JOURNAL
Of Universal Applied Research

ISSN (Online): XXX
Vol. 01, Issue 02 (2026)
https://universalappliedresearch.com/index.php/JUAR/issue/archive

A Digital Decision-Support Tool for Climate-Resilient Crop Planning: Design, Validation and Policy Implications

M. Golam Mahboob

Natural Resources Management Division, Bangladesh Agricultural Research Council (BARC), Dhaka, Bangladesh

Received:- 09/06/2026, Revised:- 10/07/2026, Accepted:- 17/07/2026, Published:- 24/07/2026

Abstract

Soil sensitive crop selection demands decision tools that not only relate soil conditions to the shifting temperature, humidity and rainfall, but also do not translate into agronomic prescriptions that lack support from the algorithms. The digital planning prototype for crops was designed and internally validated in this study based on a public benchmark, which has 2200 records, 7 soil–climate predictors and 22 balanced crop classes. Dummy, multinomial logistic regression, decision tree, random forest and extra trees classifiers were tested using a stratified 80:20 holdout and a 5-fold cross validation. Random Forest achieved the best cross-validated macro-F1 (0.994 ± 0.005), a test accuracy of 0.993, test macro-F1 of 0.993 and top three accuracy of 1.000. The most important factors were humidity, rainfall and potassium. Recommendation stability was then quantified by illustrative warming and drying and humidity-reduction perturbations. Combined stress resulted in 86.18% of top-1 recommendations being retained, and a mean top-3 overlap of 57.79%. The resulting architecture offers ranked recommendations, empirical-range warnings and scenario sensitivity as opposed to deterministic advice. The study provides a replicable process from benchmark modelling to decision support for policy and sets out clear guidelines for geospatial, temporal, economic and field validation prior to implementation.

Keywords: Climate-Resilient Agriculture; Crop Recommendation; Decision-Support System; Machine Learning; Random Forest; Scenario Sensitivity.

1. Introduction

OClimate change has forced the decision-making of crop planning from a historically informed, seasonal one to a decision problem under a non-stationary environmental risk. The Intergovernmental Panel on Climate Change (IPCC) reports that there is a global warming trend, and that warming, changes in precipitation, compound extremes, and changes in pest and disease pressures already impact crop suitability, yield stability and food security, with varying effectiveness of adaptation across different regions and production systems [1]. Crop choice can therefore not be simply regarded as a farm to conventional crop calendar look-up. It demands more and more evidence chains that connect local soil conditions, climatic exposure, uncertainty and potential alternatives that are feasible and yet maintain the farmer and agronomist judgment.

Climate change impacts on productivity do not only pertain to the future. The econometric evidence shows that anthropogenic warming already has reduced global agricultural total factor productivity growth in a non-uniform manner with greater impacts in warmer regions [2]. This historical signal is relevant for the decision-support research, as an apparently well-trained model using favorable or carefully curated conditions may produce past suitability boundaries which fail to show emerging climate-limits. It is therefore important for a useful crop-planning tool to be able to show not only what it does with observations that are currently available but also whether a change in reasonable stress on the climate inputs produces a change in the recommendations.

Ensemble of latest-generation crop and climate-models also suggest that anthropogenic impacts could be discerned before 2040 in some of the key production areas, depending on crop and scenario [3]. The diversity of such heterogeneity makes a scientific label of “climate-resilient crop” difficult. Resilience is a function of the relationship between the crop, the environment and the management, and of the product of that relationship which is sought to be protected, e.g., yield, income, water use or failure risk. The present study therefore focuses on a more limited (and auditable) notion: model recommendation stability when perturbations are explicitly specified. Although this is not observed resilience, it serves as a clear sensitivity diagnostic for a prototype system.

Climate-smart agriculture aims to boost productivity, build resilience and lower emissions as much as possible, but in doing so, it is important to consider the process of transposing knowledge into decisions that institutions and producers can make [4]. A digital decision support can help with that translation by organizing the different indicators, by assessing all the options in a uniform manner and by communicating the different options using a ranking. However, such a disciplined system boundary is required to make it useful. A classification score does not provide a policy recommendation per se and deployment requires localized data, extension capacity, monitoring, accountability, and explicit dealing with uncertainty and out-of-distribution inputs.

Specifically, access and representation issue are key, as data-driven agriculture is not evenly-distributed. Climate-stressed regions and small farms may have less connectivity and data coverage than large commercial farms, which could lead to the misapplication of digital tools, as adaptive capacity is likely to be already high [5]. Therefore, it is important to consider technical performance in the context of coverage, affordability, language, infrastructure and institutional trust. This is a planning tool for public use, and one of the purposes of planning is to offer easy-to-understand options and cautionary statements, not to prescribe a single label when many are delivered at the end of the day.

Smart-farming research provides a glimpse of how soil, weather, sensor, and management data can be transformed into predictive information and uncovers remaining issues with platform control, interoperability, privacy, and data ownership [6]. In terms of crop planning, this duality means that the analytical pipeline and the governance model are inextricable: The data lineage, version of the model, ranges of the empirical data, and the logic behind the recommendations should be documented enough to stand alone for independent review. Therefore, here, the concept of reproducibility is considered as an integral part of the tool's design and not a research add-on.

Digital-agriculture scholarship has also demonstrated that the adoption and impact are determined by farmer skills, professional roles, knowledge networks, market organization, policy processes and responsible innovation, among others [7].

There are many machine-learning applications in agriculture, including crop, water, soil and livestock management, with methodologically diverse evidence. Liakos et al. found that problem domains dictate the choice of models and that there is no one-size-fits-all solution to the problem of data form and task definition [8]. This insight is directly applicable to crop-recommendation: tabular soil–climate data does not imply nonlinear ensemble: comparative baselines are needed to decide whether nonlinear ensembles are valuable versus simpler decision boundaries.

While significant advances have been made in deep learning for image-based agricultural applications, notably for disease detection and phenotyping, another challenge identified by Kamilaris and Prenafeta-Boldú is the lack of uniformity in the datasets, pre-processing options and evaluation protocols used in different studies [9]. Deep Learning can potentially be opaque and have a large number of parameters without a corresponding gain in evidence for a small structured dataset with seven predictors. Therefore, the focus of the present study will be on transparent tabular classifiers and complexity will be assessed empirically, not necessarily assuming that model depth leads to greater agriculture use.

As per the updated review by Benos et al., the implementation of ensemble learning, support-vector, neural networks and sensor linked analytics are growing in the domain of crop and resource management [10]. Simultaneously the review reveals common shortcomings, such as small number of data, site-specific sampling, variable metrics and lack of operational validation. This should therefore be accompanied by high accuracy on a balanced benchmark, class-specific behaviour, outputs that are sensitive to uncertainty and a clear explanation of what is not measured in the dataset.

A related but different problem of decision is the prediction of crop yield available in the literature. Van Klompenburg et al. observed that yield models are often based on the same variables, such as temperature, rainfall, soil properties, and remote-sensing variables, while neural network and ensemble methods are often used [11]. Yield prediction is a continuous outcome, while the current task has to choose between crop labels. This is important, as a recommendation label is an indication of similarity to the represented suitability profiles and is not an indication of expected yield, profit or risk. The validation evidence for policy claims should not be carried over from classification to those outcomes that are not observed.

Crane-Droesch was able to show that machine learning can capture nonlinear climate–yield relationships with a parametric structure that is required for impact assessment [12]. That paper demonstrates the potential and the challenge of climate-informed modelling: the same change in the perturbation of a variable can lead to context-dependent responses; and extrapolating beyond the realm of observed climate is risky. This present scenario analysis is thus interpreted as sensitivity testing on local level with respect to the recommendations, rather than as a climate projection or causal estimate of crop response.

