iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VIII **Month of publication:** August 2026

DOI: https://doi.org/10.22214/ijraset.2026.84572

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

Maize Yield and Disease Risk Prediction Using Climatic and Soil Data: A Machine Learning Approach for Lesotho

Masani Selialia[1], Dr. M. Ramjee[2]
[1]*Dept. of Information Technology and Computer Applications, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India*
[2]*Dept. of CS&SE, Dept. of IT&CA, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India*

Abstract: The main staple crop in Lesotho (a country in southern Africa) is maize, but its production is extremely susceptible to unpredictable rainfall, temperature fluctuations, deficiencies in soil moisture, and frequent foliar disease pressure. As a result, smallholder farming households have ongoing yield shortfalls and food insecurity. The design, implementation, and assessment of a web-based decision-support system that uses supervised machine learning to forecast maize production and categorize disease risk using soil and climate data are presented in this study. Eight environmental and agronomic predictors: rainfall, temperature, relative humidity, soil moisture, cultivated area, week number, year, and administrative district, were used to independently train two Ordinary Least Squares (OLS) linear regression models on historical agricultural and environmental data spanning 40 years. Because of its interpretability, computational efficiency, and adaptability for implementation in environments with limited resources, linear regression was chosen. The system supports both single-record and bulk CSV prediction workflows and was built using a Flask (Python) backend, a Bootstrap 5 responsive frontend, a SQLite database for permanent storage, and Chart.js for dashboard analytics. The illness risk model had an R^2 of 0.76 and an RMSE of 0.10 across three interpretable risk categories (low, medium, and high), but the yield prediction model obtained a coefficient of determination (R^2) of 0.88 and an RMSE of 2.09. Functional testing verified dependable performance in dashboard, storage, reporting, and prediction processes. The findings show that in data-poor agricultural settings like Lesotho, interpretable machine learning models may provide affordable, easily accessible decision assistance for early disease risk awareness and maize production planning.

Keywords: machine learning, maize yield prediction, disease risk classification, linear regression, decision support system, precision agriculture, Lesotho.

I. INTRODUCTION

Lesotho's economy and food security are based on agriculture, and maize, referred to locally as "poone," is the nation's main staple crop and the main source of energy for most rural residents [1]. Due to the unique agro-ecological limits imposed by Lesotho's high-altitude, landlocked physiography, maize is grown nearly exclusively under rain-fed smallholder settings on soils that are constantly deteriorating. Only a small portion of domestic maize needs are satisfied by domestic production, and average national maize yields are consistently low. This leads to structural food insecurity, which is occasionally exacerbated by severe, climate-related shocks [2] Production is further threatened by foliar fungal diseases such as northern corn leaf blight and gray leaf spot, but Lesotho's agricultural industry presently lacks a forecasting instrument that can measure or predict disease pressure at scale. Agricultural planning and early-warning decisions are primarily dependent on expert judgment, generalized regional estimates, and retrospective survey data in the absence of locally calibrated, data-driven forecasting tools. This reactive information paradigm restricts policymakers' and farmers' ability to react proactively to climate variability and disease risk. In order to close that gap, this research presents a web-based method for predicting maize yield and disease risk that was created especially for the agroclimatic conditions of Lesotho, circumstances, and context. Through a responsive online interface with database-backed historical analytics, the system generates (i) quantitative maize yield projections and (ii) categorical disease risk classifications by applying supervised linear regression to a common collection of meteorological and soil-related factors. This work's main contributions are (1) the creation and empirical validation of two interpretable regression models for predicting disease risk and maize yield using inputs pertinent to Lesotho's agro-ecological conditions; and (2) the design and implementation of a full-stack, low-resource decision-support platform that operationalizes these models for both batch (bulk) and individual agricultural forecasting.

II. RELATED WORK

Over 4.5 billion people in about 94 developing nations rely on maize as their main source of calories, making it one of the most frequently grown cereal crops in the world [1]. The majority of smallholder farms in the Southern African Development Community (SADC) region grow maize under rain-fed conditions, and average yields are still far below research station potential. This discrepancy is linked to soil nutrient depletion, unpredictable rainfall, heat stress, and disease pressure [3], [4].

