EAJOS Vol. 2,No. 1, June 2026 ISSN (Online): 3088-5426 | ISSN (Print): 2953-2728

Eastern Africa Journal of Official Statistics

Website: https://journal.eastc.ac.tz/index.php/eajos/index

Machine Learning-based Analysis and Prediction of Factors Influencing Share of Agriculture Sector in the GDP in Tanzania from 1981-2024

[1]Siamarie Lyaro*, [2] Janeth P. Mchwampaka,[3] Asya H. Haji

[1,2,3] Eastern Africa Statistical Training Centre P. O. Box 35103, Dar es Salaam, Tanzania

*Corresponding Author's Email: siamarie.lyaro@eastc.ac.tz

KEYWORDS

Agricultural GDP, Machine learning prediction

ABSTRACT

Share of the agriculture sector in GDP has been declining overtime, suggesting a need to deploy machine learning method to analyse and predict factors influencing share of agriculture sector in GDP. The study used time series data from 1981-2024 and utilized LSTM, XGBoost, and Random Forest regression models to analyse and predict factors influencing agriculture GDP in Tanzania. Results showed that the l.STM model had the highest value of R^2 compared to XGBoost and Random Forest regression model. Furthermore, results from LSTM model showed that lending rate dynamics played a critical role in the prediction of the share of agriculture in GDP. Also, results from XGBoost showed that rural population share and exchange rate were influential factors which predicted share of agriculture in GDP. The results from Random Forest regression model showed that exchange rate and the first lag of the share of agriculture in GDP were factors which influenced share of agriculture in the GDP. In general findings suggest the need to monitor macroeconomic factors to prevent their ill effect on the agriculture sector. Also, findings suggest the need of modernizing agricultural activities to improve contribution of agriculture sector in the GDP over time.

1. Introduction

Agriculture sector constitutes agriculture, forestry, fishing and hunting sector in the ISIC Revision.4 (UN, 2024). Agriculture sector not only seeks to meet human needs for food, but also to provide employment, income, and raw materials for the manufacturing sector (Kocira and Staniak, 2025; ILO, 2025; UN, 2025; MU, 2025; Todaro and Smith, 2015). Contribution of agriculture sector in the global economy has been increasing as indicated in Figure 1.

Also, Figure 1 shows that Asian continent attained a notable increase in the gross production value in agriculture compared to others. Furthermore, Figure 1 shows that Africa had a least contribution in the gross production value in agriculture, and had been experiencing a gradual increase in the gross production value. Although Africa experienced a gradual increase in the gross production value in agriculture, contribution of every region varied as shown in Figure 2.

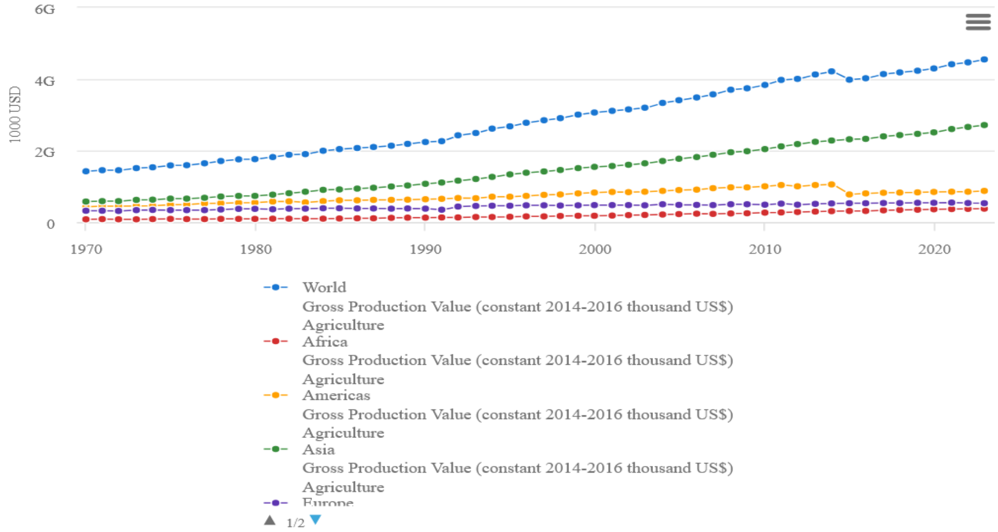

Figure 1: Gross production value in agriculture for world and continents measured in 2014-2016 constant thousand USD
Source: FAO (2025b).

