International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 5, Number 3, September 2026, *Page. 3076 - 3081*
Email : editorijhess@gmail.com

Prediction Of Daily Patient Visits At A Community Health Center Using Long Short-Term Memory (LSTM) With Laplace Differential Privacy

Azzahra[1)]*, **Sudin Saepudin**[2)], **Gina Syabani Yuda**[3)]
[1,2,3)] Department of Information Systems, Faculty of Engineering, Computers and Design, Nusa Putra University, Sukabumi, Indonesi

*Corresponding Author
E-mail: azzahra_si22@nusaputra.ac.id

Abstract

This study aims to analyze the performance of Long Short-Term Memory (LSTM) in predicting daily patient visits using limited-scale Electronic Medical Record (EMR) data and to evaluate the effect of applying Laplace Differential Privacy on prediction accuracy. The data were obtained from the EMR of Puskesmas in Sukabumi Regency, covering the period from January to December 2025. The data were preprocessed through cleaning, aggregation of daily patient visits, exclusion of holidays, and the construction of 14 days sequences. The LSTM model was evaluated using walk-forward validation with an expanding-window scheme consisting of five folds, while privacy protection was implemented using the Laplace mechanism with ε *values of* $0.5, 1.0, 2.0,$ *and* 5.0*. The results showed that the baseline model achieved an RMSE of 25.08 and an MAE of 22.19. The application of Differential Privacy increased prediction errors, with higher errors levels observed at smaller* ε *values. Increasing* ε *from 1.0 to 2.0 resulted in the largest reduction in prediction error, while further increas provided relatively small improvements. Based on these results,* $\varepsilon = 2.0$ *provided the most acceptable balance between privacy protection and prediction performance for the dataset used in this study.*

Keywords: LSTM, Patient Visit Prediction, Differential Privacy, Laplace, Electronic Medical Records

INTRODUCTION

Public health centers as primary health service facilities bear a significant operational burden and are required to provide fair, equitable, and high-quality services according to the needs of the community (Saputra et al., 2024). Unpredictable fluctuations in daily patient visits often lead to operational inefficiencies and have a direct impact on service quality (Daroedono et al., 2025) so the ability to accurately predict daily visits is an important aspect in planning medical personnel and drug availability, and proactive resource allocation (Ab Kader et al., 2023) Digital transformation through the implementation of Electronic Medical Records (EMR), provides opportunities to utilize historical data for more precise visit prediction (Aisyah et al., 2025). However, medical records contain sensitive patient information. This vulnerability is reflected in the alleged sale of COVID-19 patient data on RaidForums (Panggabean & Fitria, 2025) and the alleged leakage of BPJS Kesehatan data involving hundreds of millions of records (Yuridis et al., 2021), which contributed to the enactment of law No. 27 of 2022 on Personal Data Protection and Minister of Health Regulation No. 24 of 2022, requiring healthcare facilities to maintain patient confidentiality. This situation creates a challenge in balancing the need to develop artificial intelligence models that require detailed historical data with the obligation to protect privacy (Blanco-Justicia et al., 2023), while primary healthcare facilities are still expected to provide optimal services (Dinas Kesehatan Kabupaten Sukabumi, 2024).

This condition is evident in a primary healthcare facility in Sukabumi Regency, where daily patient visits in 2025 fluctuated between 1 and 168 visits of days.

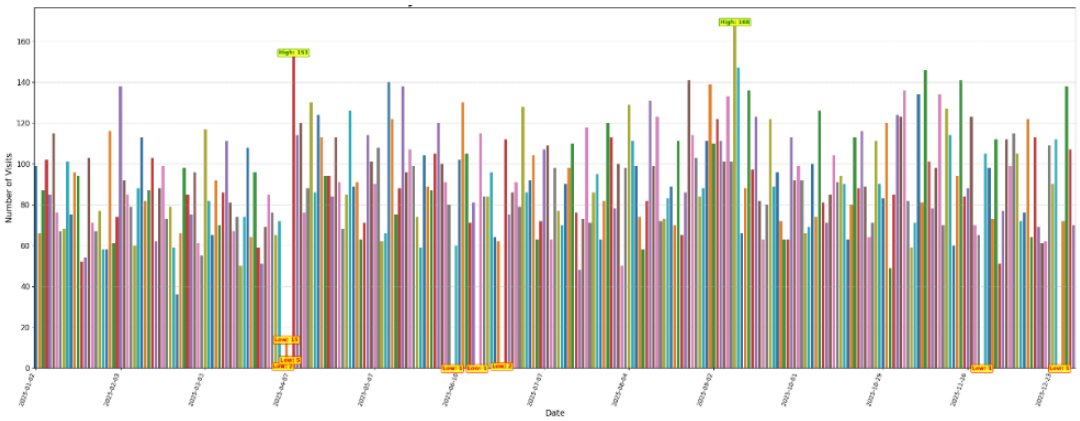

Figure 1. Frequency of Daily Patient Visits at Puskesmas Serving as the Case Study Site in 2025