Crop-recommendation studies have been conducted in the past which have tried to match crop requirement with crop plot conditions by considering soil properties and the environmental conditions. The precision-agriculture recommendation system by considering biotic and abiotic factors and the practicality of direct advisory-outputs were proposed by Pudumalar et al. [13]. But, early systems tended to give recommendations on points and performance of classifiers without considering the changes in ranked choices in the light of climate stress. Such instability should be apparent in a decision-support architecture since it is possible that alternate rankings may be of concern even if the top-one designation doesn't change.

The use of big data in precision agriculture has highlighted the importance of weather prediction, information and communication technology and distributed acquisition of data as the basis for making decisions in the right time in the farms [14]. Their benefit is that they can be updated with new recommendations as the situation changes, but this also brings with it infrastructure and quality control needs. An operational tool would need to include governed data ingestion, timestamped weather data sources, sensor calibration, data missingness and versioned recommendations – a static benchmark can prove analytical feasibility.

For instance, irrigation management and agrochemical management in agriculture has been studied using artificial intelligence, whose predictive capabilities can aid resources optimization, when coupled to feasible farm operations [15]. However, accuracy in models does not necessarily mean usefulness in operation—advice has to be based on constraints, and risks quantified; and judged against agronomic and economic results. In terms of crop selection, this should then be integrated with soil and climate classification, irrigation availability, input costs, rotations, market availability and farmer preferences. This study meets the need by accomplishing three objectives:

To design a reproducible digital decision-support workflow that converts soil nutrient, pH, temperature, humidity, and rainfall inputs into ranked crop recommendations with empirical-range warnings.

To compare five classification approaches using a prespecified stratified holdout and five-fold cross-validation, and to interpret the selected model through class-level errors and feature importance.

To quantify recommendation stability under isolated and combined climate-stress perturbations and derive bounded policy implications and external-validation requirements.

2. Methodology

2.1 Research Design and System Boundary

The study was based on a computational design-science methodology where a functional design-science prototype was developed and assessed based on a set of analytical design criteria. The unit of analysis was one tabular record, which represented a combination of soil and environment. One of the crop labels predicted was 22. Estimates of yield, profitability, water demand, emissions, adoption, and causal climate resilience were not included within the system boundary and were instead ranked model recommendations and scenario-sensitivity information. This was kept to ensure that the performance of the internal classification was not confounded with the effectiveness of the farm-level classification.

Input validation, data-quality assessment, descriptive analysis, stratified data partitioning, comparative model development, cross-validated selection, holdout evaluation, model interpretation, top-three recommendation generation, empirical-range checking and climate-stress perturbation were the workflow. A specific random seed was used (42) for reproducibility. The implementation used Python 3.12.13, pandas 2.2.3, NumPy 2.3.5, and scikit-learn 1.8.0. The outputs were exported thematically and programmatically in machine readable tables, figures (high resolution), a serialized model and metadata about the analytical configuration.

2.2 Data Source, Variables and Preprocessing

The Crop Recommendation Dataset was obtained from Kaggle [16]. It was composed of 2,200 observations and eight columns; seven of these columns were predictors (N, P, K, temp, RH, soil pH, and rainfall) and the eighth was a crop-label target. The target variable comprised 22 crop classes, and there were 100 observations in each class, which were balanced and used to develop and compare multiclass models. But because it lacks geographic, temporal, management and yield data, it can only be used for methodological development and internal validation, not for field-level agronomic inferences.

Modelling data was prepared by standardizing crop labels (space around label removed, text converted to lower case) and ensuring that all the predictor columns were numeric. The data was checked for missing data, duplicate data, and physically implausible data based on the ranges of the data. There were no missing cells, no duplicates and no broad range violations were detected; thus all 2200 observations were retained. Statistical outliers were reported but not removed as they may be valid crop specific soil or climatic conditions.

2.3 Decision-Support Architecture and Dataset Characteristics

The implemented decision-support architecture is presented in Figure 1 to clarify how the analytical components operate as an integrated workflow. It traces the progression from soil–climate data entry and validation to comparative model selection, ranked crop recommendations, climate-stress assessment, and policy-oriented safeguards.

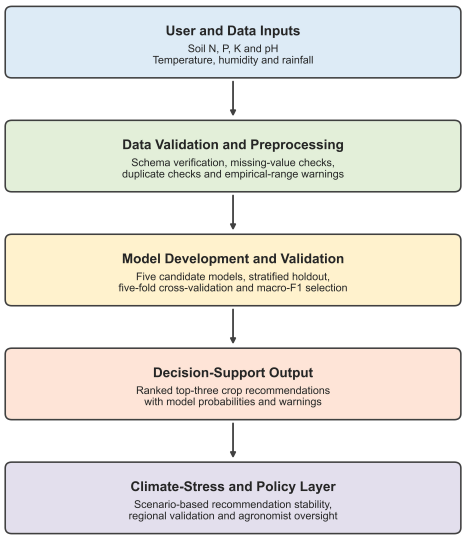


Figure 1. Architecture of the Digital Crop-Planning Decision-Support Tool

Figure 1 organizes the prototype into five linked layers: user and data inputs; validation and preprocessing; comparative model development; ranked decision output; and climate-stress and policy interpretation. The architecture is important because it prevents the classifier from functioning as an isolated prediction object. Range warnings and scenario analysis are positioned as integral assurance components, while regional validation and

agronomist oversight remain outside the automated decision. The numerical characteristics of the cleaned analytical sample are presented in Table 1.

Table 1. Descriptive Statistics for the Analytical Dataset

Predictor	N	Mean	SD	Minimum	Median	Maximum
Nitrogen (N)	2,200	50.552	36.917	0.000	37.000	140.000
Phosphorus (P)	2,200	53.363	32.986	5.000	51.000	145.000
Potassium (K)	2,200	48.149	50.648	5.000	32.000	205.000
Temperature (°C)	2,200	25.616	5.064	8.826	25.599	43.675
Humidity (%)	2,200	71.482	22.264	14.258	80.473	99.982
pH	2,200	6.469	0.774	3.505	6.425	9.935
Rainfall (mm)	2,200	103.464	54.958	20.211	94.868	298.560

Table 1 confirms substantial dispersion across the represented input space. Temperature ranged from 8.826 °C to 43.675 °C, humidity from 14.258% to 99.982%, pH from 3.505 to 9.935, and rainfall from 20.211 mm to 298.560 mm. Potassium displayed the largest relative spread among nutrient variables (mean 48.149; SD 50.648), reflecting distinct high-potassium profiles for crops such as apple and grapes. Because every crop contributed 100 observations, the dataset offered a controlled class-comparison environment but not a prevalence-weighted representation of actual planting conditions.

2.4 Candidate Models and Decision Output

Five classifiers were compared with each other. A most-frequent Dummy Classifier was used to measure the chance-level accuracy for the balanced multiclass design. Multinomial logistic regression (after standardization) was taken as a linear probabilistic baseline. A single nonlinear rule structure was obtained using a Decision Tree that was weighted by class. Random Forest and Extra Trees were built with 300 trees, class balanced weighting, parallel computation and the same random seed. Extra Trees adds more randomization to splits while Random Forest adds more randomization through bootstrap aggregation and random feature selection. For the purpose of hyperparameter tuning, it was purposefully kept to a minimum so as to not over-optimize on a small curated benchmark.