Also, Figure 2 shows that Western Africa, Northern Africa, and Eastern Africa regions contributed more in the gross production value in agriculture compared to others. Along with increased gross production value in agriculture, Western and Eastern Africa regions experienced a decreased share of agriculture in GDP as shown in Figure 3.

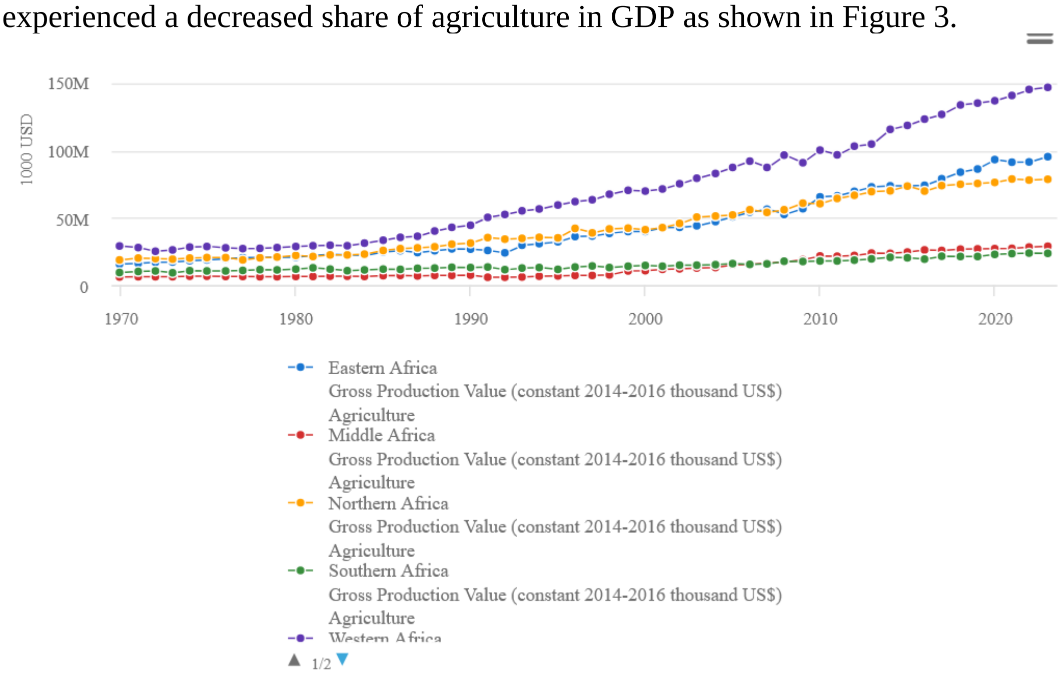

Figure 2: Gross production value in agriculture for African regions measured in 2014-2016 constant thousand USD
Source: FAO (2025b).

The Eastern African region experienced a drastic decrease in the share of agriculture in GDP from 2002 compared to others. Most of Eastern Africa countries experienced a similar decrease as shown in Figure 4. Hemrich (2021) pointed out that decrease in the share of agriculture in GDP was caused by ongoing structural transformations, while Daudi and Muba (2025) found it was due to increase in interest rates, inflation, and exchange rates.

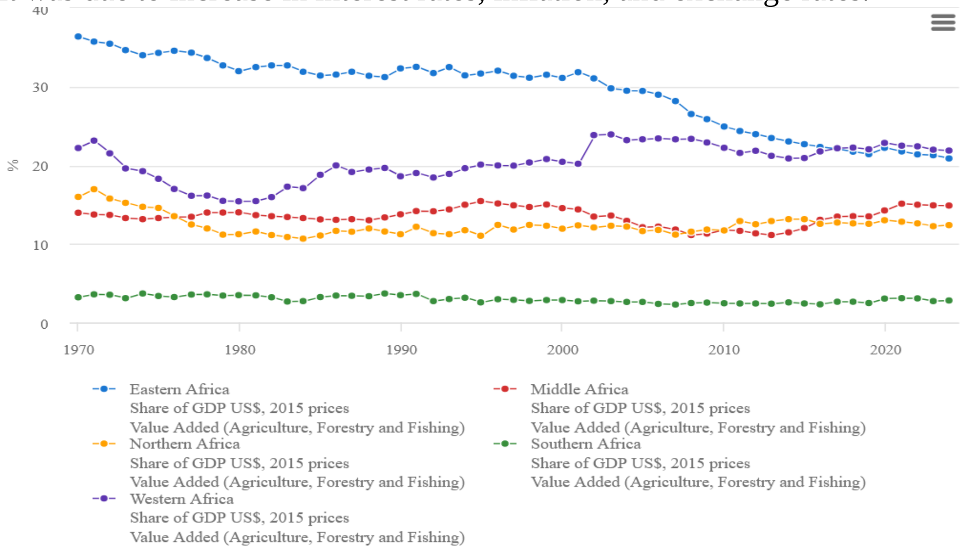

Figure 3: Agriculture's share as a percentage of GDP USD, measured in 2015 prices for African regions
Source: FAO (2025b).

