JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

PREDICTING FUTURE ENROLLMENT TRENDS AT UNIVERSITAS LANCANG KUNING USING ARIMA AND LSTM MODELS

Sutejo[1*], Yogi Ersan Fadrial[1], M. Sadar[1], Mhd Arief Hasan[1]
[1]Informatics Engineering, University of Lancang Kuning
email: *sutejo@unilak.ac.id

Abstract: This research is driven by the challenges faced by Universitas Lancang Kuning (UNILAK) in attracting applicants amidst intense competition, especially after the government's policy opened independent pathways to State Universities (PTN) from 2022-2023, which impacted private university applicant numbers. To address this and support strategic planning, this study aims to predict the trend of prospective students applying to all study programs at UNILAK for the period 2025-2027. Two time series models were employed: ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory). Applicant data from 2019 to 2024 was used to build the model. The Augmented Dickey-Fuller (ADF) test confirmed the data's stationarity with a p-value of 0.0. ACF and PACF analyses determined the ARIMA parameters as p=1, d=1, q=1. The LSTM model was trained to capture more complex data patterns. ARIMA predictions for 2025, 2026, and 2027 are 3298.66, 3362.33, and 3371.30, respectively. LSTM predictions for the same years are 3335.64, 3476.52, and 3518.42. Evaluation using Root Mean Squared Error (RMSE) showed ARIMA (RMSE=588.72) to be more accurate than LSTM (RMSE=653.96). Nevertheless, LSTM provided a more optimistic prediction. This study concludes that ARIMA is better suited for short-term planning, while LSTM can be used for more ambitious long-term strategies.

Keywords: arima; LSTM; applicants; prediction; university

Abstrak: Penelitian ini didorong oleh tantangan Universitas Lancang Kuning (UNILAK) dalam menarik pendaftar di tengah persaingan ketat, khususnya setelah kebijakan pemerintah membuka jalur mandiri ke Perguruan Tinggi Negeri (PTN) sejak 2022-2023, yang menyebabkan penurunan jumlah pendaftar di universitas swasta. Untuk mendukung perencanaan strategis, studi ini bertujuan memprediksi tren jumlah calon mahasiswa yang mendaftar ke seluruh program studi di UNILAK untuk periode 2025-2027.Dua model deret waktu digunakan: ARIMA (AutoRegressive Integrated Moving Average) dan LSTM (Long Short-Term Memory). Data jumlah pendaftar dari 2019 hingga 2024 digunakan untuk membangun model. Uji Augmented Dickey-Fuller (ADF) menunjukkan data stasioner dengan p-value 0,0. Analisis ACF dan PACF menentukan parameter ARIMA sebagai p=1, d=1, q=1. Model LSTM dilatih untuk menangkap pola data yang lebih kompleks.Prediksi ARIMA untuk 2025, 2026, dan 2027 adalah 3298.66, 3362.33, dan 3371.30. Prediksi LSTM untuk tahun yang sama adalah 3335.64, 3476.52, dan 3518.42. Evaluasi menggunakan Root Mean Squared Error (RMSE) menunjukkan ARIMA (RMSE=588.72) lebih akurat daripada LSTM (RMSE=653.96). Meskipun demikian, LSTM memberikan prediksi yang lebih optimis. Studi ini menyimpulkan ARIMA lebih cocok untuk perencanaan jangka pendek, sementara LSTM dapat digunakan untuk strategi jangka panjang yang ambisius.

Kata kunci: arima; LSTM; pendaftar; prediksi; universitas

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Higher education is vital for improving human resources. Universitas Lancang Kuning (UNILAK), a private university in Indonesia, faces significant challenges in attracting new students due to a government policy implemented from 2022 to 2023 that opened indepe ndent pathways for prospective students to enter state universities (PTNs). This policy, while providing more opportunities for some students, directly led to a decline in UNILAK's applicant numbers, as many chose PTNs due to more affordable facilities and fees, thus reducing the competitiveness of private institutions.

In response to these challenges, UNILAK has proactively implemented various improvements to attract prospe ctive students and increase applicant numbers. These efforts include enhancing the quality of education and services across all study programs, optimizing learning processes, improving infrastructure, and offering scholarships. Additionally, UNILAK plans to further develop its faculties and study programs to align with industrial and future workforce needs[1], [2].Currently, UNILAK boasts 39 study programs across 10 faculties, offering diverse options in social sciences, economics, law, and technology. This wide array of programs provides UNILAK with substantial potential for growth and attracting students from various educational backgrounds. Therefore, understanding and predicting applicant trends are crucial for UNILAK to effectively plan resource management and develop its academic offerings.

