Maize Crop Yield Prediction Model Using Machine Learning

Thobekile Chigwada, Mary Dzinomwa, Belinda Ndlovu, Khulekani Sibanda, Sibonile Moyo
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
n02162107w@students.nust.ac.zw, mary.dzinomwa@nust.ac.zw,
belinda.ndlovu@nust.ac.zw, khulekani.sibanda@nust.ac.zw, sibonile.moyo@nust.ac.zw

Abstract

Predicting crop yield in general is critical for agricultural planning, resource allocation, and food security. The recent adverse effects of climate change on agricultural productivity have given rise to crop yield prediction which is essential in assisting farmers in anticipating the worst and preparing accordingly. Maize is one of the most popular staple foods in sub-Saharan Africa, and at the same time is the most affected by climate change. Knowing the expected maize yield in advance therefore helps those communities that are heavily dependent on maize, to prepare in advance to handle the projected situation. Despite several efforts that have been done to predict maize crop yield, the topic remains an open discussion. This paper joins the debate and uses data analytics and machine learning to build a reliable predictive model for estimating maize crop yield. Using historical maize crop yield, weather, and environmental data, the Random Forest Regressor (RFR) technique is trained to capture non-linear patterns and identify complex correlations in the data. The results of the model's evaluation show an accuracy of 70.52% demonstrating the model's ability to capture a substantial portion of the variability in the data. The predictive model can be useful for maize production farmers and maize grain depots, as it gives them time to plan for the future. Policymakers may benefit from the model, as it helps them to make informed policies on grain management. Future research should consider testing this model in a live environment, before packing it for deployment on a large scale.

Keywords
Random Forest Regressor, Machine Learning, Crop Yield Prediction.

1. Introduction

Crop yield prediction is an important area of research in the field of agriculture and "has the potential to increase productivity and improve food security" (Khaki and Wang 2019). The availability of large amounts of data from numerous sources like as satellite imagery, weather stations, and sensors, has enabled the development of machine-learning models for crop yield prediction. Machine learning algorithms are designed to learn patterns from observations and relationships in data (Foote 2022). This allows them to make predictions and identify trends. For crop yield prediction, machine learning models can be trained on historical crop data to predict future yields based on a range of factors. Crop yield prediction is normally based on soil type, meteorological, environmental, and crop parameters. Climate change, soil degradation, and lack of modern tools and technology have resulted in less productivity by many farmers. Using smart farming tools like environmental sensors that gather features such as light, temperature, humidity, and soil quality in real-time enables farmers to make smart, informed decisions to enhance productivity (Ground Control 2021).

Yield loss, normal yield, and yield increment determine continuity in farming prospects. Globally, crop yield prediction is constantly evolving as new technologies and methods are developed to improve accuracy and reliability. Remote sensing technology together with machine learning algorithms is progressively being used to predict crop yields. These tools allow for the collection of data from various sources. Gornott et al. (2021) conducted a study in Germany that demonstrated that remote sensing combined with machine learning could accurately predict crop yields. Climate change is projected to have a significant impact on crop yields globally (SubbaRao et al. 2023). This has resulted in new methods of crop yield prediction being developed to account for changing weather patterns. A study by (Zhang et al. 2020) in China shows that climate variables could be used to predict crop yields. The availability of large datasets is increasingly being used to improve crop yield prediction. For example, (Liu et al. 2021) did a study in China that showed that big data analytics could be used to predict crop yields of maize, wheat, and rice. The integration of crop models with remote sensing data is another approach that (Liu et al. 2021) in China studied and concluded that it could improve crop yield.

Crop yield prediction is an important aspect of agriculture, particularly in southern Africa where agriculture is a major “source of livelihood” (Sazib et al. 2020) for many people. In Zimbabwe, for instance, agriculture contributes significantly to “the country's Gross Domestic Product (GDP) and “provides employment to about 70%” (FAO 2023) of the population.
Multiple studies on crop yield prediction in Zimbabwe and other southern African nations have focused on using various techniques such as statistical models, remote sensing, and machine learning algorithms (Paudel et al. 2021).

One study by (Munyati and Jirah 2020) examined the use of machine learning algorithms in predicting maize yields in Zimbabwe. The study used historical climate data and satellite imagery to train several machine-learning models, such as random forest, support vector machine, and neural networks. The results of this study showed that the random forest model performed better in predicting maize yields, with an accuracy of 87.5%.