Moreover, share of agriculture in GDP was affected by monetary policy rates, liquidity ratio, external debt stocks, and exports of goods and services (Ogunjinmi and Adekunle, 2025; Mohamed, 2020). Also, empirical evidence found that share of agriculture in GDP was affected by agricultural employment, expenditure in agricultural sector, gross fixed capital formation, fertilizer supply, rural population, life expectancy, and foreign direct investment (Abdelgawwad and Kamal, 2023; Sikwese *et al*., 2022; Opeyemi *et al*., 2021; De Sormeaux and Pemberton, 2011).

Conclusively, studies by Daudi and Muba (2025), Khamis and Lyaro (2025), Ogunjinmi and Adekunle (2025), Mohamed (2020), Abdelgawwad and Kamal (2023), Sikwese *et al*. (2022), and Opeyemi *et al*. (2021) reveal dominance of traditional statistical methods such as ARDL, Vector Autoregressive, and Vector Error Correction Model when explaining factors influencing contribution of agriculture sector in GDP. On the other hand, these traditional statistical methods require the link between input and output to be chosen by the user, something which may result into sub-optimal models compared to models generated by machine learning methods (Rajula et al.,2020; Ley et *al*., 2022).

Machine learning methods are increasingly being used to address time series prediction problems since they can handle many variables with complex interactions (Schmid *et al*., 2025; Ley *et al*., 2022). This implies that machine learning methods can be applied to predict factors influencing agricultural GDP growth for example in Tanzania. Currently, agriculture sector on average contributes about 24 percent of GDP in Tanzania compared to about 30 percent a decade ago, implying a its contribution decreased overtime (URT, 2025; Daudi and Muba, 2025). Although Daudi and Muba (2025) hypothesized causes of such a decrease, they used only data spanning from 1993 to 2023 without addressing non-linear relationships existed in the data set. Hence, models generated by their study were likely to yield sub-optimal predictions since they were developed by using traditional statistical methods, methods which could not handle many variables with complex interactions (Rajula *et al.*,2020; Ley *et al.*, 2022; Tjøstheim, 2025).

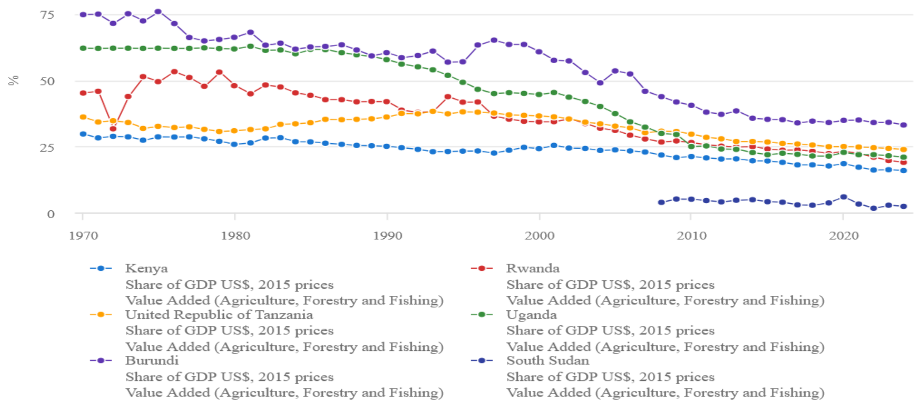

Figure 4: Agriculture's share as a percentage of GDP USD measured in 2015 prices for Eastern African countries
Source: FAO (2025b).

Therefore, this study was designed to fill identified gaps by applying machine learning method to analyse and predict factors influencing share of agriculture sector in the GDP in Tanzania from 1981-2024. Since machine learning generate high accuracy models which are useful in the formation of policies for improving performance of agriculture sector and economic growth as per SDG 2 and SDG 8.