This research aims to analyze and forecast the number of applicants at UNILAK for the upcoming years using ARIMA and LSTM prediction methods. The study seeks to provide a clear understanding of future applicant trends at UNI LAK, considering both the impact of government policies on private universities and the effectiveness of UNILAK's improvement initiatives.

A key contribution of this study is its novel comparison of ARIMA and LSTM, two distinct prediction methods. While ARIMA is established for time se ries analysis with stationary data, LSTM, a neural network, excels at handling more complex and non-stationary temporal patterns[3], [4], [5]. By comparing these methods, the research aims to identify which is more effective for predicting applicant numbers at UNILAK, offering new insights into prediction model usage for private universities navigating policy changes and shifts in student interest. This analysis also addresses the gap in studies directly linking government poli cy impacts to private universities, particularly in applicant prediction, and highlights the application of prediction models for strategic planning in Indonesian higher education.

METHOD

This research aims to analyze and predict the number of applicants at UNILAK using two time series-based prediction methods: ARIMA and LSTM. The study will also analyze the impact of government policies on the decline in applicant numbers at private universities like UNILAK, as well as UNILAK's improvement efforts.

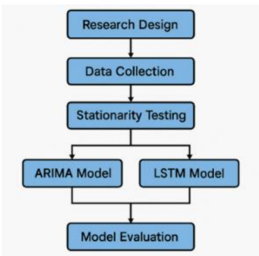

Image 1. Research Method

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Research Methodology

This study employs a descriptive-analytic research design with a quantitative approach. The research data comprises the number of prospective students, applicants, accepted students, re-registered students, and students with a student ID (NIM) at Universitas Lancang Kuning (UNILAK) from 2019 to 2024. This data, which includes information across various UNILAK study programs, is utilized to analyze applicant trends and forecast future applicant numbers. Additionally, it aids in analyzing the impact of government policies that introduced independent pathways at state universities between 2022 and 2023. Predictions for the years 2025-2027 are generated using two time series prediction methods: ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory).

Stationarity Test

The stationarity test ensures that the time series data meets the fundamental assumptions required by prediction models like ARIMA. Stationary data is characterized by a constant mean, variance, and covariance over time[6], [7], [8]. The Augmented Dickey-Fuller (ADF) Test is employed to assess data stationarity, testing the null hypothesis that the data possesses a unit root and is non-stationary. If the ADF test yields a p-value greater than 0.05, the data is considered non-stationary, necessitating differencing to remove trends. Differencing involves calculating the difference between the current and previous values in the time series to achieve stationarity. If, after differencing, the ADF test's p-value is less than or equal to 0.05, the data is deemed stationary and suitable for modeling

ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model is utilized for modeling and predicting stationary time series data. This model integrates three core components: Auto-Regressive (AR), which describes the dependence of the current data value on previous values; Integrated (I), referring to the differencing process that renders the data stationary; and Moving Average (MA), which manages the dependence between the data and prior prediction errors[9], [10], [11]. In this study, the ARIMA model was applied after confirming data stationarity through the ADF test and differencing. The ARIMA parameters (p,d,q) were determined by analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), which provided guidance for appropriate p and q values[5], [12], [13], [14].. Once these parameters were established, the ARIMA model was trained using UNILAK's applicant data to predict the number of applicants for the next three years (2025-2027).

LSTM Model

The Long Short-Term Memory (LSTM) model consists of several gates that regulate the flow of information. The primary components of an LSTM cell include:

Forget Gate: This gate determines which information from the previous cell's memory should be discarded. Its output ft is a result of the sigmoid function, controlling the extent of information to be forgotten[15], [16], [17].

Input Gate: This gate decides what new information will be added to the memory cell. Its output it) dictates how much new information will be processed[18], [19], [20], [21].

Candidate Memory Cell: This component generates a new candidate memory $\hat{C}_t$ to be incorporated into the memory cell in the subsequent step. The hyperbolic tangent (tanh) function is employed to constrain values within the range of [-1,1].

Update Cell State: The memory cell C_t is updated by combining the information to be forgotten with the new information to be added. This update involves multiplying the output of the forget gate by the previous memory and the output of the input gate by the candidate memory.

Output Gate: This gate determines which information will be released from the memory cell as the output h_t. A sigmoid function generates a value between 0 and 1, which is then multiplied by $\tanh C_t$ to control the information for prediction.

The final output of the LSTM cell (ht) is calculated based on the output gate's result and the updated memory cell.