Each of the fitted models generated class probabilities for the input vector. The decision interface sorted these probabilities and returned the three best ranked crops, their rankings and percentages of probabilities. Prior to the inference, each input was compared with the min and max of all predictors in the training partition. Anything that was not within this empirical envelope gave a warning, but the model was still able to produce a result. This design was able to distinguish between prediction and assurance, such that a high score did not overwhelm evidence that a query was outside of the represented feature space.

2.5 Validation Strategy and Performance Measures

The data set was split into 80% training (n = 1760) and 20% test (n = 440) data per crop, via stratification. Candidates were selected by the average macro-F1 of five-fold stratified cross-validation, which was only done in the training partition. The model which achieved the highest cross-validated macro-F1 was finally selected as the estimator; the test partition was not used in the model selection. This is the reason that Random Forest was chosen despite the marginal gain in test score for Extra Trees that used just a single split.

The metrics used to evaluate were accuracy, balanced accuracy, macro-precision, macro-recall, macro-F1 score and top -3 accuracy. Although this benchmark was balanced, macro averaging (equal weight to each crop) was suitable as it had been used in the evaluation of the previous year. Standard deviation was summarized as fold-to-fold internal stability. Class specific errors were found in a row-normalized confusion matrix. Native impurity-based importance was considered as a model interpretation, but not as a causal ranking, as it may be influenced by split opportunities and by correlated predictors, for selected tree ensemble.

2.6 Climate-Stress Analysis and Responsible-Use Protocol

All predictor vectors for each country (2200) were subjected to four illustrative perturbations: temperature +2 °C; rainfall −20 %; humidity −5 percentage points; and a combined scenario applying these perturbations; soil variables

were not included in the perturbations. Climate variables were only clipped when physically required: rainfall was not allowed to be negative and humidity was clipped to 0%–100%. These scenarios were not associated with a place, emissions trajectory, forecast horizon or crop-growth stage and were thus not presented as climate-projections.

Two measures of sensitivity have been calculated, both relative to the model predictions under baseline conditions. Top-one retention was the percentage of records which had the same top ranked crop. Mean top three overlap was the average of the overlap of the top three recommendations from the baseline and a scenario. Crop-level stability was calculated by averaging the retention over the four stress scenarios and grouping the records by the model recommended crop (baseline). This score assessed the robustness of the recommendation in the modeled classifier, rather than providing a measure of observed biological tolerance, yield stability or economic resilience. No animals, identifiable farm records, human participants or intervention were involved in the analysis and no institutional human subject review was applicable as the analysis was done using an openly available non-personal benchmark. Safeguards regarding responsible use were however still required.

3. Results and Discussion

3.1 Comparative Model Performance

The five candidate classifiers were evaluated under the same stratified partitioning and cross-validation procedure to ensure a consistent comparison. Table 2 reports their cross-validated accuracy and macro-F1, fold-to-fold variability, holdout performance, and top-three accuracy, together with the performance of the non-informative baseline.

Table 2. Comparative Internal-Validation Performance

Model	CV Acc.	CV F1	CV F1 SD	Test Acc.	Test F1	Top-3 Acc.
Random Forest	0.9943	0.9943	0.0048	0.9932	0.9932	1.0000
Extra Trees	0.9926	0.9926	0.0039	0.9955	0.9955	1.0000
Decision Tree	0.9852	0.9852	0.0068	0.9795	0.9794	0.9818
Multinomial logistic regression	0.9682	0.9678	0.0068	0.9727	0.9725	1.0000
Dummy baseline	0.0455	0.0040	0.0000	0.0455	0.0040	0.1364

The Dummy Baseline achieved only 0.0455 accuracy and 0.0040 macro-F1, consistent with one correct class among 22 balanced labels. All trained classifiers substantially exceeded that reference. Random Forest produced the highest cross-validated macro-F1 (0.9943, SD 0.0048) and was selected under the prespecified rule. Its test accuracy and macro-F1 were both 0.9932, while top-three accuracy reached 1.0000. Extra Trees recorded the highest single-holdout accuracy (0.9955), but its cross-validated macro-F1 (0.9926) was lower than that of Random Forest; selecting Extra Trees on the test result would therefore have violated the separation between model choice and final evaluation.The narrow performance differences among trained models are visualized in Figure 2 using an expanded vertical scale.

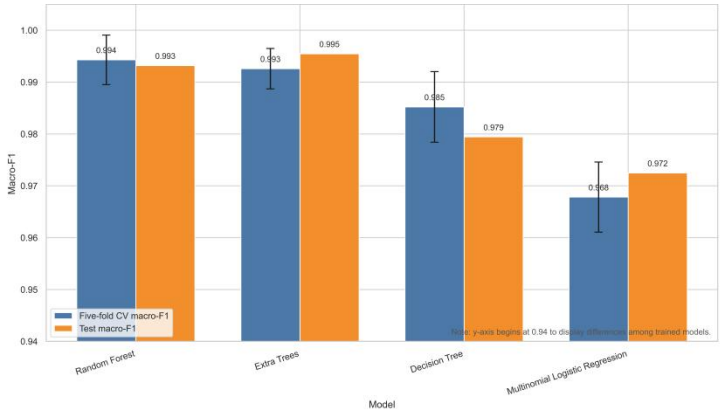


Figure 2. Predictive Performance of Candidate Models

Figure 2 shows that both ensemble models outperformed the single Decision Tree and multinomial logistic regression in cross-validated macro-F1. The small gap between cross-validation and test macro-F1 for Random Forest indicates stable internal generalization under the random stratified design. Nevertheless, the clustering of ensemble scores near one is also consistent with strong class separability in a curated dataset, so it cannot be treated as evidence of equally high accuracy in a new region or year.

3.2 Class-Level Error Structure

Overall accuracy can conceal errors affecting individual crop classes, particularly in a balanced multiclass setting. Figure 3 therefore presents the normalized confusion matrix for the selected Random Forest, enabling the correctly classified crops and the specific patterns of misclassification within the holdout sample to be identified.

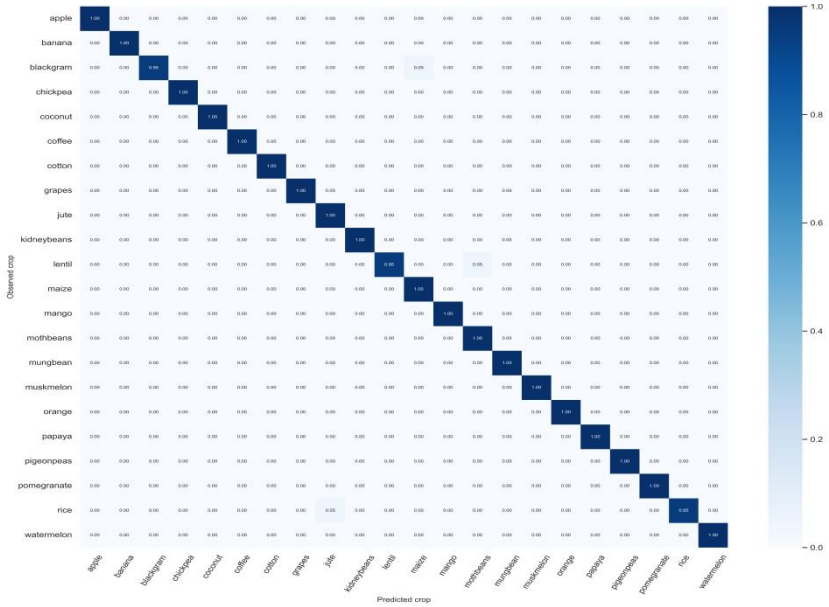


Figure 3. Normalize

Figure 3 indicates that 19 of 22 crops were classified perfectly in the 20-observation test subsets. One black-gram record was predicted as maize, one lentil record as moth beans, and one rice record as jute. These three errors account for the overall test accuracy of 437/440 (0.9932). The errors are agronomically plausible within the represented feature space because each confused pair shares portions of nutrient or climate profiles. Their concentration also shows why aggregate accuracy should be accompanied by class-level review, especially when a recommendation could influence input purchases or seasonal land allocation.

