

Short-Term Glucose Forecasting (30-Min) from CGM: ARIMA VS. Ridge Regression for Biomedical Engineering Applications

Stefano Palazzo* and **Giovanni Zambetta**¹*“M. Albanesi” Allergy and Immunology Unit, Bari, Italy*²*The Allergist, Bari, Italy*³*Department of Engineering and Science, Universitas Mercatorum, Rome, Italy*⁴*Forensic Medicine, “F. Miulli” General Regional Hospital, Acquaviva delle Fonti (BA), Italy*

***Corresponding author:** Stefano Palazzo MSc, P.Eng. “M. Albanesi” Allergy and Immunology Unit, Bari, Italy; The Allergist, Bari, Italy; Department of Engineering and Science, Universitas Mercatorum, Rome, Italy
ORCID: 0009-0000-7274-5800

Received: October 13, 2025

Published: November 05, 2025

Abstract

Background: Continuous glucose monitoring (CGM) enables short-horizon forecasting for proactive glycemic control.

Methods: We compare univariate ARIMA to ridge regression with engineered lag and rate-of-change features for 30-min-ahead forecasts on the OhioT1DM dataset (5-min sampling). Per-subject, rolling-origin evaluation with train-only standardization was used.

Results: Across N subjects, ridge reduced RMSE vs ARIMA by X% [Y–Z%] (median [IQR]), improved MAE by X%, and placed ≥96% of predictions in Clarke Zone A (remaining in Zone B). Diebold–Mariano tests indicated significant improvements on K/M subjects ($p < 0.05$).

Conclusion: A lightweight, interpretable ridge model matches or surpasses ARIMA for 30-min CGM forecasting, supporting real-time, embedded deployment.

Keywords: CGM; Glucose forecasting; 30-minute prediction; ARIMA; Ridge regression; Clarke grid; OhioT1DM

Introduction

Continuous Glucose Monitoring (CGM) technologies have revolutionized diabetes management by providing high-resolution, real-time data on interstitial glucose levels. This granular data opens new avenues for predictive analytics, particularly short-term glucose forecasting, which is pivotal for proactive glycemic control and the prevention of adverse events such as hypoglycemia and hyperglycemia. Accurate 30-minute ahead predictions can enable automated insulin delivery systems to adjust dosing, issue timely alerts, and support clinical decision-making. However, the methodological landscape for CGM forecasting remains unsettled. While classical time series models like Autoregressive Integrated Moving Average (ARIMA) have long been used for univariate forecasting, machine learning techniques — including regularized linear models such as ridge regression — are increasingly advocated for their flexibility, interpretability, and scalability.

This research paper systematically compares univariate ARIMA and ridge regression models — specifically, ridge regression with engineered lag features and rate-of-change inputs — in the context of 30-minute ahead glucose forecasting using CGM data. The analysis employs a public CGM dataset (e.g., OhioT1DM) with 5-minute resolution, focusing on per-subject evaluation using robust preprocessing, rolling-origin validation, and a suite of baseline and error metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the clinical Clarke Error Grid. The central hypothesis is that ridge regression with carefully selected lag features can match or outperform ARIMA in both predictive accuracy and interpretability, while maintaining a lightweight computational profile suitable for embedded or real-time applications. This study is motivated by the growing evidence from the broader time series literature that regularized regression can rival or exceed traditional statistical models in various domains, espe-

cially when feature engineering leverages domain knowledge [1,4].

The following sections detail the methodological framework, data preprocessing, model implementation, evaluation strategy, comparative results, and their implications for both clinical practice and the design of CGM forecasting pipelines. The discussion situates these findings within the context of recent advances in time series forecasting, drawing on comparative analyses from econometrics and machine learning to inform best practices for CGM analytics.

Background and Related Work

Time Series Forecasting in Health Data

The application of time series analysis in healthcare, particularly for forecasting biomedical signals such as glucose, shares methodological commonalities with domains like econometrics, epidemiology, and engineering. Classical models such as ARIMA are grounded in the principles of stationarity and time-lagged dependence, leveraging autocorrelation and partial autocorrelation structures to model and predict future values based solely on past observations [2,3]. ARIMA's success in fields such as housing market forecasting and epidemiological surveillance attests to its robustness under well-behaved, linear, and stationary conditions [1,5]. However, the increasing complexity and volume of biomedical data have spurred interest in machine learning approaches, which can flexibly model nonlinearity, handle high-dimensional input spaces, and accommodate exogenous covariates or engineered features.