Limited data access due to privacy considerations poses a challenge to the development of deep learning models (Alzubaidi et al., 2023). Although time-series forecasting methods have been widely used to predict patient visits (Bai et al., 2023) and Differential Privacy has been recognized as an approach for protecting medical data (Liu et al., 2023), their integration with limited-scale medical record data in primary healthcare facilities remains rarely implemented. Based on this condition, this study integrates Long Short-Term Memory (LSTM) with Laplace Differential Privacy (DP). LSTM was selected for its ability to capture nonlinear relationships in fluctuating healthcare data, with better performance than traditional statistical methods (Zhao et al., 2022), while laplace DP provides privacy protection through the addition of controlled mathematical noise. This approach is used to analyze prediction performance on a limited-scale dataset while considering patient data privacy (Mohammadi et al., 2026).

This study aims to analyze the performance of the LSTM model in predicting daily patient visits using limited medical record data and to evaluate the impact of Differential Privacy on prediction accuracy and stability. The novelty of this study lies in integrating LSTM with Laplace DP for limited-scale patient visit data and analyzing the effect of varying epsilon ($\boldsymbol{\varepsilon}$) values on the balance between privacy and model accuracy. The contribution of this study is a framework for evaluating patient visit prediction performance on limited-scale data while considering privacy protection, with the potential to support more efficient resource planning in primary healthcare facilities while ensuring compliance with data protection regulations.

RESEARCH METHODS

This study evaluates the performance of a time-series forecasting model for daily patient visits through data-driven modeling and simulation. Long Short-Term Memory (LSTM) is used as the baseline forecasting model and subsequently integrated with the Laplace Differential Privacy mechanism to analyze the effect of varying privacy parameter ε values on the balance between privacy protection and prediction accuracy in limited-scale medical record data. The data were obtained from the Electronic Medical Records (EMR) at a primary healthcare facility in Sukabumi Regency, covering the period from January to December 2025 with a total of 26.070 individual patient visit records. The experiments were conducted using Python on Google Colab, with TensorFlow used for model development and Scikit-learn for normalization and time-series data splitting. Data preprocessing included data cleaning and format stand duplicate removal, aggregation of individual visit records into daily visit counts, and restriction to normal operational days by excluding national holidays and collective leave days. The data were then transformed into sequences using a 14-day window size to represent recurring weekly patterns in patient visits.

Data splitting was performed using walk-forward validation with an expanding window scheme consisting of 5 folds (TimeSeriesSplit, n_splits=5, test_size=10), with all previously observed historical data used as training data in each fold. Normalization was performed using MinMax Scaling within the range [0, 1], with the normalization parameters calculated from the training data in each fold to prevent data leakage into the test data. The resulting predictions were then transformed back to the original scale using inverse transformation before evaluation. The LSTM model was used as the baseline because of its ability to capture temporal dependencies and nonlinear patterns in time-series data (Lindemann et al., 2021), as well as its competitive performance on limited datasets (Shrestha & Pradhanang, 2023). The model was developed using TensorFlow of 0.2, and a Dense output layer with one neuron. The Adam optimizer with a learning rate of 0.001 and Mean Squared Error loss function was used, with a maximum of 100 epochs, a batch size of 8, and EarlyStopping with a patience of 10 and restore_best_weights=True. This model served as the baseline prior to the application of the privacy mechanism.

Laplace Differential Privacy was applied using an output perturbation approach, namely by adding noise to the model predictions, with the following mechanism:

$$L(x) = f(x) + Laplace\left(0, \frac{\Delta f}{\epsilon}\right)$$

Where f(x) denotes the model output before noise addition, Δf denotes the function sensitivity, and ε denotes the privacy parameter. The experiments were conducted using four ε values, namely 0.5, 1.0, 2.0, and 5.0, representing relatively high to low levels of privacy protection. Evaluation was performed using 5 fold walk-forward validation with the same architecture as the baseline model, along with the addition of ReduceLROnPlateau. Model performance was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Evaluation was performed for each fold, and the mean values were then used to compare the performance of the baseline model with that of the model incorporating Differential Privacy.

