의 예측 안정성과 해석 가능성을 실증적으로 제시하였다. 이러한 결과는 환율 예측을 위한 실무적·정책적 활용 측면에서 의미가 있다. 금융기관은 사건·감성 지표를 보조 피처로 활용함으로써 환위험 관리 및 단기 환헤지 전략의 정밀도를 높일 수 있으며, 정책 당국은 시장 심리 변동을 조기 감지하는 데이터 기반 정책 보조지표로 활용할 수 있을 것이다.

다만, GDELT 이벤트 데이터의 노이즈와 중복성, 그리고 비정상 구간(예: 지정학적 충격)에서의 예측 불안정성은 본 연구의 한계로 남는다. 향후 연구에서는 다양한 뉴스·SNS 기반 감성 데이터의 통합, 시점별 중요도 동적 가중화, Transformer 기반 시계열 모델과의 비교 등을 통해 사건·감성 정보의 활용 가능성을 더욱 확장할 수 있을 것으로 기대된다.

부록

*A. 변수 상관관계 히트맵*

<그림 5>는 변수 간 상관관계 히트맵을 제시한 것이다. 결과를 보면 환율과 주가, 금리, 원자재 가격 등 주요 거시·시장 지표들은 서로 강한 양(+)의 상관 혹은 음(–)의 상관을 보이며, 전형적인 금융 변수들 간 공분산 구조가 나타난다. 뉴스 감성 지표들 역시 서로 밀접한 상관관계를 형성하고 있음을 확인할 수 있다.

패턴, 스크린샷, 사각형, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 5> 변수 상관관계 히트맵**

# Data Availability

The data can be provided on request.

# Declaration of Competing Interest

The authors declare no conflict of interest.

# CRediT authorship contribution statement

**Jaeheung Park:** Conceptualization, Software, Validation, Data Curation, Writing Original Draft, Preparation.

**Kyungwon Kim:** Methodology, Preparation, Software, Writing Review and Editing, Visualization, Supervision.

# References

[1] 한국은행. *환율의 정의*. 한국은행. https://www.bok.or.kr/portal/main/contents.do?menuNo=200407

[2] 김인준, & 이영섭. (2019). *국제경제론*. 박영사.

[3] 한국경제. (2023, 10월 22일). [다산칼럼] 환율 예측이 늘 빗나가는 까닭. *한국경제신문*. https://www.hankyung.com/article/2023102265501

[4] 한국은행. *환율의 변동요인*. 한국은행. https://www.bok.or.kr/portal/main/contents.do?menuNo=200407

[5] Plakandaras, V., Papadimitriou, T., & Gogas, P. (2015). Forecasting daily and monthly exchange rates with machine learning techniques. *Journal of Forecasting*, *34*(7), 560-573.

[6] 임현욱, 정승환, 이희수, & 오경주. (2021). 국고채, 금리 스왑 그리고 통화 스왑 가격에 기반한 외환시장 환율예측 연구: 인공지능 활용의 실증적 증거. *지식경영연구*, *22*(4), 71-85.

[7] Cao, W., Zhu, W., Wang, W., Demazeau, Y., & Zhang, C. (2020). A deep coupled LSTM approach for USD/CNY exchange rate forecasting. *IEEE Intelligent Systems*, *35*(2), 43-53.

[8] Qureshi, A. M. (2025). ML FORECASTING OF EXCHANGE RATES MACHINE LEARNING PREDICTION OF EXCHANGE RATES USING MACROECONOMIC INDICATORS Analysis and Use of Various ML Techniques (SVM, Random Forest). *Available at SSRN 5287273*.

[9] Wang, J., Wang, X., Li, J., & Wang, H. (2021). A prediction model of CNN-TLSTM for USD/CNY exchange rate prediction. *Ieee Access*, *9*, 73346-73354.

[10] Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock price prediction using news sentiment analysis. 2019 IEEE fifth international conference on big data computing service and applications (BigDataService)

[11] Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019.

[12] 정가연, 이혁제, 이준영, & 이제혁. (2024). 금융 특화 감정분석 모델과 딥러닝 시계열 예측 모델을 활용한 코스피 지수 예측. *대한산업공학회지*, *50*(4), 240-250.

[13] Ding, H., Shi, X., Deng, R., Faroog, S., Dewi, D. A., Abdullah, S. N., & Malek, B. A. (2024). Eur/usd exchange rate forecasting incorporating text mining based on pre-trained language models and deep learning methods. *arXiv preprint arXiv:2411.07560*.

[14] 김우석, & 한규식. (2021). COVID-19 가 원달러환율에 미친 영향. *금융지식연구*, *19*(1), 33-58.

[15] 오인정, & 김우주. (2022). SARIMA 와 ARDL 모형을 활용한 COVID-19 구간별 원/달러 환율 예측. *지능정보연구*, *28*(4), 191-209.

[16] GDELT Project. The GDELT Project. https://www.gdeltproject.org

[17] Schroeder, C., Winterbottom, S., Sitter, R., & Perdigones, J. L. (2013). Big Data Analysis of Human Societal Events: Indications for Forecasting Currency Exchange Rates.

[18] Consoli, S., Tiozzo Pezzoli, L., & Tosetti, E. (2020). Information extraction from the GDELT database to analyse EU sovereign bond markets. Workshop on Mining Data for Financial Applications,

[19] Blanqué, P., Ben Slimane, M., Cherief, A., Le Guenedal, T., Sekine, T., & Stagnol, L. (2022). Monitoring Narratives: an Application to the Equity Market. *Amina and Le Guenedal, Théo and Sekine, Takaya and Stagnol, Lauren, Monitoring Narratives: An Application to the Equity Market (April 7, 2022)*.

[20] Gupta, S., Ranjan, R., & Singh, S. N. (2024). Comprehensive study on sentiment analysis: From rule-based to modern llm based system. *arXiv preprint arXiv:2409.09989*.

[21] Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2018). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, *32*(6), 74-80.

[22] Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

[23] Nasiopoulos, D. K., Roumeliotis, K. I., Sakas, D. P., Toudas, K., & Reklitis, P. (2025). Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models. *International Journal of Financial Studies*, *13*(2), 75.

[24] Shen, Y., & Zhang, P. K. (2024). Financial sentiment analysis on news and reports using large language models and finbert. 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS),

[25] Kang, J.-W., & Choi, S.-Y. (2025). Comparative investigation of gpt and finbert’s sentiment analysis performance in news across different sectors. *Electronics*, *14*(6), 1090.

[26] Barbella, M., & Tortora, G. (2022). Rouge metric evaluation for text summarization techniques. *Available at SSRN 4120317*.

[27] Zhang, M., Li, C., Wan, M., Zhang, X., & Zhao, Q. (2024). ROUGE-SEM: Better evaluation of summarization using ROUGE combined with semantics. *Expert Systems with Applications*, *237*, 121364.

[28] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[29] Zheng, W., & Chen, G. (2021). An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization. *IEEE Transactions on Cybernetics*, *52*(12), 13902-13914.

[30] Qu, Y., & Zhao, X. (2019). Application of LSTM neural network in forecasting foreign exchange price. Journal of Physics: Conference Series,

[31] Agustin, F., & De Melin, P. (2024). Comparison of GRU and CNN Methods for Predicting the Exchange Rate of Argentine Peso (ARS) against US Dollar (USD). *International Journal Artificial Intelligent and Informatics*, *2*(1), 9-16.

[32] Islam, M. S., & Hossain, E. (2021). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters*, *3*, 100009.

[33] Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN‐LSTM‐based model to forecast stock prices. *Complexity*, *2020*(1), 6622927.