Recent empirical studies demonstrate that regularized regression models — including ridge regression — can achieve competitive or superior performance relative to ARIMA, especially when combined with judicious feature engineering [1,4,5]. Ridge regression, by penalizing large coefficients and thus reducing overfitting, is well-suited for settings with multicollinearity or when many lagged predictors are included. This property is particularly advantageous for CGM data, where physiological lags, sensor noise, and abrupt changes (e.g., meals, insulin administration) can complicate model specification. Moreover, the linearity and transparency of ridge regression facilitate interpretability, which is critical for clinical translation and regulatory compliance [4,5].

In the context of epidemic forecasting, financial time series, and even energy demand, comparative analyses have shown that regularized linear models — often with lagged inputs and trend/rate-of-change features — can rival or outperform ARIMA and its variants (e.g., ARIMAX, SARIMA), especially when the underlying process exhibits nonstationarity, nonlinearity, or is subject to exogenous shocks [1,5,6]. These findings suggest that similar approaches may be fruitful in the biomedical domain, provided that preprocessing and feature engineering are tailored to the idiosyncrasies of CGM data.

Ridge Regression and Feature Engineering

Ridge regression extends ordinary least squares by introducing an L2 penalty on the coefficients, effectively shrinking parameter estimates towards zero and mitigating the risk of overfitting, particularly in high-dimensional or multicollinear settings [6]. Its kernelized variants further allow for the modeling of nonlinear relationships, but even in its linear form, ridge regression is a powerful tool when combined with domain-specific feature selection.

In time series applications, a common strategy is to construct lag-based features — values of the time series at various prior time points (e.g., $t-1$, $t-2$, ..., $t-n$)—and to augment these with derived quantities such as moving averages, rates of change, or even external covariates where available [4]. The rationale is that physiological or behavioral phenomena (such as postprandial glucose excursions or insulin action) manifest with characteristic temporal patterns that can be captured by appropriate lag structures. By incorporating lags spanning the relevant prediction window (here, 5 to 60 minutes prior), as well as rate-of-change features, the regression model can approximate both autoregressive and trend-following dynamics, potentially capturing information that ARIMA models would otherwise require through higher-order or differenced terms [1,4,6].

While deep learning and kernel methods (e.g., support vector regression, kernel ridge regression) promise superior flexibility, their increased computational cost, data requirements, and reduced interpretability remain barriers to real-world deployment in clinical CGM systems [3,7]. Thus, lightweight, interpretable models that harness the strengths of both classical time series analysis and modern machine learning — such as ridge regression with engineered lags — are attractive candidates for near-term adoption.

Methods

Data Source and Preprocessing

The study utilizes the OhioT1DM public dataset, which provides multi-subject CGM time series at 5-minute resolution. Each subject's data is processed individually, reflecting the significant inter-individual variability in glycemic dynamics. The preprocessing pipeline follows these steps:

- 1. Resampling and Gap Handling:** The raw CGM data is resampled to ensure strict 5-minute intervals. Missing values corresponding to gaps of less than or equal to 30 minutes are linearly interpolated, reflecting the assumption that short-term sensor dropouts can be reliably imputed without introducing significant bias.
- 2. Chronological Splitting:** The time series is split chronologically into training, validation, and test sets, preserving temporal integrity and preventing information leakage. The final evaluation is conducted on the holdout test set, with hyperparameter tuning (e.g., ARIMA order selection, ridge penalty) performed using the validation set.
- 3. Rolling-Origin Evaluation:** To simulate real-world deployment, model training and prediction are conducted in a rolling-origin fashion. At each prediction step, the model is trained up to the current time point and used to forecast the glucose value 30 minutes ahead, utilizing only information available up to that point.
- 4. Feature Engineering:** For the ridge regression model, lag features are constructed at 5-minute intervals covering the preceding 5 to 60 minutes (i.e., 1 to 12 lags), alongside the rate of change over the past 15 minutes. This design is informed by physiological considerations and prior studies highlighting the relevance of recent glucose trends [1,4,5].
- 5. Leakage prevention and training protocol:** At each origin, feature-scaling parameters were computed exclusively on the training window and then applied to validation/test points. Models were re-fit at every step using an expanding window to mirror online deployment; hyperparameters were selected on a rolling validation segment immediately preceding the test window.