Another study by (Mhizha et al. 2021) used statistical models to predict maize yields in Zimbabwe. The study used climate data, soil data, and maize yield data from 2008 to 2017 to develop a statistical model for predicting maize yields. The study found that the model had an accuracy of 84.4%, indicating that statistical models can be effective in predicting crop yields in Zimbabwe. According to (Kaptymer et al. 2019) “Agriculture practices and technologies are still very limited in Africa for a host of reasons”. Climate-smart agriculture faces a number of obstacles in Africa, due to the lack of understanding of the practical concept, as well as a lack of data and information, as well as proper analytical tools at the local and national levels (Kaptymer et al. 2019). “Climate change is threatening agricultural productivity and some of Zimbabwe’s key agricultural challenges include low soil fertility, reliance on rain-fed systems” (FAO 2023), poorly functioning markets, and farmers having “limited access to credit, knowledge, and best practices” (Nyahunda and Tirivangasi 2019). According to (Phiri 2020) “Zimbabwe has been slow to embrace satellite-based systems for its agricultural sector” which provides one of the key attributes (remote sensing data) of crop yield prediction. They also state that poor road infrastructure, poor internet connectivity, and lack of electricity for some communities have been a challenge for farmers to get timeous expert advice on farming.

In developing countries such as Zimbabwe, there is a significant challenge in accurately predicting crop yield in agriculture. Zimbabwe is an agriculture-dependent country, with over two-thirds of the population engaged in some form of farming (FAO 2023). However, crop production is liable to unpredictable weather patterns, soil degradation, and limited access to essential resources. Accurate crop yield predictions are crucial for farmers and policymakers in making informed decisions about resource allocation, crop management, and food security.
The following questions guide this research:

Table 1. Research Questions

No.	Question Text
RQ1	Do farmers have access to accurate information for decision-making?
RQ2	Is there too much data that traditional methods cannot efficiently analyse?
RQ3	Are there prediction models currently being used by farmers in Zimbabwe?

To answer the research questions posed in Table 1, we discuss the related literature in Section 2 and build the model in Section 3. Section 4 discusses data collection for the study and training of the model. The results are then presented in Section 5, and the conclusion is done in Section 6.

2. Literature Review

Crop yield prediction has been an important area of research in agriculture for many years (Nyéki and Neményi 2022). Farmers have relied on “their experience and intuition to estimate crop yields” (Nuthall and Old 2018), but this method is often inaccurate and can lead to poor decision-making. It is now possible to predict crop yields with a higher accuracy using data-driven approaches because of the increasing availability of agricultural data and the development of machine learning techniques.

2.1 Related Work

Several researches have considered the use of machine learning algorithms for crop yield prediction. Renuka and Terdal (2019) found that most studies used regression techniques such as linear regression, support vector regression, and random forest regression for crop yield prediction in the studies they conducted. Other studies by (Jhajharia et al. 2023) have used more advanced techniques such as deep learning and neural networks to predict crop yield. They state that crop yield is assumed to be a non-linear function of environmental variables. The research conducted by (Paudel et al. 2023) showed that deep learning has the potential to learn features automatically and produce reliable crop yield forecasts.

One of the significant challenges in crop yield prediction is data availability and data quality. Singh and Sinha (2019) conducted a review of crop yield prediction and indicated that data pre-processing is a critical step in addressing the challenge. Data pre-processing is "cleaning, transforming and integrating" (Deepak 2023) raw data into formats that can be used by machine learning algorithms. This step is important in ensuring the accuracy and reliability of models. However, some variable data may be missing, insufficient, inaccurate, inconsistent, or in formats that cannot be processed by the algorithms. This impacts the process of cleaning data. The selection of relevant features(variables) is also essential in ensuring the accuracy of crop yield prediction models. Features that can have a significant impact on crop yields are weather conditions, soil quality, and crop type. Several studies have highlighted the use of remote sensing data to extract features like vegetation indices and land surface temperature for crop yield prediction (Ji et al. 2021).

Real-time monitoring and decision-making are the other challenges in crop yield prediction. Several studies have explored the use of integrated systems that "allow for real-time data collection, analysis, and decision-making" (Wang et al. 2021) to address these challenges. These systems include sensors for data collection, cloud-based platforms for data analysis, and decision-support tools for farmers.

