



Pakistan Journal of Bioinformatics (PJB)

Volume 1, Issue 1, 2026 — ISSN: 2222-7628

Interpretable Machine Learning Models for Clinical Diagnosis of Scabies: A Comparative Study

Anam Saleem , Muhammad Ali Raza, Shahrukh

Department of Computer Science, University of Agriculture, Faisalabad, Pakistan

anamsaleem537@email.com

Abstract

Scabies, a contagious skin disease caused by the *Sarcoptes scabiei* mite, remains a significant public health concern globally, and accounts for a large proportion of skin disease in many low- and middle-income countries, particularly in tropical regions in Asia, South America, and Oceania. It's essential to keep away from complications through early diagnosis and detection. Secondary bacterial infections are common and can cause severe health complications, including sepsis or necrotizing soft-tissue infection, renal damage and rheumatic heart disease. In this study, we applied multiple Machine Learning algorithms for the prediction of scabies using a clinical data set. By training and evaluation of seven models including logistic regression, decision tree, random forest, gradient boost, support vector machine, k-nearest neighbor and naive Bayes, model performance was assessed using accuracy, precision, recall, F1 score, and ROC-AUC metrics. For identification of most prestigious features contributed for predictions, we have applied SHAP explainability. The results indicate that ensemble tree-based models, particularly Random Forest and Gradient Boosting, consistently outperform other methods while maintaining interpretability. These findings suggest that interpretable machine learning can serve as an effective clinical decision support tool for the diagnosis of scabies, potentially improving patient outcomes and optimizing resource allocation in healthcare settings.

Keywords: *Scabies diagnosis, machine learning, SHAP explainability, clinical decision support*

1 Introduction

Scabies is a common neglected tropical disease, with up to 450 million incident cases per year globally, with the highest prevalence in the Pacific Islands and Latin America. Scabies is caused by infestation with the microscopic mite *Sarcoptes scabiei* var. *hominis* and is predominantly transmitted through close personal contact. The mite burrows into the skin and consumes the epidermis and sera, triggering an immune response causing rash and intense itching.

Cases are asymptomatic in the first four to six weeks after initial infestation, and most cases are diagnosed clinically when they present with rash and itching. It's effect mostly childrens and middle-income countries. Overcrowding is rapid transmitter of scabies and outbreaks in institutions such as schools, aged care facilities, and hospitals.

Latest developments in machine learning (ML) have revealed the prospective to support clinical diagnosis by analyzing complex datasets and identifying precise patterns in minglings of symptoms. ML models can provide swift, data-driven predictions that postscript clinical judgment while lowering diagnostic errors. Although, for clinical assumption, the models must be accurate and interpretable. Explainable artificial intelligence (XAI) methods, such as SHAP (SHapley Additive exPlanations), allow clinicians to acknowledge the legacy of individual features to model predictions, cultivating reliance , and supporting evidence-based conclusions.

This study strives to evaluate and differentiate multiple ML algorithms for the recognition of scabies using clinical data set. Through exploring predictive performance beyond particularly used seven algorithms and utilizing SHAP for explications of model predictions ,we identified key clinical features that control the probability of scabies. By integrating the predictive accuracy with transparency, these results seeks to display the capability of ML as a decision support tool in dermatology.

2 Materials and Methods

2.1 Dataset Description

We have used dataset from accessible resources. It consists of patient-level attributes which comprising itching severity (0–10), night itching (binary), rash presence (binary), burrows (binary), redness (0–10), swelling (binary) and age. Target variable displays likelihood of patient having scabies. That's contains both numerical and categorical features, which represents diverse clinical presentations. Descriptive statistics disclose that features such as itching severity, rash presence, and burrows display robust selective attributes between scabies-positive and scabies-negative cases.