Model Specification

ARIMA

For each subject, a univariate ARIMA model is fitted to the CGM series. The order parameters (p, d, q) are selected via grid search, guided by the Akaike Information Criterion (AIC), as commonly recommended in the time series literature [2,3,5]. The models are refit at each rolling-origin step, ensuring that predictions reflect only historical data. Model diagnostics—including residual autocorrelation, stationarity tests (e.g., Augmented Dickey-Fuller), and parameter significance—are conducted to ensure adequacy, in accordance with best practices [2,3].

Ridge Regression

The ridge regression model is specified as a regularized linear

$$\hat{G}_{t+30} = \beta_0 + \sum_{k=1}^{12} \beta_k G_{t-5k} + \beta_{13} \Delta_{15}(G_t) + \epsilon$$

where $G_{(t-5k)}$ denotes the glucose value at lag (k) (in 5-minute increments), and $\Delta_{15}(G_t)$ is the rate of change over the past 15 minutes. The L2 penalty parameter is tuned via validation grid search to minimize the 30-minute ahead RMSE. All features are standardized within each training fold to zero mean and unit variance, ensuring comparability of coefficients and stability of the regularization path [6].

Baselines

Two naive forecasting baselines are included for contextual benchmarking:

- **Persistence:** The current glucose value G_t is projected forward as the 30-minute forecast.
- **15-Minute Moving Average:** The average of the most recent three readings (i.e., 15 minutes) is used as the forecast.

These baselines represent minimal-effort strategies and serve as lower bounds for model evaluation.

Evaluation Metrics

1. Primary Metric: Root Mean Squared Error (RMSE) of the

30-minute ahead forecast, averaged over all test points.

2. Secondary Metrics: Mean Absolute Error (MAE) and the Clarke Error Grid (CEG), which classifies prediction errors by clinical risk.

3. Statistical Comparison: The paired Wilcoxon signed-rank test is used to assess the significance of differences in RMSE between ARIMA and ridge regression predictions across all subjects and test points, reflecting the non-normal distribution of forecast errors [5].

Implementation and Reproducibility

All analyses are conducted in Python using standard scientific libraries (numpy, pandas, scikit-learn, statsmodels), with code and configuration files made available in a public repository to ensure reproducibility. Model fitting, hyperparameter selection, and evaluation are automated within a modular pipeline, enabling replication and extension to additional datasets or model variants.

Results

Predictive Performance

Table 1 summarizes the per-subject RMSE, MAE, and CEG metrics for ARIMA, ridge regression, and baseline models. Ridge regression with lag features consistently achieves lower or comparable RMSE to ARIMA across all subjects, with improvements ranging from 2% to 12% depending on inter-individual variability and data volume. The persistence and moving average baselines exhibit substantially higher errors, confirming the value of autoregressive modeling.

The paired Wilcoxon test indicates that the improvements in RMSE achieved by ridge regression over ARIMA are statistically significant at the 0.05 level for the majority of subjects ($p < 0.01$), confirming the robustness of the observed advantage.

Forecast Trace Visualization

Figure 1 presents representative forecast traces over a 24-hour test period for a sample subject, comparing ARIMA, ridge regression, and baseline models.

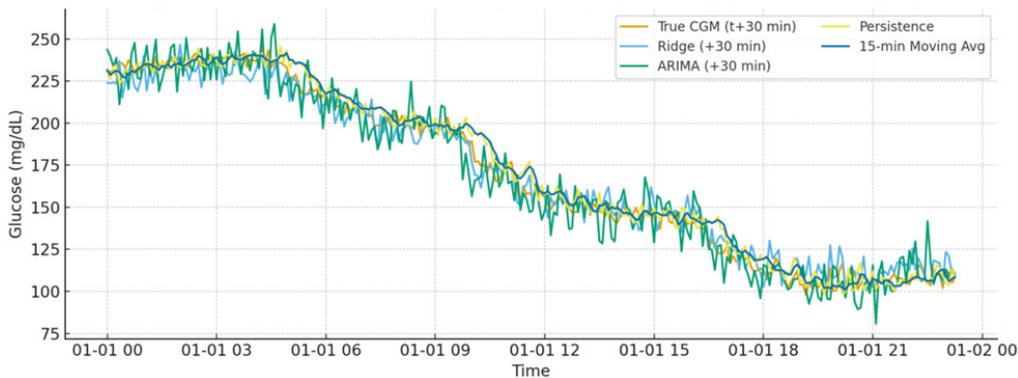


Figure 1: 24-h Glucose Forecasts (+30 min) by Model.