RESULTS AND DISCUSSION

After preprocessing, the dataset consisted of 288 operational days which were subsequently transformed into 274 sequences with an input length of 14 days.

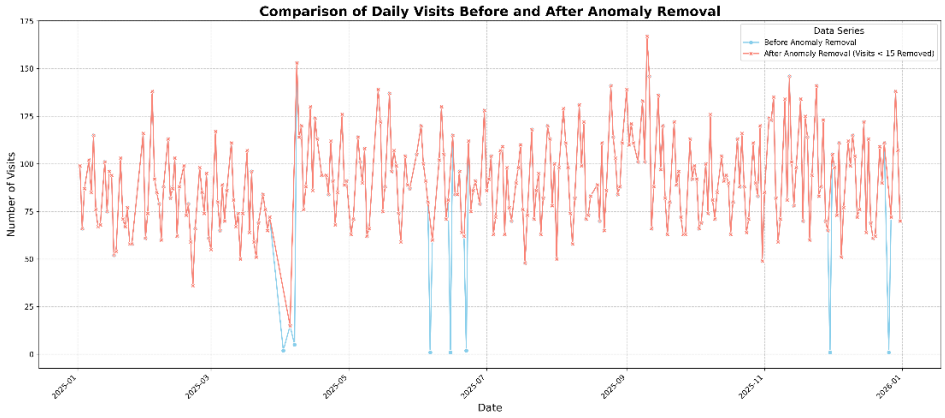

Figure 2. Comparison of Daily Patient Visits Before and After Anomaly Removal

Data splitting using walk-forward validation with an expanding window scheme resulted in five folds while preserving the chronological order of the data. Evaluation of the LSTM baseline model showed that RMSE decreased from the first to the third fold, then slightly increased in the fourth fold and increased again in the fifth fold. RMSE values were higher than MAE across all folds, indicating the presence of some predictions with relatively large errors. Overall, the model tended to produce predictions close to the mean values and was less responsive to extreme fluctuations. This is consistent

with findings that LSTM can provide competitive performance on limited datasets, although its ability to capture extreme fluctuations may be affected by data limitations (Shrestha & Pradhanang, 2023).

Table 1. Baseline Model Evaluation Results

Fold	RMSE	MAE
1	**30.53**	**29.14**
2	**26.12**	**21.75**
3	**20.52**	**17.65**
4	**20.83**	**19.11**
5	**25.97**	**23.29**
Average	**25.08**	**22.19**

The application of Laplace DP showed that the noise level affected prediction performance. Smaller ε values resulted in higher prediction errors, whereas increasing ε generally reduced RMSE and MAE. The standard deviations also tended to be lower at higher ε values, particularly from ε=2.0 onward, indicating more consistent performance across folds. This pattern suggests that stronger privacy protection through greater noise addition reduces prediction utility.

Table 2. Evaluation Results of LSTM Model with Laplace DP

Epsilon	RMSE ($\pm$ std)	MAE ($\pm$ std)
0.5	**41.85** $\pm$ 3.67	**35.62** $\pm$3.85
1.0	**39.47** $\pm$4.78	**33.53** $\pm$3.41
2.0	**27.28** $\pm$1.57	**23.83** $\pm$1.15
5.0	**26.46** $\pm$0.87	**23.35** $\pm$0.95

The relationship between ε and prediction performance is presented in Figure 3. The largest reduction in error occurred as ε increased from 1.0 to 2.0, whereas increasing ε from 2.0 to 5.0 resulted in only a relatively small reduction in error. Considering that a larger ε provides weaker privacy protection, $\varepsilon = 2.0$ offers a better trade-off between privacy and accuracy for the dataset used in this study. This selection is context-dependent and does not indicate that $\varepsilon = 2.0$ is universally optimal value.