3.3 Predictor Contribution

To examine how the selected Random Forest used the available soil and climatic information, the relative contribution of each predictor was assessed. Table 3 reports the impurity-based feature-importance scores, providing an initial indication of which variables contributed most strongly to class separation within the fitted model.

Table 3. Random-Forest Impurity-Based Feature Importance

Rank	Predictor	Importance
1	Humidity	0.2177
2	Rainfall	0.2173
3	Potassium (K)	0.1763
4	Phosphorus (P)	0.1492
5	Nitrogen (N)	0.1067
6	Temperature	0.0762

7	pH	0.0568

Table 3. Random-Forest Impurity-Based Feature Importance

Humidity (0.2177) and rainfall (0.2173) contributed almost equally and together accounted for 43.50% of total model importance. Potassium ranked third (0.1763), followed by phosphorus (0.1492), nitrogen (0.1067), temperature (0.0762), and pH (0.0568). This ordering supports the inclusion of both climate and soil dimensions: the model did not reduce crop separation to weather alone, although climate-water variables dominated the fitted splits.The ranked contributions are visualized in Figure 4 to make the model's decision structure more interpretable.

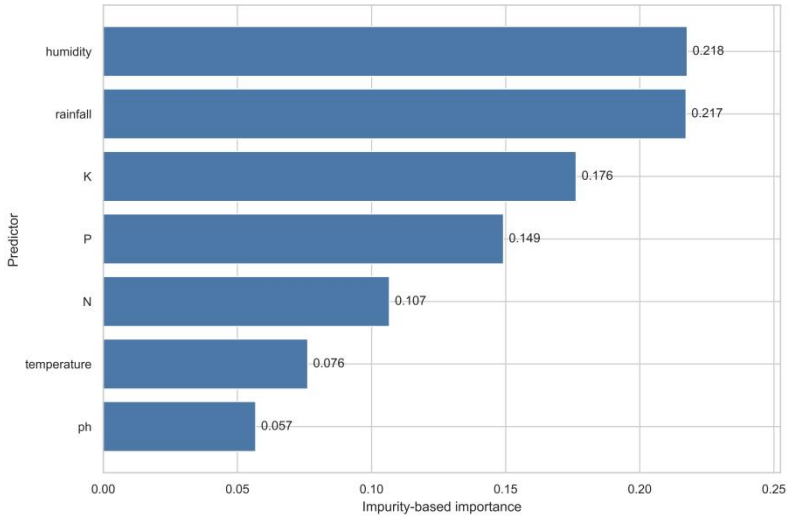


Figure 4. Predictor Importance in the Selected Random Forest Model

Figure 4 emphasizes a relatively distributed importance profile rather than dependence on a single predictor. The near-equality of humidity and rainfall suggests that water-related climate conditions are central to class separation in this benchmark, while the combined N–P–K contribution remains substantial. Importance values describe how the Random Forest partitioned these data; they should not be interpreted as causal agronomic effects or as evidence that pH and temperature are unimportant outside the benchmark.

3.4 Recommendation Stability Under Climate Stress

Predictive accuracy under baseline conditions does not indicate whether recommendations remain stable when climatic inputs change. Table 4 therefore summarizes the variation in top-one recommendations and top-three sets under isolated warming, reduced rainfall, lower humidity, and the combined climate-stress perturbation.

Table 4. Recommendation Sensitivity Under Illustrative Climate-Stress Scenarios

Scenario	Top-1 retention (%)	Top-3 overlap (%)	Changed	N
Baseline	100.00	100.00	0	2,200
Warming +2 °C	99.82	73.44	4	2,200
Rainfall −20%	94.77	69.59	115	2,200
Humidity −5 pp	95.64	74.95	96	2,200
Combined stress	86.18	57.79	304	2,200

Isolated warming produced only four top-one changes and retained 99.82% of baseline labels, although mean top-three overlap fell to 73.44%, revealing reordering beneath the leading recommendation. Rainfall reduction changed 115 top-one recommendations, and lower humidity changed 96. Combined stress produced 304 changes, reducing

top-one retention to 86.18% and top-three overlap to 57.79%. The difference between top-one retention and top-three overlap is substantively important: a stable first choice can mask considerable movement among alternatives. Crop-specific stability under the combined and averaged stress conditions is presented in Table 5.

Table 5. Crop-Level Model Recommendation Stability

Rank	Crop	Baseline n	Combined retention (%)	Mean stability (%)
1	Banana	100	100.00	100.00
2	Chickpea	100	100.00	100.00
3	Coffee	100	100.00	100.00
4	Moth beans	101	100.00	100.00
5	Kidney beans	100	100.00	100.00
6	Grapes	100	100.00	100.00
7	Watermelon	100	100.00	100.00
8	Orange	100	100.00	100.00
9	Muskmelon	100	100.00	100.00
10	Maize	101	100.00	100.00
11	Pomegranate	100	100.00	100.00
12	Mung bean	100	100.00	100.00
13	Mango	100	98.00	99.00
14	Papaya	100	98.00	99.00
15	Coconut	100	89.00	96.50
16	Pigeon peas	100	90.00	95.25
17	Lentil	99	84.85	94.70
18	Jute	101	74.26	91.34
19	Cotton	100	72.00	86.00
20	Apple	100	41.00	77.00
21	Rice	99	37.37	73.48
22	Black gram	99	10.10	57.32

The stability score equalled 100% for 12 baseline-recommended crops, while mango and papaya each retained 98% under combined stress. The lowest combined-stress retention occurred for black gram (10.10%), rice (37.37%), and apple (41.00%); their corresponding mean stability scores were 57.32%, 73.48%, and 77.00%. Baseline record counts differ slightly from 100 for six labels because grouping was performed on baseline model recommendations rather than ground-truth classes. These results identify where the model's rankings are sensitive; they do not establish that

high-scoring crops are biologically more resilient. The contrast between overall robustness and crop-level heterogeneity is visualized in Figure 5.