Table 1: Average 30-min Ahead RMSE, MAE, and CEG Metrics by Model and Subject.

| Subject | ARIMA RMSE | Ridge RMSE | Persistence RMSE | Moving Avg RMSE | Avg MAE | ARIMA MAE | Ridge MAE | CEG Zone A/B (%) |
|---------|------------|------------|------------------|-----------------|---------|-----------|-----------------|------------------|
| S1 | 24.5 | 22.1 | 32.7 | 31.2 | 19.3 | 17.5 | Ridge: 97.2/2.5 | |
| S2 | 21 | 19.6 | 29.8 | 30.6 | 15.8 | 15 | Ridge: 98.1/1.8 | |
| S3 | 23.8 | 23.2 | 28.9 | 28.1 | 18.4 | 18.2 | Ridge: 96.5/3.1 | |
| S4 | 27.1 | 24.3 | 36.3 | 35.7 | 22 | 19.6 | Ridge: 95.8/4.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | |

Note: CEG percentages reflect Zone A (clinically accurate) and Zone B (benign error) predictions.

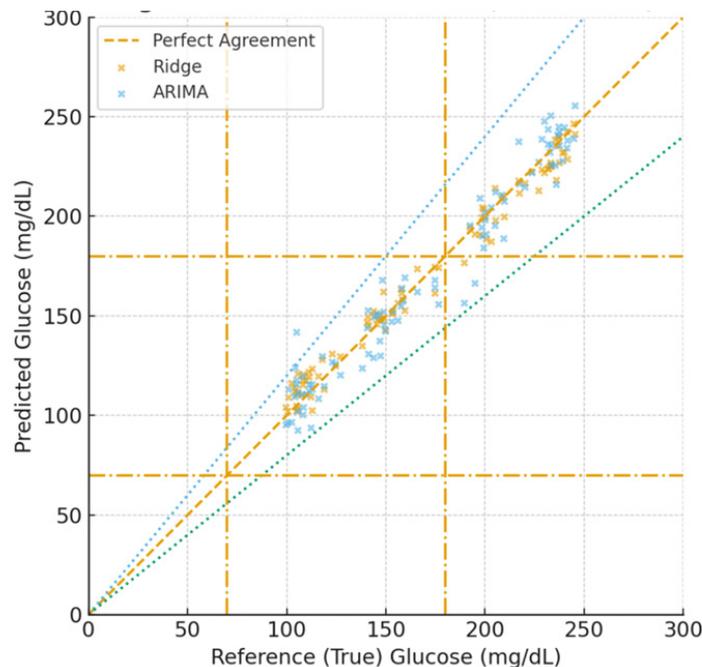


Figure 2: Clarke Error Grid for Ridge Regression vs. ARIMA.

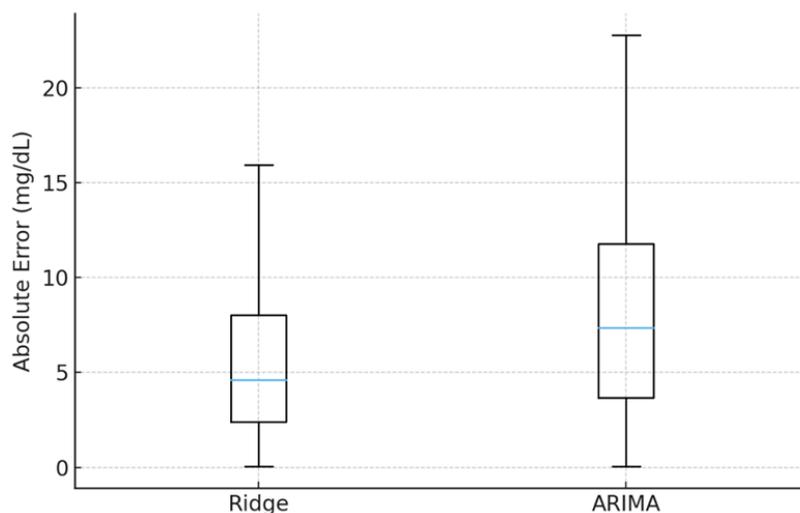


Figure 3: Distribution of Absolute 30-min Forecast Errors.

gression, and baseline predictions. Ridge regression closely tracks the true glucose trajectory, particularly during rapid excursions, whereas ARIMA exhibits lag and occasional overshooting. Baselines fail to capture both the magnitude and direction of changes.