Model Evaluation

Model evaluation is conducted after training the ARIMA and LSTM models to measure each model's prediction performance. This evaluation assesses how well each model predicts actual values, utilizing error metrics such as Root Mean Squared Error (RMSE). RMSE is one of the most commonly used metrics to quantify the error between predicted and actual values, indicating the magnitude of the model's error in the same units as the data.[3], [5], [12], [13], [22].
RMSE Formula :

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \quad (1)$$

Y_t : the actual value at time t,
$\hat{Y}_t$: the predicted value at time t
n : the number of data points (time periods)

The interpretation of RMSE is as follows: a low RMSE value indicates high prediction accuracy, meaning the difference between predicted and actual values is relatively small. Conversely, a high RMSE value suggests that the model produces less accurate predictions with larger errors. This can imply that the model requires further tuning, such as adjusting parameters or selecting more appropriate features, to enhance its accuracy and performance. A high RMSE may also indicate that the model has not fully captured the data's patterns or that significant factors have not been accounted for.

RESULT AND DISCUSSION

In this study, an analysis and prediction of the number of applicants at Universitas Lancang Kuning (UNILAK) were conducted using two time series models, namely ARIMA and LSTM. Based on the number of applicants from 2019 to 2024, predictions for the years 2025-2027 were made using both models. The results from both models were compared to determine which model is more accurate in predicting the number of applicants at UNILAK.

Results of Stationarity Test

In the first stage, the stationarity of the data was tested using the Augmented Dickey-Fuller (ADF) Test. Based on the ADF test results, an ADF statistic of -1697.32 and a p-value of 0.0 were obtained, indicating that the data is stationary (p-value < 0.05). Therefore, no further differencing steps are needed to make the data stationary.

Table 1. ADF Test Results:

ADF Statistic	p-value	Conclusion
-1697.316422	0.0	Data is stationary

ARIMA Analysis Results

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

In this stage, the ARIMA (Auto-Regressive Integrated Moving Average) model was applied to predict the number of applicants at Universitas Lancang Kuning (UNILAK) for the period 2025-2027. Before applying the ARIMA model, an analysis was conducted using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the optimal ARIMA model parameters, namely p, d, and q.

ACF and PACF Graphs

ACF Graph: This graph shows the correlation between the current data and the data at previous lags. If the points in the ACF graph exceed the boundaries of the shaded area (in light blue), it indicates a significant correlation at a specific lag, providing an indication of the value for q.

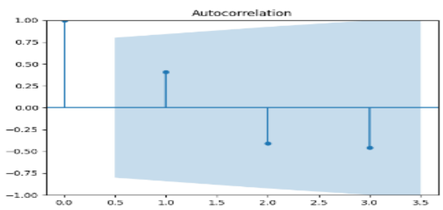

Image 2. ACF graph

Explanation: From the ACF graph, it can be seen that a significant correlation occurs at the first lag, with little correlation at the subsequent lags. This suggests that the MA(1) model (q=1) may be the most suitable for this data.

PACF Graph: This graph shows the correlation between the current value and the previous value, after removing the influence of more distant values. If the points in the PACF graph exceed the boundaries of the shaded area, it indicates a significant correlation at a specific lag, providing an indication for the value of p.

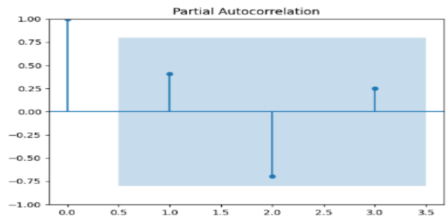

Image 3. PACF Graph

Explanation: From the PACF graph, it can be seen that significant correlation occurs at the first lag, with a sharp decline afterward. This suggests that the AR(1) model (p=1) is more suitable for this data.

ARIMA Model Selection

Based on the results from the ACF and PACF graphs, the ARIMA(1,1,1) model was chosen, which means:
p=1: AR model with one lag.
d=1: Data is differenced once to make it stationary.
q=1: MA model with one lag.

ARIMA Prediction Results

After the ARIMA parameters were determined, the ARIMA model was applied to predict the number of applicants for the years 2025–2027. The following are the prediction results obtained:

Table 2. ARIMA Prediction Results

Year	ARIMA Prediction
2025	3298.66
2026	3362.33
2027	3371.30

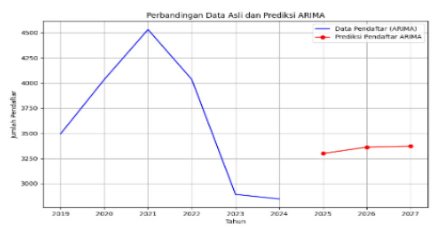

Image 4. ARIMA Prediction Results

The chart below shows a compar-

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ison between the actual number of applicants (2019–2024) and the ARIMA model predictions for the years 2025–2027.