2. Materials and Methods

a) Study area

The study area is Tanzania, an Eastern African country, the country which is located at a latitude of -6.369028 and a longitude of 34.888822. The country has been experiencing a decreased share of agricultural sector in GDP overtime (URT, 2025; Daudi and Muba, 2025).

b) Research Design

The study was a quantitative one. The study used annual time data spanning from 1981-2024 in the analysis and prediction of factors influencing share of agriculture sector in the GDP in Tanzania from 1981-2024 by using Machine Learning method. This design was also applied by (Rajula *et al*.,2020; Ley *et al*., 2022; Tjøstheim, 2025).

c) Data Source

The study used annual time series data spanning from 1981-2024. Data were obtained from the Bank of Tanzania (BoT), and the National Bureau of Statistics (NBS).

d) Data Analysis Methods

i. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is suitable for multivariate time series forecasting (Štrimaitis et al., 2025). LSTM is variant of Recurrent Neural Networks which uses sophisticated gating mechanisms enable them to execute multivariate time series forecasting, and as well as analysing bi-directioanal relationships among variables (Hall and Rasheed, 2025; Tjøstheim, 2025). The LSTM model processes data sequentially and retains information from previous periods through memory cells. In this study, the model was trained using five-year sequences, allowing it to learn delayed and cumulative effects of macroeconomic conditions on share of agriculture sector in GDP.

ii. XGBoost

XGBoost is a Tree-Based Machine Learning method, is suitable for analysing non-linear relationships which exist in the data set (Hall and Rasheed, 2025). XGBoost, an ensemble learning algorithm based on Classification and Regression Trees, demonstrates high computational efficiency and predictive accuracy, and high ability to handle missing data automatically (Shangguan, 2025). In this study XGBoost was trained on scaled features for regression prediction of the target variable. The model was capable of handling interactions and overfitting control.

iii. Random Forest Regression

Random Forests is an ensemble learning method that constructs multiple decision trees and combines their outputs through averaging for regression tasks or majority voting for classification tasks (Hall and Rasheed, 2025). In this study Random Forest model was trained on scaled features for regression prediction of the target variable. The model captured non-linear relationships and feature importance.

iv. Data Pre-processing and feature engineering

Before processing data, all variables were renamed. Column names were standardized and renamed for clarity. Furthermore, missing values were handled by using interpolation or imputation. Data scaling was conducted purposely to standardize all input features to ease comparison of machine learning models. Sequence for LSTM model was created by using data for five consecutive years to capture temporal dependencies the LSTM model. The sequence of five consecutive years provides sufficient historical context to model the influence of past years on the target variable while keeping enough samples for training.

v. Model performance tests

Performance of models was tested in the study by using RMSE, MAE, and R^2 for time-series cross-validation. Model performance is essential for optimizing a machine learning model, and reports how well a machine learning model carries out the task for which it was designed based on various metrics (IBM, 2026).

3. Results and Discussion of Findings

This section represents findings and discussion of factors influencing agriculture GDP in Tanzania using LSTM, XGBoost and, Random forest regression models.

a) Model evaluation and accuracy

Results presented in Table 1 show that R^2, RMSE and MAE were employed as metrics of error analysis for regression models. These metrics provide comprehensive analysis for model evaluation (Hall and Rasheed, 2025). Furthermore, results in Table 1 show that LSTM model significantly outperformed the Random Forest model and the XGBoost, suggesting that temporal dependencies are central in explaining changes in the share of agriculture sector in GDP in Tanzania. The LSTM model had the highest value of R^2 compared to others, and had minimal values of RMSE and MAE respectively.

Table 1: Analysis of models' accuracy

Model type	RMSE	MAE	R^2
LSTM	1.96	1.29	0.97
XGBoost	6.53	4.86	0.74
Random Forest	7.89	5.34	0.69

Source: Authors' compilation (2026).

b) Long Short-Term Memory (LSTM) Predictions

Results presented in Figure 5 show prediction of the share of agriculture in GDP by using LSTM model.

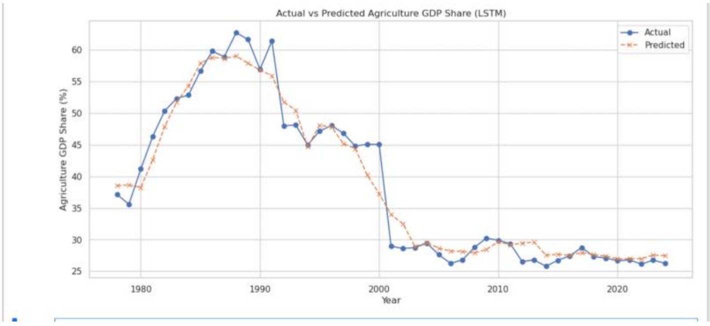

Figure 5: **LSTM predictions of the share of agriculture in GDP**

Source: Authors' compilation (2026).