Machine learning has been used to forecast agricultural productivity in an increasing amount of literature. A substantial research gap for smallholder systems in sub-Saharan Africa was noted by Van Klompenburg et al. [5], who conducted a systematic review of fifty machine-learning-based yield prediction studies and discovered that models combining multiple data types (climatic, soil, and remote sensing) consistently outperformed single-source models. Similar to the results obtained in this work using linear regression, Jeong et al. [6] used Random Forest to forecast maize and soybean production and reported R^2 values between 0.75 and 0.88. For county-level yield prediction, Cai et al. [7] compared OLS, Lasso regression, and Random Forest. They discovered that regularized linear methods provided a favorable balance of interpretability and accuracy, especially under small sample sizes—a factor directly related to Lesotho's data-scarce environment.

In terms of disease, Ward, Nutter, and Esker [8] established a methodological precedent for climate-based disease risk prediction by creating regression-based models for gray leaf spot risk in the U.S. Corn Belt using temperature and humidity data. While Mahlein et al. [10] examined the integration of environmental risk modeling with field-level detection for early warning systems, an approach conceptually similar to the current work, Battilani et al. [9] used ensemble regression to forecast aflatoxin contamination risk in maize under current and projected climate scenarios.

Despite this growing amount of research, no published machine learning system has been created and verified, especially for Lesotho's maize industry, nor has any system integrated yield and disease risk prediction into a single web-based, deployable decision-support tool that is calibrated to a common soil feature space and climate for the nation. That gap is filled by this study.

III. MATERIALS AND METHODS

A. Study Area and Data

The study focuses on the lowland agro-ecological zone of Lesotho, a landlocked, high-altitude country in southern Africa, where maize is grown. The training and validation dataset used for model development was created by compiling and preprocessing historical agricultural and environmental observations spanning 40 years.

B. Input Variables

Based on their proven relevance to maize production and disease dynamics in the agroclimatic literature, eight environmental and agronomic variables, summarized in Table II, were chosen as model inputs: rainfall, temperature, relative humidity, soil moisture, cultivated area, week number, year, and administrative district. The yield and illness risk models share a feature space since they both employ the same predictor set.

Table II. Input Variables For The Prediction Models

Variable	Description / Unit
Rainfall	Cumulative rainfall (mm)
Temperature	Mean ambient air temperature (°C)
Humidity	Average relative humidity (%)
Soil Moisture	Volumetric soil moisture content (%)
Cultivation Area	Land area under maize cultivation (ha)
Week Number	Agricultural week within the season (1–52)
Year	Calendar year of observation
District	Administrative district within Lesotho

C. Model Development
A yield prediction model that estimated maize grain yield and a disease risk prediction model that estimated a continuous disease risk score were both independently trained using Ordinary Least Squares (OLS) multiple linear regression. Because of its interpretability, low computational overhead, and proven suitability for continuous numerical prediction in deployment settings with limited data and computing resources, qualities that align with Lesotho's agricultural information infrastructure, linear regression was chosen over more complicated algorithms. To make the disease risk score generated by the second model easier to understand for non-technical end users like farmers and extension workers, it is standardized to a percentage and mapped to one of three interpretable categories: low, medium, or high risk.

D. Evaluation Metrics
The Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) were used to assess the model's performance. While RMSE represents the average size of prediction error in the target variable's original units, R^2 measures the percentage of variation in the target variable that the model explains. Regression-based agricultural prediction models are frequently rigorously evaluated using these two complementary criteria [6], [11].

IV. SYSTEM ARCHITECTURE AND IMPLEMENTATION
A full-stack web application is used to operationalize the prediction models. The Flask web framework, which was selected for its lightweight footprint and flexibility for implementing machine learning inference behind a REST-style interface, was used to develop the server-side application functionality in Python. HTML5, CSS3, Bootstrap 5, and JavaScript were used in the frontend's construction to create a responsive experience that works on desktop, tablet, and mobile devices. significant factor to take into account because, in rural Lesotho, mobile internet connection is frequently the main means of connectivity. Due to its zero-configuration deployment and adaptability for single-server applications in low-resource environments, SQLite was chosen as the persistent data store for prediction inputs, outputs, and metadata.