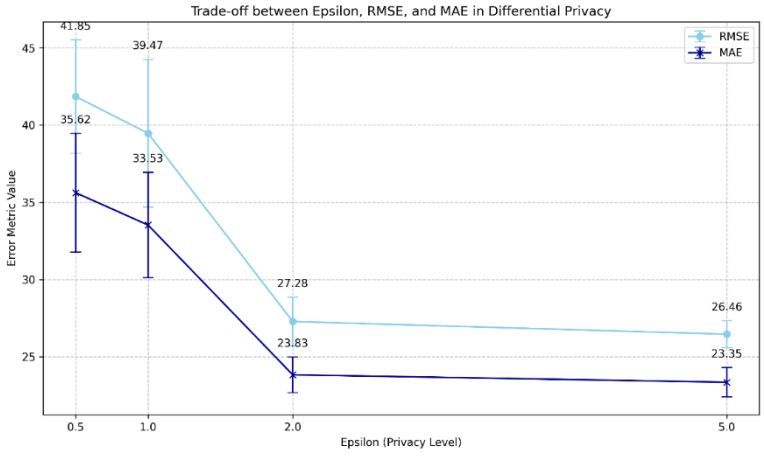

Figure 3. Trade-off Between ε, RMSE, and MAE

The prediction visualization in figure 4 and figure 5 shows that the baseline model tended to follow the mean values and was less responsive to sharp fluctuations. At smaller ε values, the prediction showed greater deviations from the actual values, whereas increasing ε resulted in patterns that increasingly approached the baseline. Nevertheless, all scenarios showed limitations in capturing extreme peaks and thoughts.

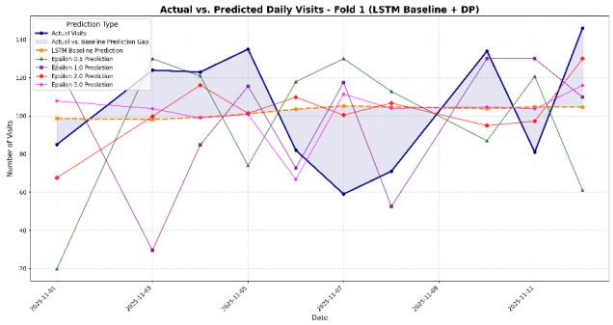

Figure 4. Comparison of Actual Values and Predicted Values in Fold 1

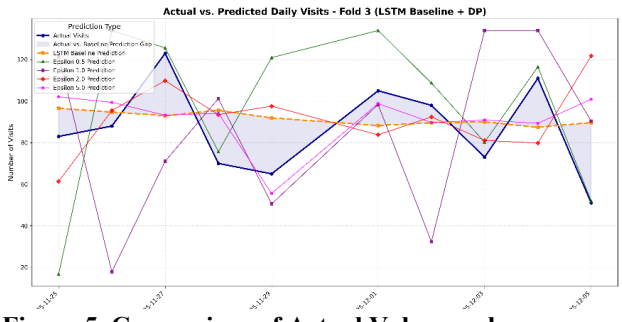

Figure 5. Comparison of Actual Values and Predicted Values in Fold 3

Overall, the prediction visualizations show that the baseline model tended to produce predictions close to the mean and was less responsive to sharp daily fluctuations. Models with smaller ε values showed greater deviations from the actual values, whereas larger ε values produced prediction patterns closer to the baseline with more controlled variability. However, the models remained limited in capturing extreme peaks and troughs, indicating that the observed privacy-accuracy trade-off was also influenced by the variability and limited size of the dataset.

CONCLUSION

The LSTM model was able to capture the general pattern of daily patient visits in the limited-scale electronic medical record dataset, although it remained less responsive to extreme fluctuations. These results indicate that medical record data available at the Puskesmas can serve as a basis for predicting patient visits and understanding general visit patterns. The application of Laplace Differential Privacy to the LSTM model was feasible for the limited-scale medical record dataset but resulted in higher prediction errors than the baseline model. Stronger privacy protection led to greater degradation in prediction performance, indicating a trade-off between privacy and predictive accuracy. The most substantial reduction in prediction error occurred as ε increased from 1.0 to 2.0, while further increasing ε provided only a relatively small improvement in accuracy. Therefore, $\varepsilon = 2.0$ provided the most acceptable balance between privacy protection and prediction performance for the dataset used in this study.