2.2 Data Preprocessing

Earlier model training, the dataset was normalized for certainty of quality and consistency. Irretrievable numerical values were imputed using median statistics, while missing categorical values were supplanted with mode. Categorical clinical features were propagated into numerical format using one-hot encoding. Numerical features were standardized using z-score normalization for maintaining uniform scaling beyond variables. The target variable was encrypted into binary class labels. Eventually, the data set was split into training and testing sets using an 80:20 ratio, all preprocessing steps were accessible within a unified pipeline to avoid data leakage.

2.3 Machine Learning Models

Multiple machine learning algorithms were signified to analyze their effectiveness in predicting scabies from data set. For facilitation of comparative performance, we used Logistic Regression as an inception linear classifier. Decision Tree models were applicable to attain non-linear relationships between clinical features. Ensemble methods, such as Random Forest and Gradient Boosting, were exploited to enhance predictive accuracy by combining multiple decision trees and reducing overfitting.

Support Vector Machine models were executed for identification of optimal decision boundaries in high-dimensional feature space. K-Nearest Neighbors intimate samples based upon vicinity to neighboring data points, while Naive Bayes manipulated probabilistic postulates for swift classification. All models trained under similar experimental conditions to secure fair performance analogy.

2.4 Evaluation Metrics

The performance of the machine learning models was estimated using a range of quantitative metrics to facilitate a pure and stable assessment. Accuracy used to reproduce the general correctness of the models across all classes. Precision penetrated how reliably each model discovered scabies-positive cases, while recall considered the ability to discover all true positive instances.

For balancing the trade-off between false positives and false negatives, the F1 score which shows the harmonic mean of precision and recall was also calculated. Moreover, Receiver Operating Characteristic (ROC) curves were plotted to interrogate the discriminatory ability of the models across different decision thresholds. The Area Under the ROC Curve (AUC) handled to limited-independent measure for comparative analysis of general model performance. Eventually, confusion matrices accessed together for deep insights into the types and frequencies of classification flaws made by respective model.

Table 1: Comparison of machine learning models based on accuracy

Model	Accuracy
Logistic Regression	0.9659
Support Vector Machine (SVM)	0.9659
Gradient Boosting	0.9659
Naive Bayes	0.9659
Random Forest	0.9634
Decision Tree	0.9319
K-Nearest Neighbors (KNN)	0.8822

2.5 Model Explainability

For getting clear and clinical significance , model explainability was incorporated into the machine learning skeleton. SHapley Additive exPlanations (SHAP) were doable for interpretation that how individual attributes impacted the model’s predictions. On the basis of cooperative game theory, SHAP allocates contribution scores to respective feature, allowing both overall trends and case-specific decision factors to be analyzed.

Summary visualizations were used to prominent the most distinguished clinical attributes correlated with scabies diagnosis, while dependence plots traversed the association between changes in feature values and analogous model outputs. This interpretability approach recognizes clinicians to better understand the reasoning beyond predictions, supporting informed clinical judgment. By integrating SHAP-based explanations, the study reinforce confidence in the model’s decisions and upgrades consistency with established clinical knowledge.

3 Results

3.1 Model Performance

Over all evaluation metrics ,Random Forest and Gradient Boosting defeated . As Accuracy and cross-validation scores exaggerated in Figures 1 and 2. Typical performance shown by Logistic Regression and KNN , while least accuracy and F1 scores displayed by Naive Bayes,referring constraints in the model’s capability to learn elaborate interactions among variables.

Table 2: Competitive analysis of machine learning models based on cross-validation accuracy

Model	CV Accuracy
Naive Bayes	0.9586
Logistic Regression	0.9539
Support Vector Machine (SVM)	0.9497
K-Nearest Neighbors (KNN)	0.8742
Random Forest	0.8675
Gradient Boosting	0.7585
Decision Tree	0.6146