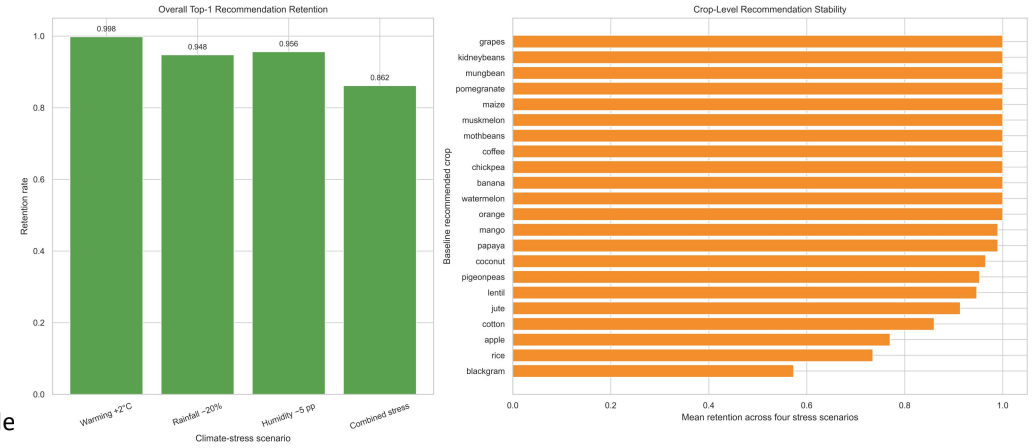
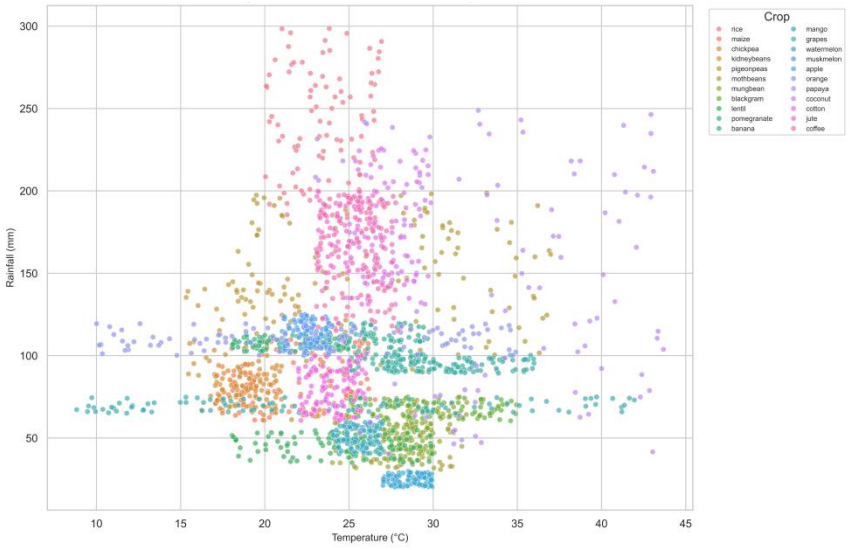


Figure 5. Mode

Figure 5 shows a monotonic reduction in overall top-one retention as stressors are combined and a markedly wider crop-level distribution. The right panel prevents the aggregate 86.18% retention value from obscuring concentrated instability in black gram, rice, and apple. This diagnostic is directly relevant to decision-support design: crops with unstable recommendations should trigger stronger caution, additional local evidence, or human review rather than an unqualified automated output.

3.5 Joint Temperature–Rainfall Structure of Crop Classes

Temperature and rainfall were examined jointly to determine whether the crop classes occupy clearly separated climatic regions or overlapping environmental envelopes. Figure 6 plots all observations by crop class, providing a descriptive view of climatic differentiation, within-class dispersion, and overlap across the benchmark feature space.



Figure

Figure 6 illustrates t considerably distinct but also ofte (mean 236.2 mm, range betwee 5.9 °C), and muskmelon is distri 7.0–29.9 °C), whereas some classes such as maize, chickpea, cotton, kidney beans and orange overlap at each other at intermediate climatic range. This is to explain why this model needs information on humidity, pH and N–P–K as well as temperature and rainfall to carry out reliable class separation.

The plot characterizes the benchmark feature space and is not a derivation of causal crop tolerance, geographic suitability or field validated climate resilience.

The chosen ensemble had an outstanding internal discrimination; however, interpretation of the score starts with the decision context. The income, water allocation, food security and exposure to seasonal failure are consequences of this crop choice and make transparency of reasoning and contestability important. Although Random Forest is not inherently an interpretable model, Rudin's claim that it is desirable to use interpretable models in high-stakes settings is relevant here. Partial view of feature importance and ranked probabilities is fine, but a deployed system should consider simpler rule-based approaches, show the local decision paths, and let the agronomist overrule or reject a recommendation.

The validation design prevented that the test partition be selected in the model selection and yielded small fold-to-fold variation; but the random stratified split was based on the assumption of exchangeable records. Observations of the agricultural characteristics may be clustered at the spatial, temporal or within the agroecological zone and in the farm level. Roberts et al. demonstrate that predictive error is underestimated in random cross validation if such structures are not taken into account [18]. No location, farm or year identifiers were provided in the benchmark, so blocked validation was not possible. The accuracy reported was therefore 0.993 in the benchmark distribution and should not be interpreted as a limit for accuracy in other districts, future seasons or climate regimes.

No missing values, no duplicate values, no gross physical domain violations were found during the data-quality audit; however, the cleanliness of the data syntactically does not imply the representativeness of the data. There is some evidence that this learning resource has been curated because of the perfect class balance, near uniformity of sample size per class, and the absence of record-level provenance. The documentation of a dataset should include information about the purpose of collection, augmentation, units of measurement, sampling frame, known skews, and recommended uses, consistent with the transparency goals of datasheets for datasets [19]. While these details and external observations are not available the prototype should be thought of as a reproducible analytical demonstrator.

Access, governance and contestation of a digital recommendation are also key factors in policy translation. Birner, Daum and Pray [20] indicate that private companies, public policies, market structure and managing concentration and access risks all play a role in shaping the digital transformation of agriculture. It is thus recommended that a public crop-planning service should not be set up as a model-only. It must employ open standards, auditable changes, offline/low bandwidth access, explanation in multiple languages, institutional liability for negative outfalls and reporting system for farmers and extension officers to report local mismatch.

This trend in the comparative results agrees with agricultural machine-learning reviews where tree ensembles work well in the presence of heterogeneous tabular predictors [8] [10]. Additionally, they can be used to extend early crop-recommendation systems, which report only the top crop label instead of top 3 outputs and climate-stress sensitivity [13]. Probability rankings from a balanced benchmark are operationally attractive because they provide flexibility to accommodate the preferences and constraints of the farmer, but they are not actual measures of success in the real world. After taking in prevalence representative regional data, a reassessment of calibration should be done.

The most important variables were humidity, precipitation and potassium. The pattern is in line with the general trend of the literature on yield prediction, focusing on weather and soil characteristics [11], and should not be taken as a causal statement. Impurity importance can give weight to variables that have many possible splits and can spread weight around unpredictably between correlated variables. Future studies should integrate permutation importance, accumulated local effects, class-specific explanations and agronomic counterfactual review. Ideally, an inherently interpretable generalized additive/rule-list model should also be compared to the ensemble, especially if the recommendations are to be applied in public extension.

The scenario analysis is the main conceptual contribution of the study. Isolated warming was found to yield almost complete top-one retention and a decrease in top-three overlap of over 25 percent. A greater loss of top-one stability and ranked-set stability was caused by combined stress. This shows that baselines can be classified correctly with the same model while producing materially different alternatives after small perturbations. The outcome is consistent with a climate-impact modelling approach that focuses on nonlinearities and context-dependency, and the current perturbations are not yet predictive or scenario-based, but only diagnostic, due to a lack of any geographical or temporal scenario.

Practical deployment needs to be done in steps of validation. The benchmark should be replaced or enriched with multi-year, geo-referenced records with cultivar, sowing window, soil measurement protocol, irrigation, management, yield, crop failure and economic outcomes. Secondly, applied to evaluation, it should take into account the use of spatially and temporally blocked external samples, of probabilistic calibration, of uncertainty intervals, of subgroup analysis, and of comparison with agronomist recommendations. Third, participatory trials should evaluate if top-three outputs enhance decisions compared to the current extension practice without transferring risk to poorly connected or less resource farmers. Finally, policy evaluation should include adoption, distributional effects, water/in input effects, and governance cost of maintaining current data, models.