REFERENCES

Ab Kader, N. I., Yusof, U. K., & Khalid, M. N. A. (2023). Forecasting Inpatient and Outpatient Visits for Depressive Disorders: A Comparative Study of Deep Learning Approaches. *Applications of Modelling and Simulation*, *7*(Vol. 7 No. 2), 168–177. https://doi.org/10.31224/osf.io/9p7m5

Aisyah, D. N., Setiawan, A. H., Mayadewi, C. A., Lokopessy, A. F., Kozlakidis, Z., & Manikam, L. (2025). Understanding Health Information Systems Utilization Across Public Health Centers in Indonesia: Cross-Sectional Study. *JMIR Medical Informatics*, *13*, 1–15. https://doi.org/10.2196/68613

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A Survey on Deep Learning Tools Dealing

with Data Scarcity: Definitions, Challenges, Solutions, Tips, and Applications. *Journal of Big Data*, *10*(1), 1–82. https://doi.org/10.1186/s40537-023-00727-2

Bai, L., Lu, K., Dong, Y., Wang, X., Gong, Y., Xia, Y., Wang, X., Chen, L., Yan, S., Tang, Z., & Li, C. (2023). Predicting monthly hospital outpatient visits based on meteorological environmental factors using the ARIMA model. *Scientific Reports*, *13*, 2691. https://doi.org/10.1038/s41598-023-29897-y

Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., & Muralidhar, K. (2023). A Critical Review on the Use (and Misuse) of Differential Privacy in Machine Learning. *ACM Computing Surveys*, *55*(8). https://doi.org/10.1145/3547139

Daroedono, E., Aurellita, N., Savero, I., Sirait, G. P., Imanuel, R., & Kuheba, E. G. (2025). The Curative Role of the Main Primary Health Facility in Indonesia: Analyzing Patient Visits at a Single PUSKESMAS with a Focus on General and Maternal-Child Health Clinics. *Asian Journal of Research in Medical and Pharmaceutical Sciences*, *14*(2), 68–78. https://doi.org/10.9734/ajrimps/2025/v14i2306

Dinas Kesehatan Kabupaten Sukabumi. (2024). Profil Kesehatan Kabupaten Sukabumi Tahun 2024. In *Buku*. Dinas Kesehatan Kabupaten Sukabumi.

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, *99*, 650–655. https://doi.org/10.1016/j.procir.2021.03.088

Liu, W., Zhang, Y., Yang, H., & Meng, Q. (2023). A Survey on Differential Privacy for Medical Data Analysis. *Annals of Data Science*, *May*, 1–15. https://doi.org/10.1007/s40745-023-00475-3

Mohammadi, M., Vejdanihemmat, M., Lotfinia, M., Rusu, M., Truhn, D., Maier, A., & Tayebi Arasteh, S. (2026). Differential privacy for medical deep learning: methods, tradeoffs, and deployment implications. *Npj Digital Medicine*, *9*(1). https://doi.org/10.1038/s41746-025-02280-z

Panggabean, M. V., & Fitria, A. (2025). Perlindungan Hukun Data Pribadi di Indonesia (Kasus Kebocoran Badan Penyelenggara Jaminan Sosial Kesehatan). *Arus Jurnal Sosial Dan Humaniora*, *5*(3), 3678–3688. https://doi.org/10.57250/ajsh.v5i3.1741

Saputra, A., Hajar, S., & Sari, M. T. (2024). Analisis Kebijakan Kesehatan Dalam Meningkatkan Mutu Pelayanan Kesehatan Puskesmas Di Kota Medan. *Jurnal Ilmu Administrasi*, *15*(02), 210–227. https://doi.org/10.23969/kebijakan.v15i02.10182

Shrestha, S. G., & Pradhanang, S. M. (2023). Performance of LSTM over SWAT in Rainfall-Runoff Modeling in a Small, Forested Watershed: A Case Study of Cork Brook, RI. *Water*, *15*(4194), 1–15. https://doi.org/https://doi.org/10.3390/w15234194

Yuridis, A., Informasi, A., Studi, I., Penjualan, K., & Pasien, D. (2021). Juridical Analysis of Illegal Information Access: Case Study on Sales of Data Patients Covid - 19. *SOEPRA Jurnal Hukum Kesehatan*, *7*(1), 1–13. https://doi.org/10.24167/shk.v7i1.268

Zhao, D., Zhang, H., Cao, Q., Wang, Z., He, S., Zhou, M., & Zhang, R. (2022). The Research of ARIMA, GM(1,1), and LSTM Models for Prediction of TB cases in China. *PLoS ONE*, *17*(2 February), 1–18. https://doi.org/10.1371/journal.pone.0262734