Singla and Jindal (2019) conducted a review of machine-learning techniques used for crop yield prediction and highlighted the importance of data preprocessing and feature engineering in improving model accuracy. The study highlighted the need for weather, environmental, crop history, and satellite imagery data for crop prediction using machine learning algorithms.

Huang et al. (2020) proposed a big data platform for crop yield prediction that integrates multiple data sources that include weather data, environmental data, and remote sensing data. They used machine learning algorithms like random forest and support vector machines to predict crop yield for different crops in China. Their results showed that the proposed platform can provide accurate and timely information for farmers to make informed decisions.

In the article by (Fountas et al. 2015), they discussed the importance of farm management information systems (FMIS) as a means for collecting and analyzing data "for precision agriculture". They indicated that FMIS can provide real-time data on crop yield, soil moisture, and nutrient levels and that these can be used for future crop yield predictions. They also highlighted the need for standardized data collection and sharing protocols to ensure the interoperability of FMIS.

According to (van Klompenburg et al. 2020), accurate prediction helps farmers to make informed decisions on when and what to grow. Machine learning is viewed as an important tool for assisting judgments in crop yield prediction. This includes decisions about which crops to cultivate and what to consider during the crop-growing season. Crop yield prediction is a challenge in precision and smart agriculture and "requires the use of several datasets as crop yield depends on many factors such as climate, weather, soil, use of fertilizer, and seed variety" (van Klompenburg et al. 2020).

Shaikh et al. (2022) identify security and privacy as the major challenges and future scope of data mining in smart agriculture. Also, agricultural information systems frequently have issues with data availability and quality, (Shaikh et al. 2022). For example, real-time data presents some complications in data handling.

3. Methods and model building

In this section, we detailed the methodology employed to construct the maize crop yield prediction model using machine learning algorithms like Linear Regression and Random Forest regressor based on their suitability for regression tasks and their ability to handle different levels of complexities in the data. The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology was outlined.

3.1 CRISP-DM methodology

The CRISP-DM methodology guided the research to ensure a systematic and comprehensive model development process because of its iterative and structured approach. It consists of six key stages which include "Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment" (Smart Vision 2023).

3.2 Data processing and feature selection

A successful machine-learning model depends on the quality and readiness of the input data. In the Data Understanding stage, we collected which included historical crop yield data and meteorological data. The data underwent preprocessing to handle missing values and outliers to ensure data quality. Feature engineering was conducted to extract relevant information such as season length, rainfall, and temperature indices so as to capture crucial variables that influence maize crop yield.

3.3 Model development

This stage encompassed the selection of machine learning algorithms.

3.3.1 Linear Regression Algorithm

This is a fundamental algorithm in statistical modelling that establishes a baseline for prediction models. Despite its simplicity, Linear Regression can capture linear relationships between features and target variables. While this algorithm provides a straightforward interpretation of feature importance, its performance may be limited when dealing with complex non-linear relationships essential in agricultural systems.

3.3.2 Random Forest Regressor

Since agriculture data is of nonlinear nature, the RFR algorithm was selected to enhance prediction accuracy. By creating an ensemble of decision trees, RFR can capture complex interactions between features which results in improved predictive power. It also has an ability to mitigate overfitting which makes it well-suited for handling agricultural data. The RFR was initialised with 100 estimators which are the number of the decision trees. Other hyperparameters include the maximum depth of each tree (max_depth), and the number of features to consider at each split (max_features).

3.4 Model iteration and refinement

The CRISP-DM methodology's flexibility enabled model iteration and refinement. The iterative process aimed to strengthen the models' predictive power and align it more effectively with the complexities of maize crop yield prediction.

4. Data Collection and Preprocessing

Data collection in this study involved gathering relevant information and observations from the Meteorological Services Department of Zimbabwe and the Ministry of Agriculture. The steps included identifying relevant data sources, collecting data samples, preprocessing and cleaning the data, and organizing it in a format suitable for training machine learning models. However, not all required variables were provided.

4.1 Data quality assessment

Data quality was assessed using the six dimensions of data quality and the results of this assessment are as follows:

- **Accuracy**: The data was deemed fairly accurate as it was provided by the recording institutions.
- **Conformity:** The datasets contained standard formats and correct data types for all data.
- **Integrity:** All data was assumed to be of relatively high integrity as it is provided by authentic sources.
- **Completeness:** The three datasets used in this study complement each other to form a complete dataset. Individual datasets provided pieces of required information which were then merged to form a complete dataset.
- **Uniqueness:** No overlap or duplicates were found in the datasets.