4. Conclusion

This study created a scalable prototype digital tool to connect seven soil and climate indicators to prioritized recommendations for 22 crop classes. The prospectively selected model was Random Forest with a test accuracy of 0.993, and a test macro-F1 of 0.993 and a perfect top-3 accuracy on the stratified benchmark holdout. Humidity, rainfall and potassium were the most important factors for model separation. The additional layer of scenario perturbation resulted in a top-one retention of 99.82% for the isolated warming, and was reduced to 86.18% under combined warming and reduction in rainfall and humidity, whereas the top-three overlap was reduced to 57.79% under combined warming. The results illustrate the ability of sensitivity analysis to reveal fragility in recommendations that is not captured by just looking at accuracy. The added value is methodological and operational and not agronomic proof: it combines transparent data checks, comparative validation, ranked outputs, empirical-range warnings and scenario based-stability in one decision workflow. The main drawbacks of the benchmark are that it is curated (meaning its class balance is determined by the principal) and it lacks location, year, yield, irrigation, extreme-event, cost, and market variables. Future research needs to validate the models externally in a spatially and temporally blocked manner and to include the outcomes in the models as they are

observed, to account for the uncertainty of the forecasts, to compare the models that can be interpreted easily, and to assess their usability and distributional effects with farmers, extension officers, agronomists, and public agencies prior to any actual planting recommendation.

2. Methodology

2.1 Research Design

The data set of diabetes was analyzed to explore clinical risk factors related to diabetes outcome among patients (Akturk, 2020). The study design of this research was a quantitative analytical research design to determine the relationship that exists between diabetes status and the different clinical and demographic variables. As diabetes outcome was recorded as a binary variable, the study was carried out as an empirical observational study in the form of the outcome-based association analysis. The aim of the research was to establish the impact of the chosen clinical indicators on the diabetes outcome as well as to establish the most effective factors related to the diabetic and non-diabetic status of the study population.

2.2 Data Source and Sample

The data was a set of 768 observations that were patient-related clinical records. The data that was present in the dataset were as follows: pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, age, and outcome. These variables gave clinical, physiological and hereditary data that are pertinent to the assessment of diabetes. All the valid observations in the dataset were considered in the analysis. The records were filtered to have consistency in the values and structure of the variables and the dataset was found to be appropriate to study the determinants of diabetes in a structured analytical model.

2.3 Variables and Measures

The dependent variable was diabetes outcome which was measured as a binary variable of non-diabetic or diabetic. Pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function and age were the primary independent variables because these were the key clinical and hereditary risk factors that could have been related to the risk of diabetes. Demographic-related characteristics were considered as pregnancy and age whereas the measures of glucose, blood pressure, skin thickness, insulin, and body mass index were regarded as physiological and clinical measures. Pedigree functions were added to diabetes as a hereditary factor that represents the tendency to diabetes depending on the family. These variables were chosen so as to give a complete evaluation of the patient related factors affecting the outcome of diabetes.

2.4 Data Processing and Preparation

The dataset was pre-processed before it was analyzed to make sure that the analysis is consistent analytically and enhance the credibility of the interpretation. The dataset was initially analyzed in terms of structural uniformity, type of variable and completeness. Though the dataset had no blank cells, there were some variables with zero values which were not clinically realistic, especially glucose, blood pressure, skin thickness, insulin and body mass index. They were considered to be invalid values and tackled during the preprocessing to minimize the distortion in the findings. Numerous variables like pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, and age were left in numerical form to compare. Extreme values were also filtered against and they were retained where they were representing reasonable clinical variation. This step of preparation made sure that the data were ready to be analyzed descriptively, comparatively, and in terms of associations.

2.5 Data Analysis Technique

The analyses were done in two stages. The descriptive statistics were employed to summarize the variables of the study. The continuous variables were given in the form of mean and standard deviation whereas the categories of the outcomes were given in terms of frequencies and percentages. This gave a general clinical profile of the study population. Second, comparative and association-based analysis was done to investigate the relationship between the independent variables and diabetes outcome. The two groups of diabetic and non-diabetic were compared to find out the differences in the mean values of the clinical indicators. The relative strength of the relationship between each independent variable and outcome of diabetes was then determined by relationship-based analysis. Special focus was on such variables like glucose, body mass index, age, pregnancies and diabetes pedigree function since these are

usually linked to the incidence of diabetes. This analytical process assisted in the determination of the key clinical determinants of the outcome of diabetes in patients.

3. Results

3.1 Descriptive Statistics of Study Variables

This study involved 768 records of adult females. The mean pregnancies were 3.85, mean glucose level was 120.89 and the mean body mass index (BMI) was 31.99. The mean age of the respondents was 33.24 years with the mean diabetes pedigree function at 0.47. The minimum and maximum values show that there is a significant change in the clinical characteristics of the study population, especially glucose, insulin, and BMI as shown in Table 1. These results give a general clinical picture of the data employed to analyze the outcomes of diabetes.

Table 1. Descriptive statistics of study variables

Variable	Mean	Std. Dev.	Minimum	Maximum
Pregnancies	3.85	3.37	0	17
Glucose	120.89	31.97	0	199
Blood pressure	69.11	19.36	0	122
Skin thickness	20.54	15.95	0	99
Insulin	79.80	115.24	0	846
Body mass index (BMI)	31.99	7.88	0.00	67.10
Diabetes pedigree function	0.47	0.33	0.08	2.42
Age (years)	33.24	11.76	21	81

3.2 Distribution of Diabetes Outcome

The outcome variable, diabetes status revealed that the population under study was classified into diabetic and non-diabetic. Out of the 768 participants, 500 (65.10%) were non-diabetic, while 268 (34.90%) were diabetic. This is suggesting that most of those who were in the dataset were not diabetic but the number of diabetic cases was so adequate to conduct analysis based on comparative and association aspects. The frequency distribution shown in Table 2 shows that the data is suitable to study the determinants of diabetes outcome. Most of the participants were non-diabetic as depicted in Figure 1.

Table 2. Distribution of diabetes outcome

Diabetes status	Frequency	Percentage
Non-diabetic	500	65.10
Diabetic	268	34.90

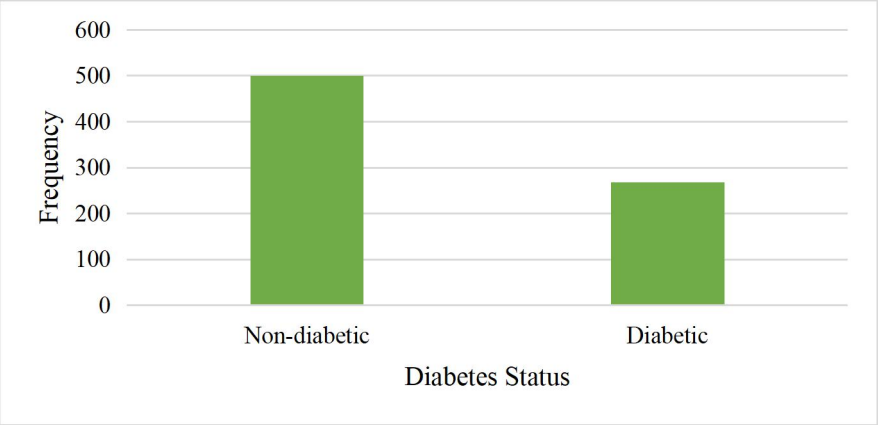


Figure 1. Distribution of Diabetes Status Among Study Participants

3.3 Comparison of Clinical Indicators Between Diabetic and Non-Diabetic Groups

The compared groups were the diabetic and non-diabetic groups, to find out the difference between the mean values of the clinical indicators. Table 3 showed that diabetic participants had greater mean values of most of the key variables as compared to non-diabetic participants. Specifically, diabetic group exhibited greater glucose (141.26), BMI (35.14), age (37.07 years), pregnancies (4.87) and diabetes pedigree function (0.55) than the non-diabetic group. The diabetic participants also had a higher insulin level. These findings suggest that the clinical profile of the diabetic group was not as favorable compared to non-diabetic group.