This implies that LSTM predictions for the share of agriculture in GDP Y=(Y_{t+1},Y_{t+2}, ,…, Y_{t+i}) are a function of time series X=(X_1, X_2,,…, X_i), X_i being lending rate, inflation, exchange rate,expenditure on agriculture, gross capital formation, foreign direct investment, export of goods and services, value of fertilizer supplied, share of the rural population, total external debt, first lag of the share of agriculture in GDP, second lag of the share of agriculture in GDP, interaction of the inflation and lending rate, and interaction of total external debt and exchange rate.

Regarding feature contribution, SHAP analysis of LSTM sequences showed that lending rate had the highest contribution, suggesting that temporal lending rate dynamics play a critical role when sequential dependencies are explicitly modeled. Lending rate, a monetary policy rate passing-through to retail bank interest rates, usually meets the short and medium-term financing needs of the private sector (Mbowe, 2015;WB, 2026). Monetary policy rate had negative impact on agricultural performance in the long run underscoring the critical role of maintaining lower policy rates to support agricultural growth (Daudi and Muba, 2025, Olusola and Adekunle, 2025; Wagao et al, 2025). Lower policy rates led to agriculture sector development (Exaud and Mwaitete, 2022).

c) XGBoost Predictions

Results presented in Figure 6 and in Figure 7 show that exchange rate was the leading influential factor which predicted share of agriculture in GDP by XGBoost model. Daudi and Muba (2025) found that exchange rate had a statistically significant negative impact on the growth of the agricultural sector, while Waagao *et.al* (2025) found exchange rate has a positive and significant impact on agricultural GDP in the long-run. This implies that exchange rate can either increase or decrease agriculture GDP, hence there is need to monitor monetary policies to ensure exchange rate will have a positive impact on agriculture GDP over time.

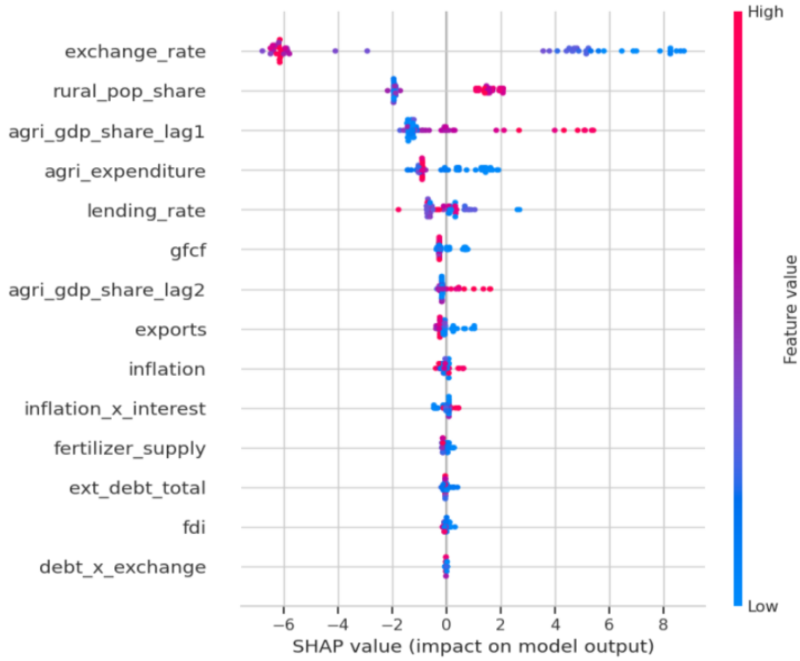

Figure 6: Feature Contributions (SHAP Analysis) by XGBoost

Source: Authors' compilation (2026).

Also, results in Figure 6 and in Figure 7 showed that rural population share was also the other influential factor, next to the exchange rate, which predicted share of agriculture in GDP by XGBoost model. De Sormeaux and Pemberton (2011) found that rural population significantly determined the percentage contribution of agriculture to GDP. Furthermore, Majaba and Amani (2025) found that labour force influenced positively economic growth implying that the rural population serves as the fundamental driver of agricultural production, the production which is very labor intensive in most of the African countries. Abdelgawwad and Kamal (2023) found that in the long run, every 1% increase in agricultural employment results in an increase in the agricultural GDP, implying that agriculture sector is labour intensive.