Figure 1. System Architecture of the Lesotho Maize Prediction Platform

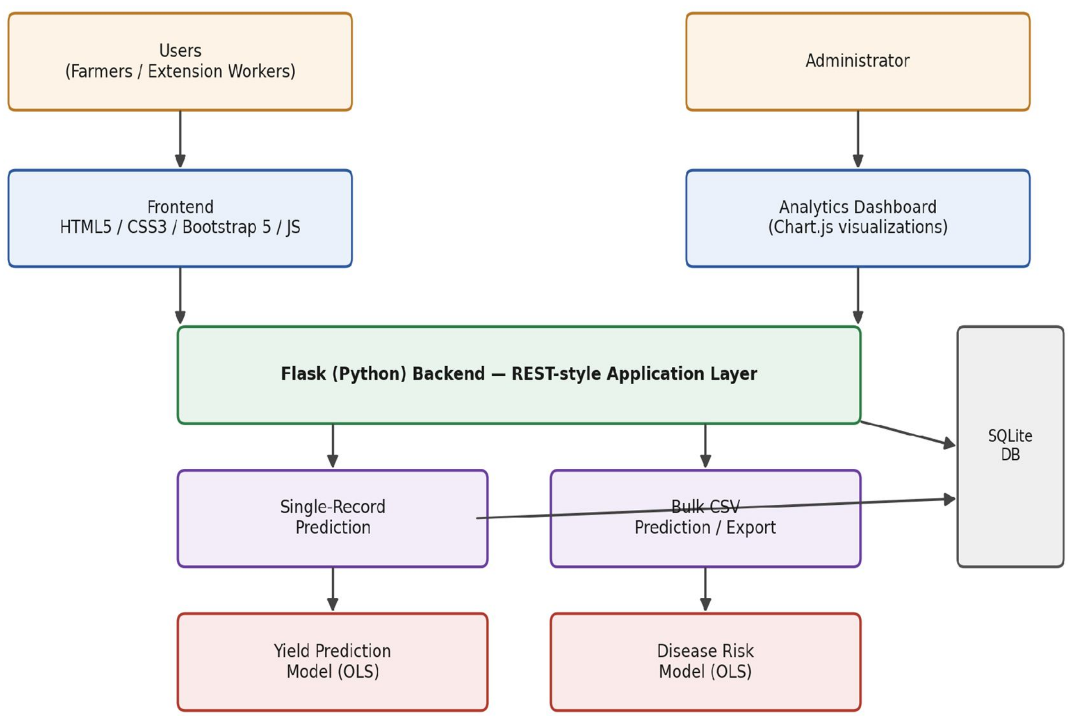


Figure 1. System architecture of the Lesotho maize prediction platform, showing the Flask backend, SQLite persistence layer, and the single-record and bulk-CSV prediction pathways feeding the two OLS models.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

Two prediction modes are supported by the system. In the single-record mode, an immediate maize yield estimate and disease risk categorization are obtained by manually entering the eight input variables. When using the bulk mode, users submit a structured CSV file with several agricultural observations. The system verifies the file's completeness and structure, applies both models to each record, and provides a downloaded CSV report with all of the forecast results. To facilitate further historical research, all predictions, whether single or bulk, are saved in the SQLite database.

In addition to tabular record views, filtering, and CSV export for offline study, an administrator dashboard created using Chart.js offers weekly, annual, and district-level visualizations of expected yield and disease risk. Farmers and extension users can view the prediction interface without logging in, but only authorized administrators have access to the dashboard.

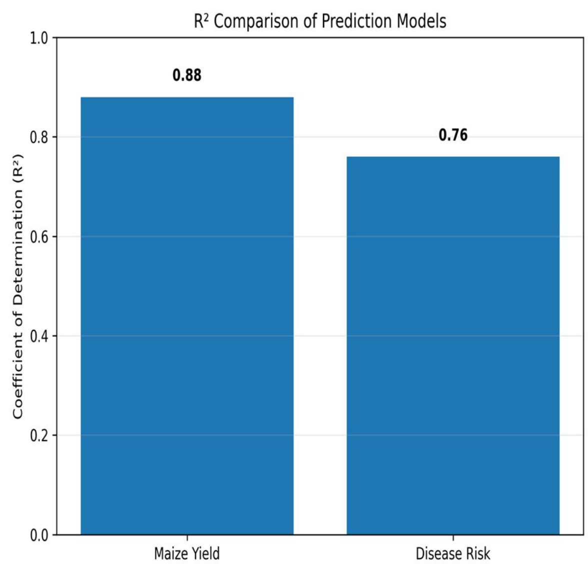
V. RESULTS AND DISCUSSION

A. Model Performance

The two trained models' performance is summed up in Table I. With a low average prediction error in relation to the target variable's scale, the yield prediction model's R^2 of 0.88 and RMSE of 2.09 show that the chosen predictors account for around 88% of the variance in observed maize yield. R^2 values above 0.80 are often considered suggestive of a well-fitting regression model [6], [12], and this finding is in line with performance levels reported for maize yield prediction elsewhere in the literature.