LSTM Prediction Results

At this stage, modeling was carried out using LSTM (Long Short-Term Memory) to predict the number of applicants at Universitas Lancang Kuning (UNILAK) for the period 2025–2027. The LSTM model was used due to its ability to handle time series data with more complex patterns and long-term dependencies that may not be captured by the ARIMA model.

Preparing Data for the LSTM Model

Before training the LSTM model, applicant data was scaled using MinMaxScaler to a range between 0 and 1, then reshaped to fit the LSTM input format with a time step of 1. The LSTM model was constructed using Keras, featuring 50 LSTM units and a single output Dense layer to produce a single prediction value. The model underwent training for 100 epochs with a batch size of 32, optimizing with mean squared error (MSE) as the loss function. Following training, predictions for the number of applicants for 2025–2027 were generated from the scaled data and subsequently converted back to their original scale using inverse scaling with MinMaxScaler.

Tabel 3. Hasil Prediksi LSTM:

Tahun	Prediksi LSTM
2025	3335.64
2026	3476.52
2027	3518.42

The chart above shows a comparison between the actual number of applicants (2019–2024) and the LSTM model predictions for the years 2025–2027. This hart illustrates how the LSTM model forecasts the number of applicants for the upcoming years, with slightly higher predictions compared to the ARIMA model. The LSTM predictions indicate a more significant increase in the number of applicants, suggesting that the model captures more complex patterns in the data.

Image 5. LSTM Prediction Results

Comparison of ARIMA and LSTM Prediction Results

After analyzing and predicting the number of applicants at Universitas Lancang Kuning (UNILAK) using both models, ARIMA and LSTM, the results from these two models need to be compared to understand the differences and strengths of each.

Table 4. Comparison of ARIMA and LSTM Prediction Results

Year	ARIMA Prediction	LSTM Prediction	Difference
2025	3298.66	3335.64	37.0
2026	3362.33	3476.52	114.19
2027	3371.30	3518.42	147.12

From the table above, it can be seen that the LSTM model predicts a higher number of applicants compared to the ARIMA model for all three prediction years (2025-2027). The LSTM prediction shows a more significant increase, with the largest difference in prediction reaching 147.12 in 2027.

The following chart shows a visual comparison between the ARIMA and LSTM model predictions, helping to better understand how the two models predict trends in the number of applicants.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 3, Juni 2025, hlm. 541 – 548 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i3.3865
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

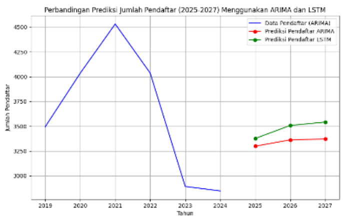

Image 6. Comparison Chart of ARIMA and LSTM Prediction

This chart shows that both models predict an increase in the number of applicants, but LSTM predicts a sharper increase compared to ARIMA.

Evaluation of Prediction Results

To evaluate the accuracy of both models, the Root Mean Squared Error (RMSE) is used as a metric to measure how far the model predictions are from the actual data. A lower RMSE value indicates a more accurate model in predicting the data.

Table 5. RMSE Evaluation:

Model	RMSE	Interpretation
ARIMA	588.72	Prediction is fairly accurate
LSTM	653.96	Prediction is less accurate

From the RMSE table, it can be seen that ARIMA has a lower RMSE value compared to LSTM, indicating that ARIMA is more accurate in predicting the number of applicants based on historical data. Although LSTM is more flexible, it has a higher RMSE, suggesting that the model may be too optimistic in predicting a surge in applicant numbers.

CONCLUSION

This study compares the ARIMA and LSTM time series prediction models to forecast the number of applicants at Universitas Lancang Kuning (UNILAK) for the period 2025-2027. The results indicate that the ARIMA model provides more stable outcomes with a slight increase in applicant numbers, reflecting a more consistent and reliable pattern for short-term forecasting. In contrast, the LSTM model offers a more optimistic forecast with a larger projected increase in applicants, but it comes with a higher RMSE, suggesting that the model is more sensitive to data fluctuations and potentially overly optimistic in predicting future trends. While ARIMA is more accurate in terms of RMSE and suitable for short-term planning, LSTM has the potential to capture more complex patterns and is beneficial for long-term strategies, especially if UNILAK aims for a significant improvement in applicant numbers. Therefore, this study's findings provide valuable insights for UNILAK in developing student recruitment strategies, considering both accuracy and the potential for future applicant growth.