- **Consistency**: Individual datasets contained consistent data. However, the layouts in the datasets were different and had to be transposed using Microsoft Excel tools before being loaded into Jupyter Notebook.

4.2 Data Cleansing and Visualisation

Crop data and weather data were merged into one dataset and then normalised on a scale of 0 to 1 to remove feature importance since they contained different value ranges. Data exploration was carried out to identify patterns or relationships to help explain the data and identify the correlation of the independent variables with the target variable (crop yield). Figure 1 shows a moderate negative correlation between temperature and yield (-0.481), and a moderate positive correlation between rainfall and yield (0.510).

```
Temperature-Yield correlation: -0.4811257777444185
Rainfall-Yield correlation: 0.5098926052666266
```

Figure 1. Correlation of temperature and rainfall with Yield

Two important features were analyzed with the aim of communicating the insights derived from the analysis in a clear and concise manner. Figure 2 shows the relationship between yield, temperature, and rainfall drawing a conclusion that rainfall is correlated to crop yield compared to temperature.

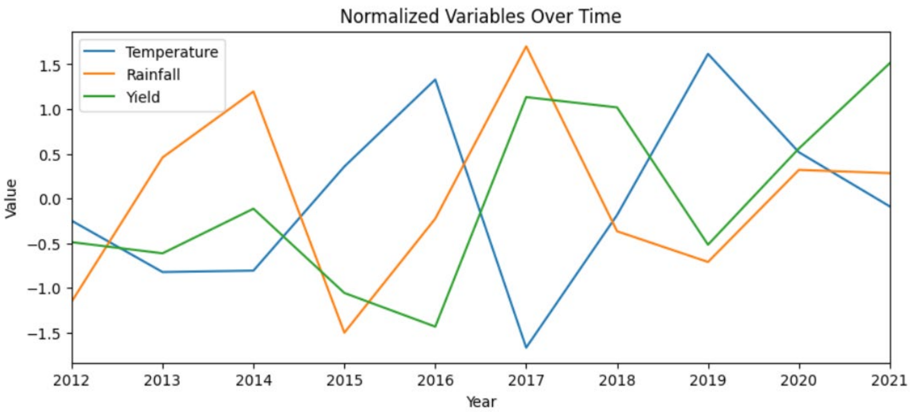

Figure 2. Relationship between yield, temperature, and rainfall over ten years

4.3 Model training and evaluation

The prepared dataset was partitioned into training and testing sets to facilitate model training and evaluation. The two algorithms were trained and evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) metrics. The data was split into training and testing sets using a ratio of 9:1 to enable the algorithms to learn the underlying patterns and relationships between the input features.

5. Results

The outcomes of the prediction models were presented and their implications for farmers' decision-making process stated.

5.1 Model performance evaluation

The MAE, MSE, and R2 were used to evaluate the models and provided information on their accuracy. The two algorithms comparison is shown in Table 2 below.

Table 2. Comparing Linear Regression and Random Forest Regressor

Evaluation Metric	**Machine Learning Algorithm**	
	Linear regression	Random forest
Accuracy (R^2 score)	0.5925	0.70516

MSE	4469565.703469417	3233876.394825
MAE	1821.620449308044	1476.1725000000001

The Linear Regression R-squared value gave an approximate 59.3% accuracy showing that the variance in the dependent variable (crop yield) is explained by the independent variables that were used in the model. The results of the Random Forest Regressor model of 70.5% achieved a relatively lower Mean Squared Error and a moderate R-squared value indicating a reasonable level of accuracy and a reasonable amount of variance explained by the independent variables. This suggests that the RFR model captures a reasonable portion of the variability in the data, indicating moderate predictive power.

Using the model, it is possible to make future yield predictions using forecasted values by inputting the forecasted parameter values. The model also has the ability to output the predicted value for the required year. Figure 3 shows an example of the input code for given variables in the model.

```
# Usage example
year_input = input("Enter Year: ")
temp_input = input("Enter Temperature: ")
rain_input = input("Enter Rainfall: ")
pesticide_input = input("Enter Pesticide: ")

file_path = "ModelDataSet_test.csv"

# Call the function with the input values
input_values = [year_input, temp_input, rain_input, pesticide_input]
append_to_csv(input_values, file_path)
```

Figure 3. User Interaction code

5.2 Implications for decision makers

Accurate maize crop yield predictions have a significant importance for farmers' decisions. Timely and dependable forecasts enable efficient resource distribution and other crop management strategies. The model's anticipation can function as early warning system, enabling proactive measures to counter potential yield reductions and economic setbacks.