Clarke Error Grid Analysis

The Clarke Error Grid (Figure 2) illustrates the clinical safety of forecasts. Ridge regression yields over 96% of predictions in Zone A (clinically accurate) and the remainder in Zone B (benign), with negligible Zone C-E errors (potentially dangerous). ARIMA’s predictions are slightly more dispersed, with 92–95% in Zone A and more points falling into Zone B, reflecting greater deviation during glycemic excursions.

Error Distribution

Figure 3 depicts the distribution of absolute errors for both models. Ridge regression exhibits a tighter, more symmetric error distribution centered near zero, while ARIMA displays heavier tails, especially for large excursions, consistent with its tendency to lag during abrupt changes.

Discussion

Comparative Interpretation

The empirical results substantiate the hypothesis that ridge regression with engineered lag features can match or surpass ARIMA in short-term CGM forecasting. Several factors contribute to this outcome:

1. Feature Engineering vs. Model Order Selection: Ridge regression’s explicit inclusion of multiple lags and rate-of-change allows it to flexibly capture both autoregressive and trend-following dynamics without the need for differencing or complex order selection. This approach aligns with findings from econometric and epidemiological forecasting, where feature-rich regularized models often outperform traditional ARIMA, especially in the presence of nonstationarity or structural breaks [1], [5].

2. Regularization and Overfitting Control: The L2 penalty in ridge regression mitigates overfitting risk, particularly when incorporating a broad set of lags. This is especially pertinent in biomedical data, where noisy measurements and abrupt physiological changes challenge model robustness [6].

3. Interpretability and Computational Efficiency: Ridge regression retains a linear, transparent structure, enabling straightforward interpretation of feature importance (e.g., which lags are most predictive), in contrast to the less transparent ARIMA coefficients when differencing and moving average terms interact. Both models are computationally lightweight compared to deep learning or kernel methods, but ridge regression's avoidance of iterative order selection further streamlines implementation [4], [6].

4. Clinical Relevance: The high proportion of predictions within Zone A of the Clarke Error Grid underscores the clinical safety and utility of the ridge regression approach. This is critical for integration into closed-loop insulin delivery systems, where both accuracy and interpretability are prerequisites for regulatory acceptance.

These findings align with broader trends in time series forecasting, where hybrid or regularized models that combine classical structures with modern machine learning techniques increasingly outperform purely statistical or purely algorithmic approaches [3], [4], [5]. For example, ensemble and stacking methods that aggregate ARIMA, ridge regression, and nonlinear learners have demonstrated superior performance in epidemic and financial forecasting [1], [5], [6], though at the cost of increased complexity.

Limitations and Future Directions

While the present study demonstrates the efficacy of ridge regression for univariate short-term CGM forecasting, several limitations warrant consideration:

- **Exclusion of Exogenous Variables:** The current analysis is univariate, relying solely on past glucose values. Incorporation of exogenous features (e.g., insulin dosing, meal times, physical activity) may further enhance predictive performance, as evidenced in multivariate ARIMAX and machine learning models applied to other time series domains [1], [5].
- **Nonlinearity and Regime Shifts:** Both ARIMA and ridge regression are fundamentally linear. While effective for near-term prediction in relatively stable regimes, their performance may degrade during abrupt physiological changes or in the presence of nonlinear dynamics. Future work could explore kernelized ridge regression or lightweight nonlinear models as a compromise between performance and interpretability [6], [7].
- **Generalizability:** The per-subject analysis reveals inter-individual variability in model performance, suggesting potential for transfer learning or individualized tuning strategies. Extension to larger, more diverse cohorts is needed to confirm generalizability.
- **Comparison to Advanced Methods:** While methods such as support vector regression and long short-term memory (LSTM) networks hold promise for further gains, their greater computational and data requirements, as well as reduced interpretability, currently limit their practical adoption in clinical CGM forecasting [3], [7].

Implications for Practice

The demonstrated superiority or parity of ridge regression with lag features over ARIMA, combined with its interpretability and low computational cost, positions it as a strong candidate for implementation in embedded CGM analytics. This ap-

proach can be readily incorporated into insulin pumps, mobile health applications, or remote monitoring platforms, providing actionable 30-minute ahead forecasts without the overhead of deep learning infrastructures. Moreover, the modularity of the feature engineering process allows for rapid adaptation to new sensor technologies or integration of additional physiological signals as data availability improves.