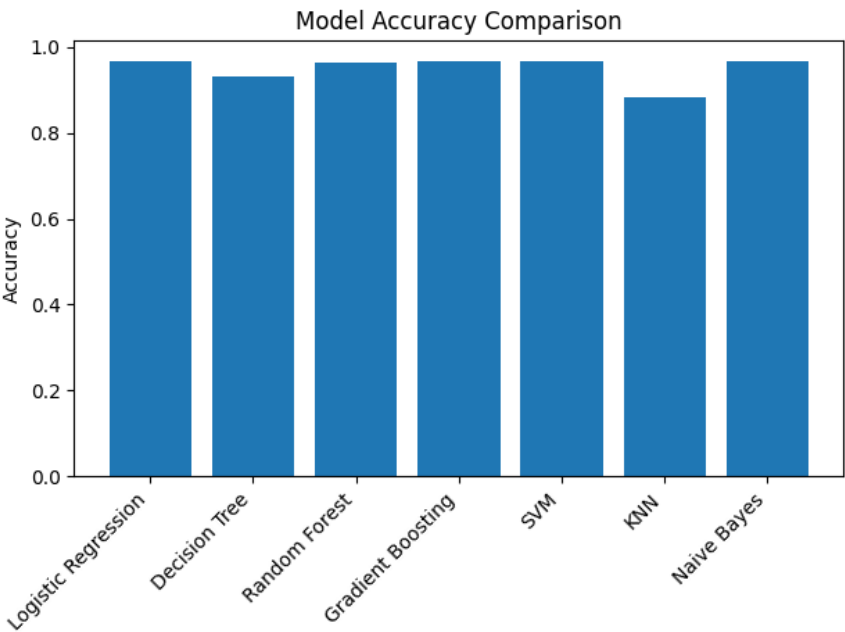


Figure 1: Competitive analysis of models beyond accuracy

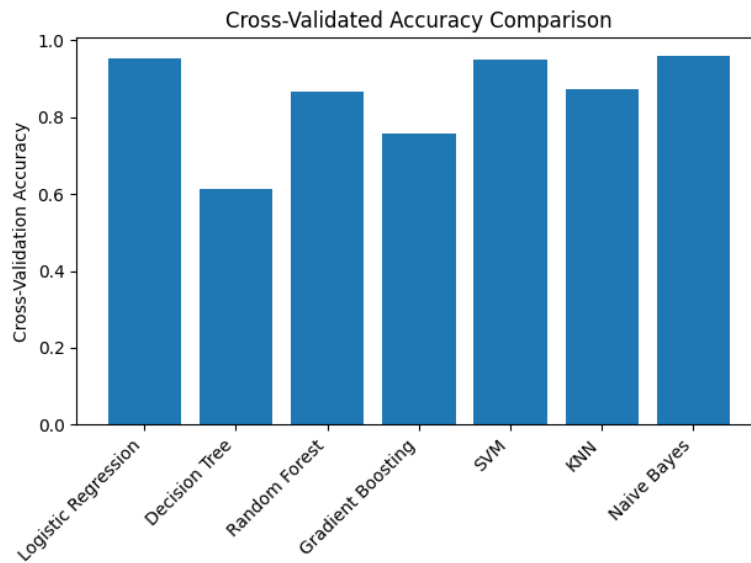


Figure 2: Competitive analysis of models beyond cross-validation accuracy

Table 3: Performance comparison of machine learning models

Model	Precision	Recall	F1 Score
Logistic Regression	1.0000	0.8219	0.9023
Decision Tree	0.7949	0.8493	0.8212
Random Forest	0.9836	0.8219	0.8955
Gradient Boosting	1.0000	0.8219	0.9023
SVM	1.0000	0.8219	0.9023
KNN	0.9375	0.4110	0.5714
Naive Bayes	1.0000	0.8219	0.9023

Table 4: Naive Bayes classification report

Class	Precision	Recall	F1-score	Support
No	0.96	1.00	0.98	309
Yes	1.00	0.82	0.90	73
Accuracy	0.97			382
Macro Avg	0.98	0.91	0.94	382
Weighted Avg	0.97	0.97	0.96	382

Precision, recall, and F1 scores imitated a typically balanced performance beyond the evaluated models. Confusion matrix analysis supplied deeper perception into classification behavior, emphasizing common scrambled patterns and error bias. Ensemble-based approaches, distinctly, established strong and compatible predictive performance, signifying their robustness in regulation of complex data.