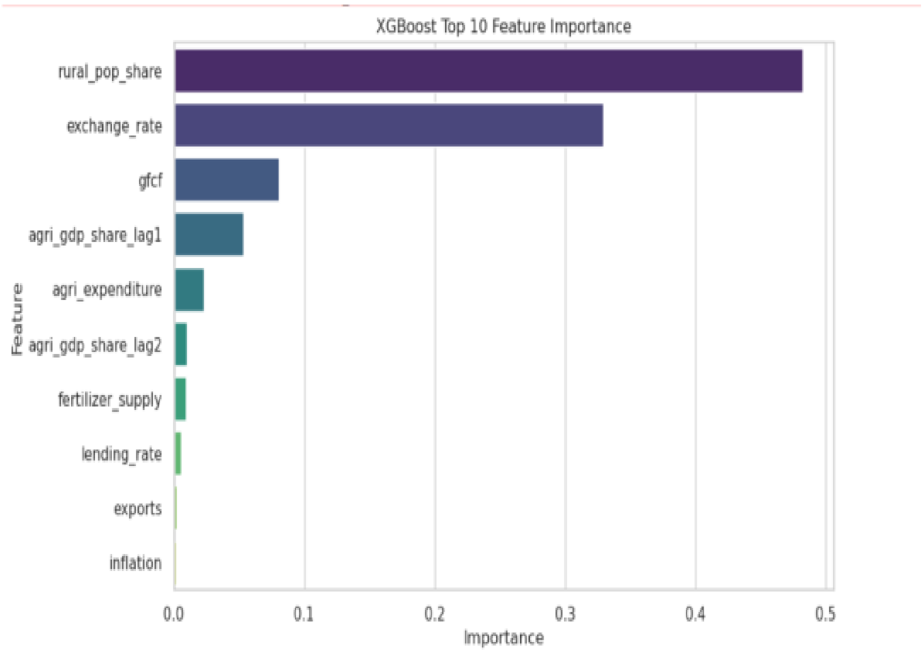

Figure 7: Top ten feature importance by XGBoost

Source: Authors' compilation (2026).

d) Random Forest Predictions

Results presented in Figure 8 show that exchange rate was the leading influential factor which influenced share of agriculture in GDP by Random Forest model. This implies that fluctuations in exchange rate influence cost of agricultural production activities which in most cases utilize imported inputs such as agro-chemicals. Hence, this suggests the need for continuous implementation of targeted input subsidy programmes as the sector is also vulnerable to external shocks (IMF, 2025; Ndifwa 2026).

Results presented in Figure 8 show that past values of share of agriculture in GDP was the other determinants of the future values of the share of agriculture in GDP. This implies that performance of agriculture sector in past years determines present performance of the sector. This relationship was captured in the Random Forest model due to the fact that the model uses randomness in selecting training instances and input features, and handles collinear data for final predictions by aggregating the predictions of individual trees, thereby mitigating over-fitting and enhancing accuracy (Muhanguzi *et al*., 2025).

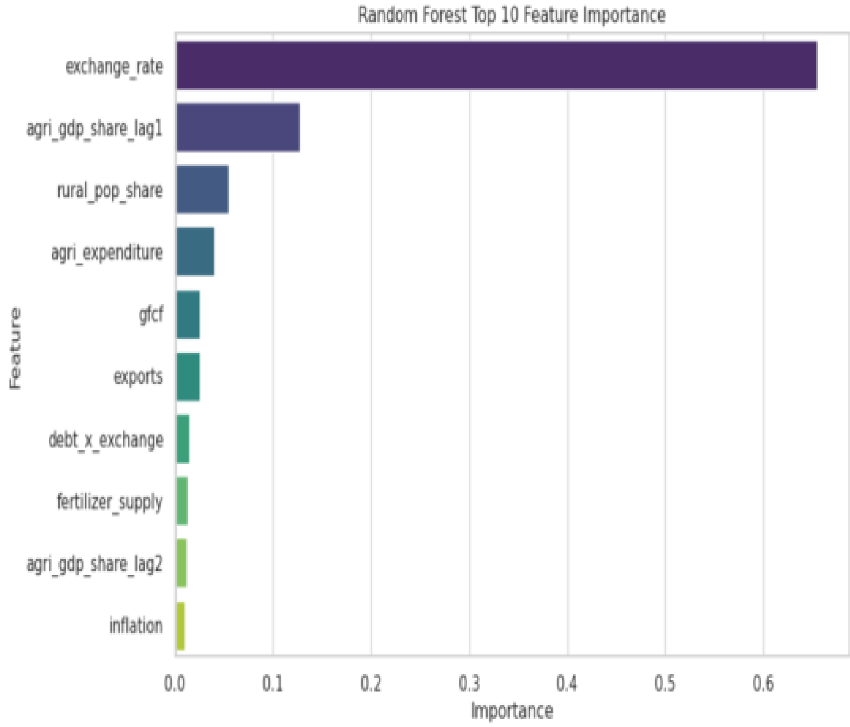

Figure 8: Top ten feature importance by Random Forest Regression Model

Source: Authors' compilation (2026).