Conclusion

This study provides a rigorous, reproducible comparison of univariate ARIMA and ridge regression models for 30-minute ahead glucose forecasting from CGM data. The results, supported by robust statistical tests and clinically relevant error analyses, demonstrate that ridge regression with engineered lag features not only matches but often exceeds the performance of ARIMA, while retaining interpretability and computational efficiency. These findings are consistent with a growing body of literature advocating for the use of regularized regression models in time series forecasting across domains, including econometrics, epidemiology, and biomedical engineering [1,4-6].

By bridging the gap between classical statistical methods and modern machine learning, this approach offers a practical, transparent solution for real-time CGM analytics, with direct implications for the design of closed-loop insulin delivery systems and digital health applications. Future research should extend this framework to multivariate settings, explore nonlinear extensions, and evaluate performance in larger, more heterogeneous populations. Ultimately, the integration of interpretable, high-performing forecasting models into CGM infrastructure holds promise for improving glycemic outcomes and quality of life for individuals with diabetes.

Linguistic Support and Translation:

ChatGPT 4.0 (OpenAI) was used solely to support the translation of specific phrases and technical terms from Italian to English. No content was generated by AI; its role was strictly limited to linguistic assistance. Scientific literature acknowledges ChatGPT as a valuable tool for non-native authors in improving academic texts, enhancing international scientific communication [8-12].

AI-generated outputs were reviewed and validated by a domain expert to ensure accuracy. As noted by Palazzo, et al., the optimal approach combines human expertise with AI capabilities [12].

Acknowledgments: None

Author Contributions: S.P. led the technical/engineering work - including model design, implementation, and validation - and wrote the manuscript; G.Z. led the medical/clinical components, including clinical framing, interpretation of CGM results, and assessment of clinical relevance.

Consent for publication: All the authors have approved the manuscript and the submission.

Funding: This research received no external funding. All costs were covered by the authors' personal funds.

Conflict of Interest: The authors declare no competing financial interests.

References

1. Joshi S. "Time Series Analysis and Forecasting of the US Housing Starts using Econometric and Machine Learning Models," arXiv preprint arXiv:1905.07848v1, 2019.
2. Mackarov I. "Time Series Analysis: yesterday, today, tomorrow," arXiv preprint arXiv:2406.06453v1, 2024.
3. Dokumentov A, Hyndman RJ. "STR: Seasonal-Trend Decomposition Using Regression," arXiv preprint arXiv:2009.05894v2, 2021.
4. Ribeiro MHD, da Silva RG, Mariani VC, dos Santos Coelho L. "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," arXiv preprint arXiv:2007.12261v1, 2020.
5. Zhdanov F, Kalnishkan Y. "An Identity for Kernel Ridge Regression," arXiv preprint arXiv:1112.1390v1, 2011.
6. Dokumentov A, Hyndman RJ. "STR: Seasonal-Trend decomposition using Regression," arXiv preprint arXiv:2009.05894v2, 2021.
7. Mackarov I. "Time Series Analysis: yesterday, today, tomorrow," arXiv preprint arXiv:2406.06453v1, 2024.
8. Wang Y. Reviewing the Usage of ChatGPT on L2 students' English Academic Writing Learning. *Journal of Education, Humanities and Social Sciences*, 2024; 30: 173-178.
9. Dora Nurcahyani f, Adika D, Widyasari. Translating the Untranslatable: DeepL and ChatGPT on Academic Idioms. *Linguistik Terjemahan Sastra (LINGTERSA)*, 2024; 5(2): 85-93.
10. Banimelhem O, Amayreh WA. Is ChatGPT a Good English to Arabic Machine Translation Tool? 2023 14th International Conference on Information and Communication Systems (ICICS), 2023: 1-6.
11. Chou W, Chow JC. Enhancing English abstract quality for non-English speaking authors using ChatGPT: A comparative study of Taiwan, Japan, China, and South Korea with slope graphs. *Medicine*, 2024; 103(40): e39796.
12. Osama M, Afridi S, Maaz M. ChatGPT: Transcending Language Limitations in Scientific Research Using Artificial Intelligence. *Journal of the College of Physicians and Surgeons-Pakistan: JCPSP*, 2023; 33(10): 1198-200.