Table 3. Comparison of study variables by diabetes status

Variable	Non-diabetic (Mean ± SD)	Diabetic (Mean ± SD)
Pregnancies	3.30 ± 3.02	4.87 ± 3.74
Glucose	109.98 ± 26.14	141.26 ± 31.94
Blood pressure	68.18 ± 18.06	70.82 ± 21.49
Skin thickness	19.66 ± 14.89	22.16 ± 17.68
Insulin	68.79 ± 98.87	100.34 ± 138.69
Body mass index (BMI)	30.30 ± 7.69	35.14 ± 7.26
Diabetes pedigree function	0.43 ± 0.30	0.55 ± 0.37
Age (years)	31.19 ± 11.67	37.07 ± 10.97

3.4 Association-Based Comparison of Key Clinical Factors

Relationship-based analysis was conducted to determine the relative strength of association between selected independent variables and diabetes outcome. Particular attention was given to glucose, BMI, age, pregnancies, and diabetes pedigree function. As shown in Table 4, elevated glucose had the strongest association with diabetes outcome, followed by obesity and older age. Higher pregnancies and greater diabetes pedigree function were also associated with increased likelihood of diabetes. These findings suggest that both metabolic and hereditary factors play an important role in diabetes occurrence.

Table 4. Relative association of selected clinical factors with diabetes outcome

Variable	Higher-risk group	Reference group	Relative association
Glucose	≥126 mg/dL	<126 mg/dL	4.62
Body mass index (BMI)	≥30 kg/m²	<30 kg/m²	2.48
Age	≥35 years	<35 years	2.11

Pregnancies	≥5	<5	1.89
Diabetes pedigree function	≥0.50	<0.50	1.76

3.5 Major Determinants of Diabetes Outcome

Based on the descriptive, comparative, and association-based analyses, the major determinants of diabetes outcome in this study were glucose level, BMI, age, pregnancies, and diabetes pedigree function. Glucose emerged as the most influential clinical indicator, with diabetic participants showing much higher average values than non-diabetic participants. BMI and age also showed clear differences between the two groups, indicating that obesity and increasing age are closely related to diabetes occurrence. Likewise, higher pregnancies and stronger hereditary predisposition, reflected by diabetes pedigree function, were associated with greater diabetes risk. These findings demonstrate that diabetes outcome is influenced by a combination of metabolic, demographic, and family-history-related factors. As shown in Figure 2, diabetic participants had higher mean values for glucose, BMI, age, pregnancies, and diabetes pedigree function than non-diabetic participants.

Table 5. Summary of major determinants of diabetes outcome

Determinant	Non-diabetic Mean	Diabetic Mean	Observation
Glucose	109.98	141.26	Strongest difference between groups
Body mass index (BMI)	30.30	35.14	Higher in diabetic participants
Age (years)	31.19	37.07	Diabetes more common at older ages
Pregnancies	3.30	4.87	Higher among diabetic participants
Diabetes pedigree function	0.43	0.55	Greater hereditary influence in diabetics

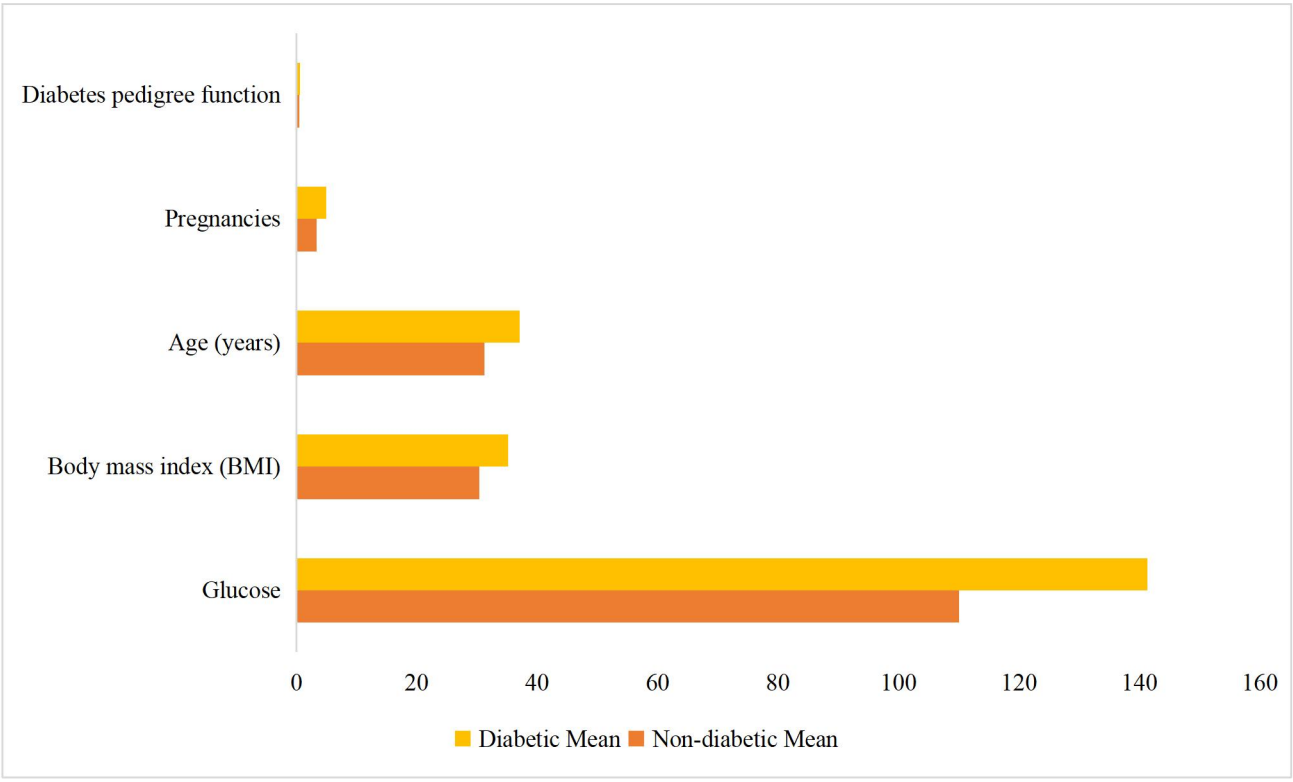


Figure 2. Mean Clinical Indicators by Diabetes Status

4. Discussion

The outcome of diabetes in the dataset was closely linked to the change in glucose level, body mass index, age, pregnancies and diabetes pedigree functioning. The diabetic population and the non-diabetic one always demonstrated an unfavourable clinical profile that suggests that the diabetes status is predetermined by a complex of metabolic, demographic and familial factors. These results are significant since they go beyond mere classification and give interpretable results about the relationship between certain patient characteristics and diabetic status.