4. Conclusions and Policy Implications

The study sought to analyse and predict factors influencing share of agriculture sector in the GDP in Tanzania by using machine learning method. The study used annual time series data from 1981-2024 from the Bank of Tanzania and the National Bureau of Statistics of Tanzania. Findings showed that the LSTM model had the highest value of R^2 compared to XGBoost and Random Forest regression model. This implies that the LSTM model provided high accuracy model for predicting factors affecting share of agriculture in GDP. Furthermore, results from the LSTM model showed that lending rate dynamics played a critical role in the prediction of the share of agriculture in GDP, implying that the banking sector has to offer affordable loans to farmers so that they can invest more and profitably in agriculture. Furthermore, results from XGBoost showed that rural population share and exchange rate were influential factors which predicted share of agriculture in GDP. This implies that there is a need to improve mechanization of agriculture to substitute human labour. Also, results from XGBoost suggest the need for continuous use of targeted input subsidy programmes for smallholder farmers. Also, results from Random Forest regression model showed that first lag of the share of agriculture in GDP was an influential factor which predicted share of agriculture in GDP. This implies that performance of agriculture sector in past years determine the present performance. Hence, there is a need to review agricultural policies periodically to improve current and future performance of the sector. Lastly, findings suggest the need to monitor and evaluate fiscal, monetary, and technological strategies to improve performance of agriculture sector in the long run.

Acknowledgements:

Authors would like to express their sincere gratitude and appreciation to the Eastern African Statistical Training Centre, Bank of Tanzania, and National Bureau of Statistics for their sincere support.

Conflict of Interest

Authors declare no conflict of interest.

References

Abdelgawwad, N., A. and Kamal, A.L.M. (2023). Contributions of investment and employment to the agricultural GDP growth in Egypt: An ARDL approach. *Economies*. 11 (8), 1-16. https://doi.org/10.3390/ economies11080215

Daudi, P.D., and Muba, S. (2025). The effect of macroeconomic factors on agricultural sector growth in Tanzania. *African Journal of Economics and Business Research*, 4(1), 18-38. Doi:https://dx.doi.org/10.4314/ajebr.v4i1.2

De Sormeaux, A., and Pemberton, C. (2011). *Factors Influencing Agriculture's Contribution to GDP: Latin America and the Caribbean*. Retrieved from https://ideas.repec.org/p/ags/cars11/187331.html

Exaud, N., nad Mwaiatete, C. (2022). Assessment of financial institutions credits on agriculture sector development in Tanzania: A Case of Arumeru district. *Accountancy and Business Review*, Vol. 14(1), 1-10,

FAO. (2025a). *Highlights*. Retrieved from https://www.fao.org/statistics/highlights-archive/highlights-detail/gross-domestic-product-and-agriculture-value- added-2014-2023/en

FAO. (2025b). *Compare Data*. Retrieved from https://www.fao.org/faostat/ en/#compare

Hall,T., and Rasheed, K. (2025). A survey of machine learning methods for time series prediction. *Applied Sciences*. 15 (11), 1-33, https://doi.org/10. 3390/app15115957

Hemrich, G. (2021). G*aps, Challenges and Progress in Boosting Agricultural Productivity and Ending Hunger*. Retrieved from https://www.un.org/deve lopment/desa/dspd/wp-content/uploads/ sites/22/2021/05 /Hemrich_Paper.pdf

IBM. (2026). *What is Model Performance?* Retrieved from https://www. ibm. com/ think/topics/model-performance

ILO. (2025). *Agriculture, Aquaculture, Plantations & Other Rural Sectors.* Retrieved fromhttps://www.ilo.org/topics-and-sectors/industries-and- sectors/agriculture-aquaculture-plantations-other-rural-sectors

IMF. (2025). *Agricultural Producer Subsidies: Navigating Challenges and Policy Considerations*. Retrieved from https://www.elibrary.imf.org/ view/ journals /068/2024/002/article-A001-en.xml

Kocira, A., and Stanik, M. (2025). Role of agriculture in implementing the concept of sustainable food system. *Agriculture*, 15 (10), 1-5. https://doi.org/10.3390/agriculture15101041