6. Conclusion

The design of a crop yield prediction model involves careful consideration of data collection, preprocessing, feature engineering, model selection, training, and evaluation. Developing models that can accurately predict crop yields can be achieved by using a structured approach of collecting comprehensive and high-quality data. Conducting thorough data preprocessing and feature engineering, experimenting with different model architectures, and regularly evaluating and refining the model can also produce accurate models. However, it is important to be aware of the limitations and challenges such as data quality and availability, the complexity of crop-yield relationships, generalization to different regions and crop types, and the dynamic nature of crop-yield relationships.

References

Aurélien Géron , *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc, 2019.

Ayoub Shaikh, T., Rasool, T. and Rasheed Lone, F. , 'Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming', *Computers and Electronics in Agriculture*, 198, p. 107119, 2022. Available at: https://doi.org/https://doi.org/10.1016/j.compag.2022.107119.

Chergui, N., Kechadi, M.-T. and McDonnell, M., 'The Impact of Data Analytics in Digital Agriculture: A Review', in *2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA)*, pp. 1–13, 2020. Available at: https://doi.org/10.1109/OCTA49274.2020.9151851

Deepak Jain. (2023, May 6). *Data Preprocessing in Data Mining*. https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/

FAO (2023) Zimbabwe at a glance, Available at: https://www.fao.org/zimbabwe/fao-in-zimbabwe/zimbabwe-at-a-glance/en/

Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S., Crop Yield Prediction using Machine Learning and Deep Learning Techniques. *Procedia Computer Science*, *218*, 406–417, 2023. https://doi.org/https://doi.org/10.1016/j.procs.2023.01.023

Ji, Z., Pan, Y. and Li, N., 'Integrating the temperature vegetation dryness index and meteorology parameters to dynamically predict crop yield with fixed date intervals using an integral regression model', *Ecological Modelling*, 455, p. 109651,2021. Available at: https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109651.

Kannengiesser, U. and Gero, J. , 'MODELLING THE DESIGN OF MODELS: AN EXAMPLE USING CRISP-DM', Proceedings of the Design Society, 3, pp. 2705–2714, 2023. Available at: https://doi.org/10.1017/pds.2023.271.

Kaptymer, B.L., Abdulkerim Ute, J. and Negeso Hule, M., 'Climate Smart Agriculture and Its Implementation Challenges in Africa', *Current Journal of Applied Science and Technology*, pp. 1–13, 2019. Available at: https://doi.org/10.9734/cjast/2019/v38i430371.

Keith D. Foote. (2022, August 3). *Machine Learning Algorithms*. https://www.dataversity.net/machine-learning-algorithms/

van Klompenburg, T., Kassahun, A. and Catal, C., 'Crop yield prediction using machine learning: A systematic literature review', *Computers and Electronics in Agriculture*, 177, p. 105709, 2020. Available at: https://doi.org/https://doi.org/10.1016/j.compag.2020.105709.

Mark SubbaRao, Jonas Jaegermeyer, Ian Jones, & Laurence Schuler. (2023, April 30). Impact of Climate Change on Global Agricultural Yields. https://svs.gsfc.nasa.gov/4974#section_credits

Marko Phiri , *Crop-monitoring satellites are giving farming advice to remote areas of Zimbabwe*, *Thomson Reuters Foundation,2020*. Available at: https://gca.org/crop-monitoring-satellites-are-giving-farming-advice-to-remote-areas-of-zimbabwe/ (Accessed: 7 February 2023).

Mhizha, T., Nyamadzawo, G. mba, E. and Gotosa, J. &Mutsamba, E. (2021) 'Modeling maize yield prediction using statistical models: A case of Gwanda District, Zimbabwe', *Heliyon*, 7(e06931.), p. 4.

Munyati, T., and Jirah, S., ' A comparative study of machine learning algorithms for maize yield prediction in Zimbabwe', *Agricultural and Forest Meteorology*, (108026), p. 290,2020.

Nick Hotz (2023) 'What is CRISP-DM?' Data Science Process Alliance. Available at: https://www.datascience-pm.com/crisp-dm-2/ (Accessed: 25 May 2023).