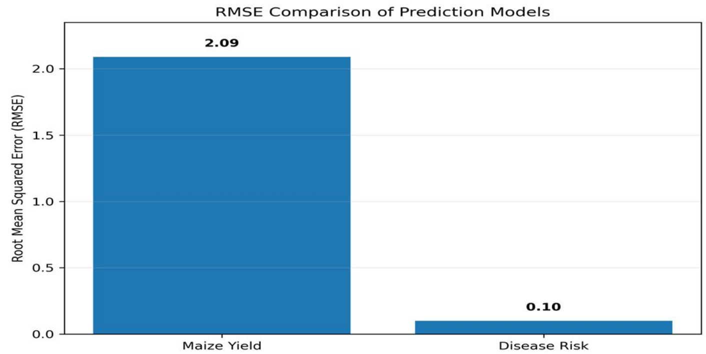
Table I. Comparative Performance Metrics, Yield, And Disease Risk Models

Evaluation Metric	Yield Model	Disease Risk Model
RMSE	2.09	0.10
R^2	0.88	0.76
Algorithm	OLS Linear Regression	OLS Linear Regression
Feature Space	Rainfall, Temp., Humidity, Soil Moisture, Area, Week, Year, District	Same as yield model



An R^2 of 0.76 and an RMSE of 0.10 were attained by the illness risk prediction model. Given the inherent complexity of plant disease dynamics, disease incidence is influenced by additional factors like pathogen inoculum load, cultivar susceptibility, and agronomic management practices that were not captured in the available training data and that represent a natural direction for future model refinement. This result, while lower than the yield model's performance, represents an acceptable level of accuracy.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com



B. System Functional Evaluation

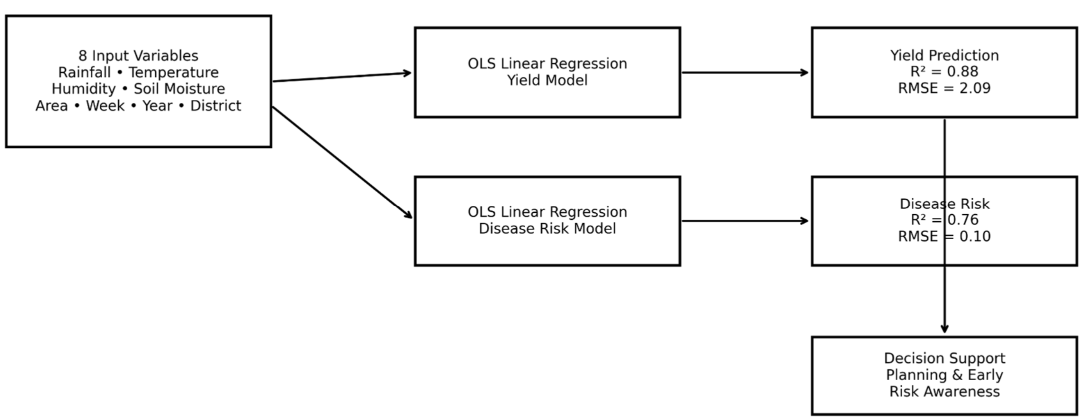
All the main system processes were functionally tested, such as single-record prediction, bulk CSV prediction, database persistence, dashboard analytics, CSV export, and user authentication. Every process that was examined was successful, with dashboard visuals reliably changing in response to new prediction records and predictions and related metadata being correctly recorded and retrieved from the SQLite database. The responsive design solution maintained functional completeness and visual coherence across device types, as confirmed by further validation of the interface across desktop, tablet, and mobile screen configurations.

C. Discussion

The findings show that in an agricultural setting with limited data, an interpretable linear regression framework in conjunction with a lightweight and accessible online architecture may provide valuable decision-support insights for disease risk awareness and maize production planning. In keeping with the theoretical presumptions underlying OLS regression, the robust yield model performance indicates that the connection between the chosen meteorological, soil, and agronomic variables and maize yield is roughly linear over the observed range of circumstances in Lesotho. The higher intrinsic complexity of illness dynamics in comparison to yield creation is reflected in the disease risk model's relatively poorer but nonetheless adequate performance.