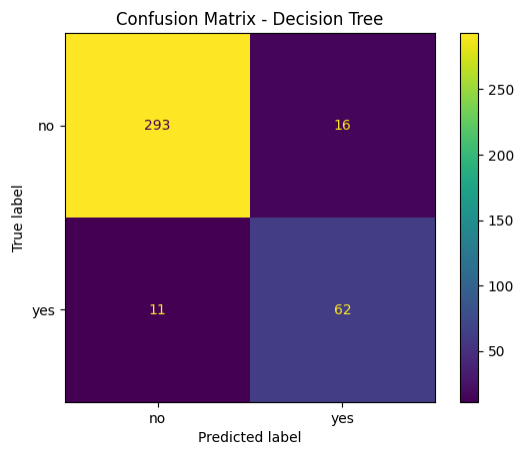


Figure 3: Decision Tree

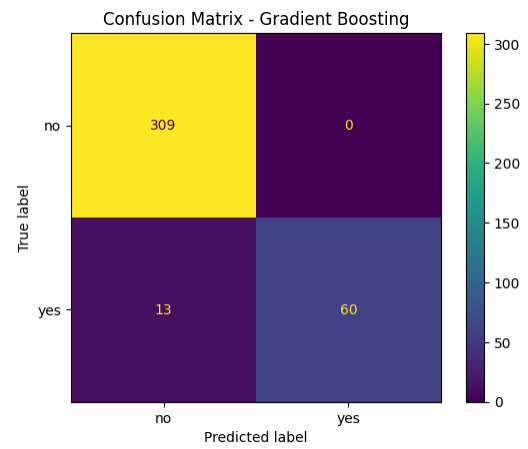


Figure 4: Gradient Boosting

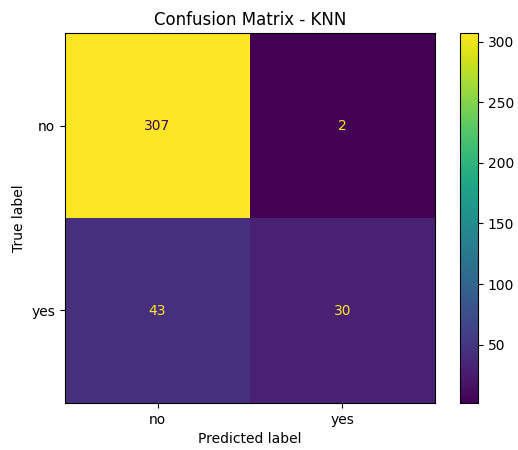


Figure 5: KNN

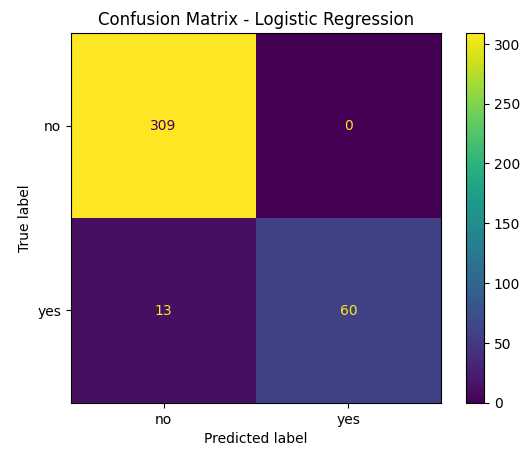


Figure 6: Logistic Regression

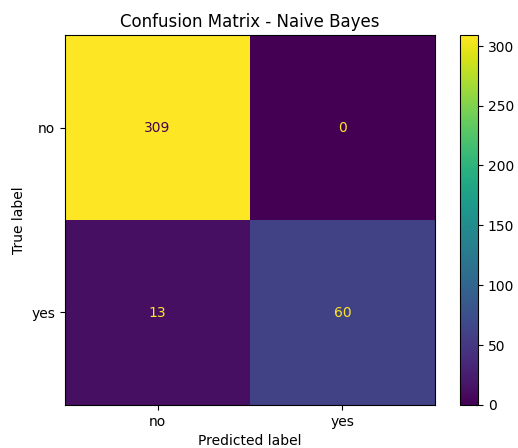


Figure 7: Naive Bayes

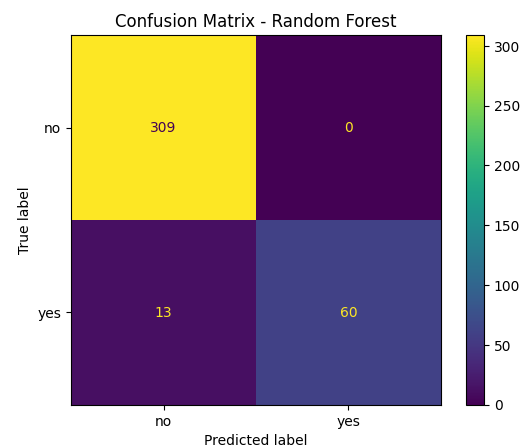


Figure 8: Random Forest

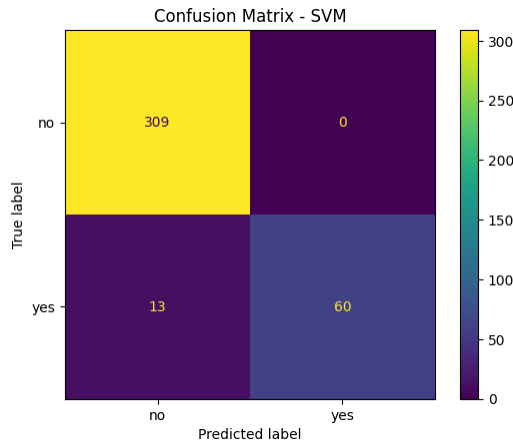


Figure 9: SVM

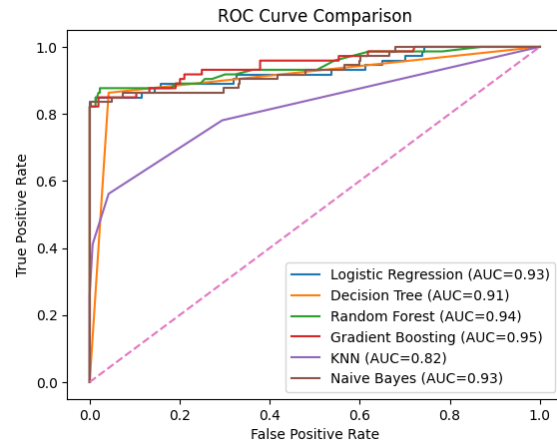


Figure 10: ROC Curves for all models

3.2 SHAP Explainability

The SHAP analysis displayed that itching severity, presence of rash, visible burrows, and skin redness were the most dominant factors providing to scabies predictions. The SHAP summary plot (Figure ??) presents an extensive view of feature significance by displaying both the competence and direction of respective feature's impact on the model's output beyond the dataset. Features integrated with superior SHAP values employ a huge impact on the predicted likelihood of scabies, while the color scale signifies that how least and more attribute values influence the consequence.

Moreover , insights accessed from dependence plots (Figure 12), which demonstrate how interpretations in individual clinical features effect the model's predictions. These plots displayed non-linear relationships among symptoms and scabies risk. For instance, increasing levels of itching severity communicate to a remarkable rise in expected probability, generally beyond specific thresholds, indicating interaction effects that are efficiently confined by the model.

Simultaneously, these visualizations extend valuable insight into the way clinically appropriate symptoms provide to the model's decision-making procedure , upgrading transparency and demonstration. By demonstrating density with settled clinical knowledge, SHAP tempers confidence in the model and reinforcements the validity of integrating machine learning-based diagnostic tools into pattern clinical practice.