Ley, C., Martin, R. K., Pareek, A., Groll, A., Seil, R., and Tischer, T. (2022). Machine learning and conventional statistics: making sense of the differences. *Knee Surgery, Sports Traumatology, Arthroscopy*.30 (3), 753–757. https:// doi.org/10.1007/s00167-022-06896-6

Khamis, J.K., an d Lyaro, S. (2025). Impact of value added by agriculture sub- sectors on national income of Tanzania from 1990 to 2023. *Eastern Africa Journal of Official Statistics.* 1(2), 1-15. DOI: https:// doi.org/10.63933/ eajos.1.2.2025.18%20

Majaba, H.D., and Amani, L.N.M. (2025). Contribution of agricultural exports to the economic growth in Tanzania from 1984 to 2023. *Eastern Africa Journal of Official Statistics.* 1(2), 1-19. DOI**:** https://doi. org/10. 63933/ eajos. 1.2. 2025.51%20

Maryville Univerisity (MU). (2025). *Why is Agriculture Important? Benefits and Its Role*. Retrieved from https://online.maryville.edu/blog/why-is-agriculture- important/

Mbowe, W.E. (2015). *Monetary Policy Rate Pass-through to Retail Bank Interest Rates in Tanzania*. Retrieved from https://www. bot.go.tz/Publications /Other/ Working%20Papers%20Series/sw/2020021122482941240.pdf

Muhanguzi, H., Bassi, F., Mwanga, Y., and Wokalada, J. (2025) , *Statistical Approaches for Service Delivery Differentials as Assessed Through a Composite Indicator: Application to Ugandan Local Governments*. Retrieved from https://doi.org/10.1371/journal.pone.0338264

Ndifwa, J. (2026). Effect of agriculture sector performance on economic growth in Sub-saharan Africa: Application of Panel CCS-ARDL. *Turkish Journal of Agriculture-Food Science and Technology*,14(2), 387-394 DOI:https://doi. org/10.24925/turjaf.v14i2.387-394.8193

Olusola, O.O., and Adekunle, E.A. (2025). Asymmetric effects of monetary policy on agricultural performance in Nigeria. *African Journal of Economic Review*, 13 (1), 95-111.

Opeyemi, G., Olusegun, S.S., Taiwo, A., and Mobolaji, A.O. (2023). Impact of agricultural input supply on agricultural growth in Nigeria. *Journal of Applied Science and Environmental Management*. 25 (7),1317-132. DOI: https://dx.doi.org/10.4314/jasem.v25i7.30

Rajula, H.S.R., Verlato, G., Manchia, M., Antonucci, N., and Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*, 56(9), 1-10. doi: 10.3390/ medicina 56090455

Shangguan, Z. (2025). An explainable machine-learning framework based on XGBoost–SHAP and big data for revealing the socioeconomic drivers of population urbanization in China. *Systems*, 13 (8), 1-21.https://doi.org/ 10.3390/systems13080679

Sikwese, N.G., Mmasa, J. J., and Mwinuka, L. (2022). The effect of recurrent budget expenditure on the growth of agricultural sector in Tanzania. *Tanzania Journal of Agricultural Sciences,* 21(1), 138-148.

Štrimaitis, R., Ramanauskaitė S., and Stefanovič, P. (2025). Automated selection of time series forecasting models for financial accounting data: synthetic data application. *Electronics*. 14(7), 1-20. https://doi.org/ 10.3390/ electronics 14071253

Tjøstheim, D. (2025). Selected topics in time series forecasting: statistical models vs. machine learning. *Entropy*. 27(3), 1-34. https://doi.org/10.3390/e27030279

Todaro, M.P., and Smith, S.C. (2015). Economic Development. (12th Ed). Pearson Education, Inc. Boston. 856pp.

UN. (2024). *SDG Indicator Metadata*. Retrieved from https://unstats.un.org/ sdgs/metadata/files/metadata-02-0a-01.pdf

UN. (2025). *Zero Hunger*. Retrieved from https://www. un.org/ sustain abledevelopment/hunger/

United Republic of Tanzania (URT). (2025). Agricultural Sector Development Strategy - II 2015/2016 –2024/2025. Retrieved from https://faolex.fao.org/ docs/pdf/tan160643.pdf

Wagao, J. H, Akyoo, M.A, and Ndyetabula, W.D. (2025). The impact of bank credit on agricultural growth (Agdp) in Tanzania. *International Journal of Development and Management Review*, 20(1), 1-23, Doi:https://dx.doi. org/10.4314/ijdmr.v20i1.1

WB.(2026). *Metadata Glossary*. Retrieved from https://databank. worldbank.org/metadataglossary/world-development-indicators/ series/ FR INR.LEND