Nuthall, P. L., & Old, K. M., Intuition, the farmers' primary decision process. A review and analysis. Journal of Rural Studies, 58, 28–38,2018. https://doi.org/10.1016/j.jrurstud.2017.12.012

Nyahunda, L. and Tirivangasi, H.M. , 'Challenges faced by rural people in mitigating the effects of climate change in the Mazungunye communal lands, Zimbabwe', *Jàmbá Journal of Disaster Risk Studies*, 11(1),2019. Available at: https://doi.org/10.4102/jamba.v11i1.596.

Nyéki, A., & Neményi, M., Crop Yield Prediction in Precision Agriculture. Agronomy, 12(10), 2460, 2022. https://doi.org/10.3390/agronomy12102460

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. , Machine learning for large-scale crop yield forecasting. Agricultural Systems, 187, 103016, 2021. https://doi.org/10.1016/j.agsy.2020.103016

Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. , Interpretability of deep learning models for crop yield forecasting. Computers and Electronics in Agriculture, 206, 107663, 2023. https://doi.org/10.1016/j.compag.2023.107663

Renuka and Terdal, Dr.S. , 'Evaluation of Machine Learning Algorithms for Crop Yield Prediction', *International Journal of Engineering and Advanced Technology*, 8(6), pp. 4082–4086, 2019. Available at: https://doi.org/10.35940/ijeat.F8640.088619.

Saltz, J.S. 'CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps', in 2021 IEEE International Conference on Big Data (Big Data), pp. 2337–2344, 2021. Available at: https://doi.org/10.1109/BigData52589.2021.9671634.

Sazib, N., Mladenova, Iliana E., & Bolten, J. D. , Assessing the Impact of ENSO on Agriculture Over Africa Using Earth Observation Data. Frontiers in Sustainable Food Systems, 4 2020. https://doi.org/10.3389/fsufs.2020.509914

Sarker, I.H. , 'Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective', *SN Computer Science*, 2(5), p. 377,2021. Available at: https://doi.org/10.1007/s42979-021-00765-8.

Smart Vision (2023) *What is the CRISP-DM methodology?* Available at: https://www.sv-europe.com/crisp-dm-methodology/ (Accessed: 25 May 2023).

Think Insights (2018) *CRISP-DM – A Framework For Data Mining And Analysis. Retrieved from https://thinkinsights.net/data/crisp-dm/.* Available at: https://thinkinsights.net/data/crisp-dm/. (Accessed: 20 May 2023).

Biographies

Thobekile Chigwada holds a Bachelor of Science Honors degree in Computer Science. She is a Master of Science student in Big Data Science at the National University of Science and Technology, Zimbabwe. She is a distinguished educator and accomplished industry professional. With a commitment to education, her students have achieved awards in IGCSE Ordinary Level and CIE Advanced Level Cambridge examinations in Computer Science. Her research initiative centers on developing a machine learning-based maize crop yield prediction model. By bridging education and agriculture, she aims to contribute transformative insights for sustainable practices.

Mary Dzinomwa holds an MSc in Information Systems and a BSc in Computer Science. She is a seasoned academic. Her research interest is in the fields of Data Analytics, Health Informatics, ICT4D, and 4IR.

Belinda Mutunhu Ndlovu is a Ph.D. in Information Systems student at UNISA. She holds an MSc in Information Systems and a BSc in Computer Science. She is a seasoned software developer and academic. She has published several papers in the fields of Data Analytics, Health Informatics, ICT4D, and 4IR.

Khulekani Sibanda is a PhD candidate at the University of Johannesburg. He obtained his Master of Science in Computer Science in 2011 and his Bachelor of Science in Computer Science in 2008 from the National University of Science and Technology. He is currently employed as a lecturer at NUST. His current research interests are in Blockchain adoption in Healthcare, Data Science and ICT4D.

Sibonile Moyo received an M.Sc. degree in Computer Science from the National University of Science and Technology, Bulawayo, Zimbabwe, in 2005, and a PhD in Computer Science from the University of South Africa in 2020. She is currently a Senior Lecturer in the Department of Informatics and Analytics, at the National University of Science and Technology (NUST), Zimbabwe. She has taught undergraduate courses in software engineering, software development, and database systems at NUST for the past 18 years. Her research interests are in software engineering, software development methodologies, and developing quality software. Dr. Moyo has supervised a number of undergraduate and postgraduate students in the area.