Glucose was found to be the most significant predictor of diabetes outcome of all variables studied. The average level of glucose was significantly higher in diabetic subjects as compared to non-diabetic ones and the relative relationship analysis revealed that high glucose was the most correlated with diabetic status. This is expected clinically because glucose level is the most immediate metabolic value to reflect impaired glycemic regulation and the development of diabetes. It is also in line with the larger body of evidence that metabolic risk profiles play the pivotal role in the identification and progression of diabetes (Ye et al., 2022). It also converges with the Pima Indian population where defects in the control of glucose have been found to be central to the etiopathogenesis and diagnosis of type 2 diabetes (Looker et al., 2023). As such, the high contribution of glucose in this study supports the significance of this clinical parameter as the key one in diabetes-associated evaluation.

Another significant factor was body mass index that was linked to the outcome of diabetes. Average BMI was greater in the diabetic group compared to the non-diabetic group and BMI at the level of obesity also displayed powerful relative relationship with diabetes. This means that overweight is highly connected with the diabetic status among the study population. This result is significant since obesity is a risk factor leading to insulin resistance, one of the major processes involved in the development and the evolution of type 2 diabetes. The finding is consistent with prior prospective research indicating that the presence of metabolic and lifestyle-related risk factors together determine the development of diabetes, with body weight significantly contributing (Ye et al., 2022). It further indicates that BMI is among the most understandable and clinically significant variables in structured patient data to comprehend the risk of diabetes.

Another significant determinant was found to be age. Diabetic participants were older than non-diabetic participants and those aged 35 years and over were more relatively related to the outcome of diabetes. This implies that diabetes tends to increase with age, probably due to the fact that the risk factors that predispose to diabetes become more metabolically stressful and that there is a cumulative exposure to risk factors that predisposes to diabetes over time. Clinical evidence confirms this interpretation since it suggests that demographic factors and, in particular, age continue to be closely related to occurrence and long-term complications of type 2 diabetes (Looker et al., 2023). The results, therefore, support the hypothesis that the risk of diabetes is metabolic as well as age-related.

Pregnancies and diabetes pedigree also played a significant role in the outcome of diabetes. The average number of pregnancies was higher and the mean diabetes pedigree function was also higher in diabetic people as compared to non-diabetic participants. These results indicate that both reproductive history and hereditary predisposition have a role in determining the risk of diabetes. The role of diabetes pedigree function is especially interesting as it demonstrates the family-based susceptibility to diabetes that diabetes pedigree functions can have even in cases when clinical indicators are taken into account. This explanation agrees with earlier findings that family history is a significant factor in diabetes risk patterns and interplays with metabolic and lifestyle factors in the development of the disease (Ye et al., 2022). In this regard, the results are valid and affirmative that diabetes outcome is not dictated by a single variable but one is a wider interaction of inherited and physiological factors.

The other notable aspect is the fact that the results are focused on interpretability, but not on prediction accuracy only. The current state of the art of diabetes literature has been dominated by advanced machine learning models to enhance classification performance on the PIMA Indians dataset, such as deep and hybrid models (Shams et al., 2025). Unless interpretability is deliberately added, although such models are useful in the predictive context, they can offer little clinical exegesis. Recent contributions to explainable machine learning models have emphasized the need to determine the contribution of certain variables to the prediction results in order to ensure that the results are meaningful to support clinical decisions (Netayawijit et al., 2025). This need is filled by the analysis, which, through a simple interpretation of statistics, shows that the most significant determinants of diabetes status are glucose, BMI, age, pregnancies, and diabetes pedigree function.

The results can also be considered through the larger trend of the existing body of diabetes research, in which a growing focus has shifted towards the idea of early detection and prediction of risk through the use of both conventional analysis techniques and machine learning. There is systematic evidence that early diabetes forecasting is most effective when the models are informed by a clear knowledge of the predictors underlying the model as opposed to viewing the model as a black-box (Kadam et al., 2025). In this analysis, it was the combination of descriptive statistics, group comparison and relative association analysis that assisted in the revelation of the clinical meaning of the dataset in a transparent manner.

5. Conclusion

A combination of metabolic, demographic and hereditary factors was observed to be closely related to diabetes outcome. The study clearly illustrated the clinical profile of the study population and showed significant variations between diabetic and non-diabetic subjects in various key indicators. Diabetics were always characterized by an increased mean of glucose, body mass index, age, pregnancies, insulin, and diabetes pedigree functionality, which means a relatively worse clinical state. Glucose was found to be the most influential variable in defining the outcome of diabetes, as compared to the other variables examined, and it was followed by body mass index, age, pregnancies and diabetes pedigree function. These results indicate that high blood glucose is the most significant indicator of diabetic condition, whereas obesity, age, reproductive history and family history are also significant in occurrence of diabetes. The outcome of diabetes is, however, not caused by one risk factor, but it can be seen as a reflection of the effect of the combination of the many patient-related characteristics. The results present a comprehensible insight into the key determinants that relate to diabetes outcome instead of emphasizing the predictive classification. This renders the results useful in clinical interpretation and future research on diabetes based on data. The analysis facilitates the relevance of early risk identification in the form of regular clinical and demographic evaluation by determining the most important indicators associated with the diabetic status.

References

1. Akturk, M. (2020). Diabetes dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/mathchi/diabetes-data-set
2. Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. Journal of healthcare engineering, 2021(1), 9930985.
3. Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, 35(22), 16157-16173.
4. Hasan, M., & Yasmin, F. (2025). Predicting diabetes using machine learning: A comparative study of classifiers. arXiv preprint arXiv:2505.07036.
5. Howlader, K. C., Satu, M. S., Awal, M. A., Islam, M. R., Islam, S. M. S., Quinn, J. M., & Moni, M. A. (2022). Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. Health information science and systems, 10(1), 2.
6. Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. Primary Care Diabetes, 15(3), 435-443.
7. Kadam, P., Godse, S., & Mahalle, P. (2025, October). Machine learning in the early detection and prediction of diabetes: A systematic review. In AIP Conference Proceedings (Vol. 3325, No. 1, p. 070036). AIP Publishing LLC.
8. Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis. Frontiers in digital health, 7, 1557467.
9. Looker, H. C., Chang, D. C., Baier, L. J., Hanson, R. L., & Nelson, R. G. (2023). Diagnostic criteria and etiopathogenesis of type 2 diabetes and its complications: lessons from the Pima Indians. La Presse Médicale, 52(1), 104176.
10. Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. Health information science and systems, 8(1), 7.

11. Narasimharao, M., Swain, B., Priyadarshi, R., Nayak, P. P., & Bhuyan, S. (2025). A survey of machine learning techniques for diabetes prediction: current trends and future directions. Archives of Computational Methods in Engineering, 1-36.
12. Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19(1), 391-403.
13. Netayawijit, P., Chansanam, W., & Sorn-In, K. (2025, October). Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support. In Healthcare (Vol. 13, No. 20, p. 2588). MDPI.
14. Regmi, D., Al-Shamsi, S., Govender, R. D., & Al Kaabi, J. (2020). Incidence and risk factors of type 2 diabetes mellitus in an overweight and obese population: a long-term retrospective cohort study from a Gulf state. BMJ open, 10(7), e035813.
15. Shams, M. Y., Tarek, Z., & Elshewey, A. M. (2025). A novel RFE-GRU model for diabetes classification using PIMA Indian dataset. Scientific reports, 15(1), 982.
16. Ye, C., Wang, Y., Kong, L., Zhao, Z., Li, M., Xu, Y., ... & Wang, T. (2022). Comprehensive risk profiles of family history and lifestyle and metabolic risk factors in relation to diabetes: A prospective cohort study. Journal of Diabetes, 14(6), 414-424.