There are a few restrictions to be aware of. First, deficiencies in agroclimatic coverage may induce systematic bias, and the accuracy of the model is limited by the completeness and representativeness of the underlying 40-year historical dataset. Second, threshold effects under heat or moisture stress are examples of non-linear feature interactions that may be agronomically significant but are not captured by linear regression. Third, instead of using real-time sensor or satellite-derived inputs, the current system uses manually entered environmental data, which adds delay and the possibility of data-entry mistakes.

Overall Results and Decision-Support Framework



VI. CONCLUSION AND FUTURE WORK

A web-based system for predicting maize yield and disease risk for Lesotho was designed, implemented, and evaluated in this work. It is based on two interpretable OLS linear regression models that use a common set of eight meteorological, soil, and agronomic factors. Using a tri-level risk categorization system, the illness risk model obtained good accuracy ($R^2 = 0.76$, RMSE = 0.10), while the yield model produced excellent predictive accuracy ($R^2 = 0.88$, RMSE = 2.09). Functional testing verified dependable functioning across all assessed processes. The system was constructed as a full-stack Flask, SQLite, and Bootstrap 5 application enabling both single-record and bulk CSV prediction, supported by an interactive Chart.js analytics dashboard.

The findings show that accessible, computationally light machine learning models can provide actionable predictive insights without requiring substantial computational resources or specialized end-user expertise when combined with appropriate web and database infrastructure. This directly addresses the needs of food security and agricultural productivity monitoring in a data-scarce environment like Lesotho.

Future research will concentrate on (i) assessing ensemble and non-linear methods, such as Random Forest, Gradient Boosting, and Artificial Neural Networks, to capture non-linear predictor interactions; (ii) integrating real-time
meteorological data through weather APIs to eliminate manual data on additional staple crops relevant to Lesotho's agricultural sector, such as wheat and sorghum entry; (iv) creating a companion mobile application to enhance accessibility for smallholder farmers in rural regions using spatially explicit Geographic Information System (GIS) capabilities; and (v) expanding the framework.

REFERENCES

[1] CIMMYT, Maize in Sub-Saharan Africa: Opportunities and Challenges. Nairobi: International Maize and Wheat Improvement Center, 2019.
[2] J. Verschuur, S. Li, P. Wolski, and F. E. L. Otto, "Climate change as a driver of food insecurity in the 2007 Lesotho–South Africa drought," Scientific Reports, vol. 11, p. 3032, 2021.
[3] J. E. Cairns et al., "Maize production in a changing climate: Impacts, adaptation, and mitigation strategies," Advances in Agronomy, vol. 114, pp. 1–58, 2013.
[4] S. Zingore, H. K. Murwira, R. J. Delve, and K. E. Giller, "Soil type, management history and current resource allocation: Three dimensions regulating variability of maize yields and nutrient use efficiencies on smallholder farms in Zimbabwe," Field Crops Research, vol. 101, no. 3, pp. 296–305, 2011.
[5] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 177, p. 105709, 2020.
[6] J. H. Jeong et al., "Random forests for global and regional crop yield predictions," PLOS ONE, vol. 11, no. 6, p. e0156571, 2016.
[7] Y. Cai et al., "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," Agricultural and Forest Meteorology, vol. 274, pp. 144–159, 2019.
[8] J. M. J. Ward, F. W. Nutter, and P. D. Esker, "Gray leaf spot of corn: A review," Plant Disease, vol. 96, no. 10, pp. 1369–1382, 2012.
[9] P. Battilani et al., "Aflatoxin B1 contamination in maize in Europe increases due to climate change," Scientific Reports, vol. 6, p. 24328, 2016.
[10] A.-K. Mahlein, T. Rumpf, P. Welke, H.-W. Dählein, L. Plümer, U. Steiner, and E.-C. Oerke, "Development of spectral indices for detecting and identifying plant diseases via hyperspectral imaging," Computers and Electronics in Agriculture, vol. 121, pp. 57–65, 2016.
[11] D. K. Ray, J. S. Gerber, G. K. MacDonald, and P. C. West, "Climate variation explains a third of global crop yield variability," Nature Communications, vol. 6, p. 5989, 2015.
[12] K. Kuwata and R. Shibasaki, "Estimating crop yields with deep learning and remotely sensed data," in Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), 2015, pp. 858–861.
[13] P. Roudier, B. Sultan, P. Quirion, and A. Berg, "The impact of future climate change on West African crop yields: What does the recent literature say?," Global Environmental Change, vol. 21, no. 3, pp. 1073–1083, 2011.