ACKNOWLEGMENTS

We would like to express our sincere gratitude to the Faculty of Computer Science, Universitas Lancang Kuning, for their generous funding and support for this research in the 2025 fiscal year. Their invaluable contribution has greatly facilitated the completion of this study. We deeply appreciate their trust and encouragement throughout the research process.

BIBLIOGRAPHY

[1] M. Sitompul, M. A. Hasan, and M. Devega, "Forecasting Simcard Demand Using Linear Regression Method," *J. Res. Dev.*, vol. 8, no. 1, 2023, doi: 10.25299/itjrd.2022 .12202.

[2] H. Sunaryanto, M. A. Hasan, and G. Guntoro, "Classification Analysis of Unilak Informatics Engineering Stud ents Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3),

Random Forest and K-Nearest Neighbors (KNN)," *IT J. Res. Dev.*, vol. 7, no. 1, pp. 36–42, Aug. 2022, doi: 10.25299/itjrd.2022.8912.

[3] S. Khan, Y. Muhammad, I. Jadoon, S. E. Awan, and M. A. Z. Raja, "Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting," *Appl. Soft Comput.*, vol. 170, p. 112765, 2025.

[4] S. Toshaliyeva, "Using the Arima model to forecast the share of railways in the industry," in *E3S Web of Conferences*, 2024, p. 2024.

[5] R. Tosepu and N. Y. Ningsi, "Forecasting of diarrhea disease using ARIMA model in Kendari City, Southeast Sulawesi Province, Indonesia," *Heliyon*, vol. 10, no. 22, 2024.

[6] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *Int. J. Eng. Bus. Manag.*, vol. 10, p. 1847979018808673, 2018.

[7] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, 2014, pp. 106–112.

[8] E. Chodakowska, J. Nazarko, Ł. Nazarko, H. S. Rabayah, R. M. Abendeh, and R. Alawneh, "Arima models in solar radiation forecasting in different geographic locations," *Energies*, vol. 16, no. 13, p. 5029, 2023.

[9] R. Ospina, J. A. M. Gondim, V. Leiva, and C. Castro, "An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil," *Mathematics*, vol. 11, no. 14, p. 3069, 2023.

[10] L. Kong *et al.*, "Time-aware missing healthcare data prediction based on ARIMA model," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 2022.

[11] X. Dong, B. Dang, H. Zang, S. Li, and D. Ma, "The prediction trend of enterprise financial risk based on machine learning arima model," *J. Theory Pract. Eng. Sci.*, vol. 4, no. 01, pp. 65–71, 2024.

[12] S. Yasmin and M. Moniruzzaman, "Forecasting of area, production, and yield of jute in Bangladesh using Box-Jenkins ARIMA model," *J. Agric. Food Res.*, vol. 16, p. 101203, 2024.

[13] L. Wang, L. Chen, S. Jin, and C. Li, "Forecasting the Green Behaviour Level of Chinese Enterprises: A Conjoined Application of the Autoregressive Integrated Moving Average (ARIMA) Model and Multi-Scenario Simulation," *Technol. Soc.*, p. 102825, 2025.

[14] U. Samal and A. Kumar, "Enhancing software reliability forecasting through a hybrid ARIMA-ANN model," *Arab. J. Sci. Eng.*, vol. 49, no. 5, pp. 7571–7584, 2024.

[15] X. Zhang *et al.*, "High-speed railway seismic response prediction using CNN-LSTM hybrid neural network," *J. Civ. Struct. Heal. Monit.*, vol. 14, no. 5, pp. 1125–1139, 2024.

[16] Y. Tang, P. Dong, Z. Tang, X. Chu, and J. Liang, "Vmrnn: Integrating vision mamba and lstm for efficient and accurate spatiotemporal forecas ting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5663–5673.

[17] S. Lu *et al.*, "Surgical instrument posture estimation and tracking based on LSTM," *ICT Express*, vol. 10, no. 3, pp. 465–471, 2024.

[18] D. O. Oyewola, S. A. Akinwunmi, and T. O. Omotehinwa, "Deep LSTM and LSTM-Attention Q-learning based reinforcement learning in oil and gas sector prediction," *Knowledge-Based Syst.*, vol. 284, p. 111290, 2024.

[19] F. Dao, Y. Zeng, and J. Qian, "Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized CNN-LSTM neural network," *Energy*, vol. 290, p. 130326, 2024