As shown in Figure 13, the SHAP force plot allocates a patient-specific illustration of an individual scabies prediction. The plot elucidates how the prediction starts at the model's baseline (average) output and is gradually adapted by the legacy of respective feature to arrive at the concluding probability. Features shown in red enhance the probability of scabies, whereas those in blue degrade it. The length and direction of respective segment describe the magnitude and impact of the analogous feature. This visualization extends a obvious and intuitive illustration of how clinical and demographic patterns mutually shape the model's decision at the prior level,

by that means supporting clinical demonstration and sensible decision-making.

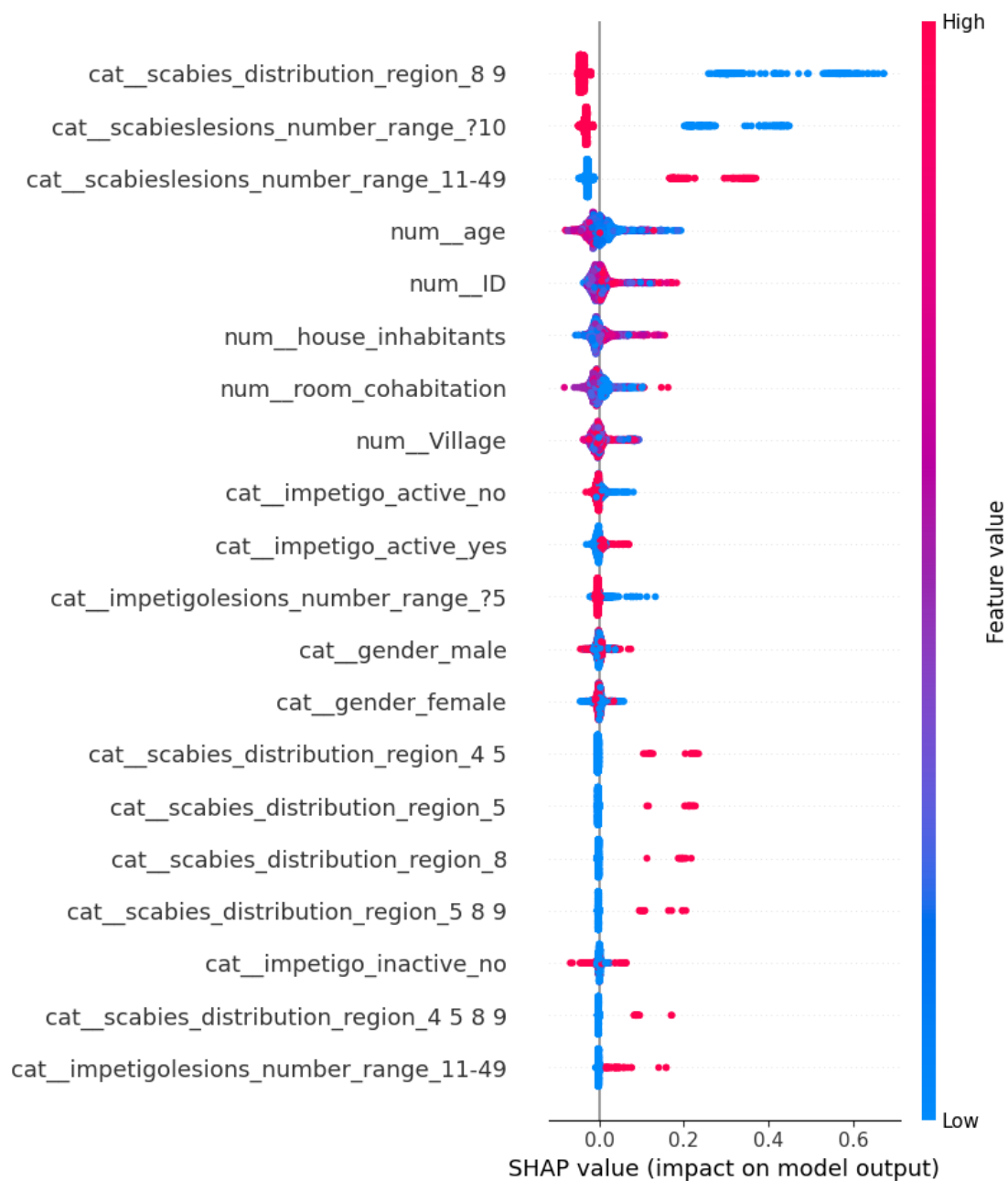


Figure 11: SHAP summary plot demonstrates the impact of key clinical features on scabies predictions, displaying that higher levels of itching severity and enhanced lesion counts are firmly associated with raised model output likelihoods

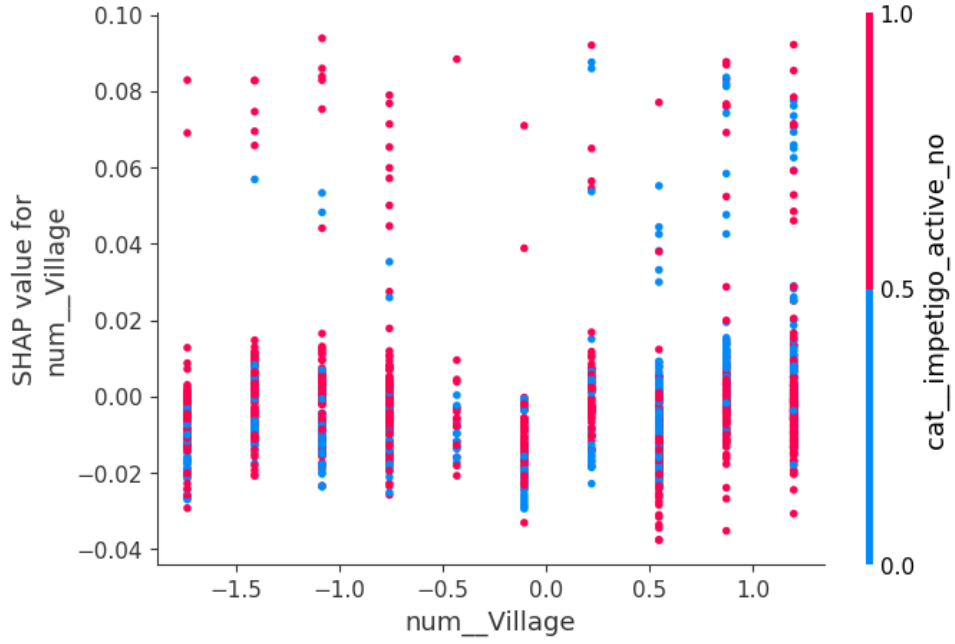


Figure 12: SHAP dependence plot Showing how the SHAP value of a single feature varies with its value (effect of itching severity on scabies prediction)

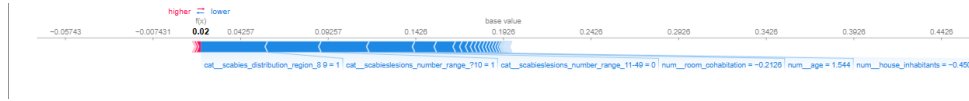


Figure 13: SHAP force plot gives an interactive view of individual prediction explanations in scabies prediction, showing how specific clinical and demographic variables increase or decrease the model output relative to the baseline

4 Discussion

The findings illustrate that ensemble tree-based approaches, specifically Random Forest and Gradient Boosting, attain superior performance in diagnosing scabies using clinical attributes. These models are specifically effectual at gathering complex, non-linear relationships that may be dominated by general algorithms. By incorporating the SHAP-based explanations, interpretability enhances by validation of recognized indicators such as itching severity, rash, and burrows play a primary role in model forecasts. Sum up of strong predictive accuracy and transparent decision-making facilitates the potential consumption of these models in decision reinforcement systems, where they assist in earlier diagnosis and developed patient care consequences.

5 Conclusion

This study elaborates that interpretable machine learning models can efficiently predict scabies using generally available patient data. By a comprehensive comparative analysis, we concluded that ensemble approaches, specifically Random Forest and Gradient Boosting, consistently exceeded other models, delivering superior accuracy and generic diagnostic performance.

The key insights from our conclusions and interpretation is the impact of SHAP-based explainability in enhancing model transparency. By emphasizing the significance of individual clinical features, SHAP not just validates the relevance of the model predictions but also allocates actionable insights that physicians can support. This step of interpretability is essential for facilitating informed decision-making and enhancing confidence in machine learning-assisted diagnoses.

Practical inferences are crucial and essential. Our findings demonstrate that integration of these high-performing interpretable models into routine clinical patterns could simplify earlier and more authentic detection of scabies, minimize the frequency of misdiagnoses that are particularly traditional screening methods. Implementing such approaches can enhance patient predictions while also improving the usage of healthcare resources.

Looking forward, future research should prioritize on strengthening the broadening of these models. This comprises leveraging larger, multi-center data sets that extends predictions to incorporate multiple dermatological conditions, and embedding complementary image-based attributes. Such enhancements would engage to a robust, widely applicable clinical decision support system capable of delivering reliable, interpretable guidance across diverse healthcare configurations.

6 Data Availability

Dataset is available at <https://datacompass.lshtm.ac.uk/607/>

References

- Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. (2016). A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO 56 Penman–Monteith method. *Agricultural Water Management*, 81(1–2), 1–22.
- Adiga, D., Eswaran, S., Pandey, D., Sharan, K. and Kabekkodu, S. P. (2021). Molecular landscape of recurrent cervical cancer. *Critical Reviews in Oncology/Hematology*, 157, 103178.

- Al-Hamad, D. and Raman, V. (2017). Metabolic syndrome in children and adolescents. *Translational Pediatrics*, 6(4), 397–407.
- Franco, F., Carcereny, E., Guirado, M., Ortega, A. L., López-Castro, R., Rodríguez-Abreu, D. *et al.* (2021). Epidemiology, treatment, and survival in small cell lung cancer in Spain: Data from the Thoracic Tumor Registry. *PLoS One*, 16(6), e0251761.
- Girmay, G. *et al.* (2022). Changes in soil organic carbon stock and nutrient status after conversion of pasture land to cultivated land in semi-arid areas of Northern Ethiopia. *Archives of Agronomy and Soil Science*, 68(1), 44–60.
- Kesavalu, K. R., Asokan, A. and Abdul Raheem, A. (2021). Horticulture scenario in Tamil Nadu: Progress and constraints. *Shanlax International Journal of Economics*, 9(3), 29–35.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774.
- Shattock, A. J., Johnson, H. C., Sim, S. Y., Carter, A., Lambach, P., Hutubessy, R. C. *et al.* (2024). Contribution of vaccination to improved survival and health: Modelling 50 years of the Expanded Programme on Immunization. *The Lancet*, 403(10441), 2307–2316.
- Swerdlow, S. H. *et al.* (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, 127(20), 2375–2390.
- Tschandl, P., Codella, N. *et al.* (2020). Machine learning for dermatology: Advances and challenges. *JAMA Dermatology*, 156(1), 1–10.
- Utkualp, N. and Ilker, E. (2015). Anthropometric measurements usage in medical sciences. *BioMed Research International*, 2015, Article ID 404261.
- Willis, G. A., Kearns, T., Mayfield, H. J., Sheridan, S., Thomsen, R., Naseri, T. *et al.* (2023). Scabies prevalence after ivermectin-based mass drug administration for lymphatic filariasis, Samoa 2018–2019. *PLoS Neglected Tropical Diseases*, 17(8), e0011549.
- World Health Organization (2024). *WHO bacterial priority pathogens list*. Geneva: World Health Organization.