KEC Conference 2025

ISSN: 3021-940X (Print) DOI: https://doi.org/10.3126/injet.v2i2.78579

Enhancing Monkeypox Diagnosis: Explainable AI-Driven Evaluation of Machine Learning Models for Reliable and Transparent Detection

Saugat Kafle[1], Smriti KC[2], Satish Kumar Karna[3], Prakash Paudel[3, *]

[1] *Samriddhi College, Bhaktapur Nepal, saugatkafle77@gmail.com*
[2]*Padma Kanya Campus, Bagbazaar, Kathmandu, kcsmriti80@gmail.com*
[3]*Nepal College of Information Technology, Balkumari, Lalitpur, satishkarna76@gmail.com*
[3]*Nepal College of Information Technology, Balkumari, Lalitpur, visitppdl@gmail.com*

Abstract

Monkeypox is a re-emerging zoonotic disease caused by the monkeypox virus (MPXV), which raises serious public health issues owing to the possibility of human-to-human transmission. Traditional diagnostic methods, such as PCR and serological testing, are effective but time-consuming and resource-intensive, limiting rapid outbreak response. Machine learning-based systems are a viable alternative, but their lack of interpretability and transparency remains a significant limitation in clinical decision-making. This study addresses the challenges by applying four machine learning algorithms: AdaBoost, Gradient Boosting Classifier, Multilayer Perceptron (MLP), and LightGBM (LGBMC). LGBMC leads with the highest recall of 90.67%, showcasing its strong ability to identify positive cases. Gradient Boosting follows at 88.78%, while MLP and AdaBoost have slightly lower recall rates of 87.46% and 87.02%, respectively. LGBMC proves to be the optimal model for applications requiring high recall, while AdaBoost performs the least effectively. Additionally, receiver operating characteristic (ROC) curve analysis evaluates model performance and reliability. This technique enhances diagnostic accuracy and provides insights into feature importance, helping healthcare professionals understand the factors influencing predictions. Through XAI methodologies, the study bridges the gap between AI-driven diagnostics and practical clinical use, ensuring increased transparency, trust, and reliability in monkeypox detection.

***Keywords*:** Monkeypox, Machine Learning Algorithms, XAI, LIME, SHAP, Optimal Model

1. Introduction

Monkeypox disease, caused by the monkeypox virus, is a highly communicable disease and has prompted the World Health Organization to declare it as a Public Health Emergency (Magsino, et al., 2024). In the current global health scenario, as the world continues to emerge from the shadow of the COVID-19 pandemic, there is growing concern regarding a new challenge: the outbreak of monkeypox (Raha D. R., et al., 2024). Monkeypox is a rare and potentially fatal disease caused by the monkeypox virus. The disease is endemic in parts of Central and West Africa, where it is occasion- ally transmitted to humans through contact with infected animals or through human-to-human transmission. Early detection of monkeypox cases is crucial for effective outbreak response and control. In recent years, Machine Learning (ML) and Deep Learning (DL) techniques have been increasingly used to analyze large amounts of data and help detect outbreaks of infectious diseases (Patel, Srinivasulu, Jani, & Sreenivasulu, 2023). Monkeypox, while occasionally transmitted from one human to another, can be disseminated through the inhalation of droplets or through contact with the skin lesions of an infected individual. The disease has incubation period usually 7–14 days and it causes fever, headache, fatigue, myalgia, widespread body aches, swelling in lymph nodes and skin lesions. The World Health Organization (WHO) has raised the alarm over the monkeypox outbreak, classifying it as a global health emergency (Region--Low, 2024). Its spread is ongoing, with an increasing number of cases reported daily. Between January 1, 2022, and August 9, 2023, the WHO reported a total of 89,308 laboratory-confirmed monkeypox cases and 152 deaths across 113 countries in all six WHO regions (Region--Low, 2024). In the fortnight leading up to August 9, 2023, there was an increase in reported cases

**Corresponding Author*

in the Western Pacific, European, and American regions (Raha A. D., et al., Corrections to "Attention to Monkeypox: An Interpretable Monkeypox Detection Technique Using Attention Mechanism", 2024).

AI-based models often function as "black boxes," making it difficult to understand how predictions are made. To improve transparency, Explainable AI (XAI) techniques like SHAP and LIME are used (Paudel, et al., 2023). SHAP highlights key features affecting predictions, while LIME simplifies complex models for easier interpretation (Bhandari, Shahi, Siku, & Neupane, 2022) (Bhandari, Shrestha, Karki, Adhikari, & Gaihre, 2024). These techniques help clinicians interpret and trust AI decisions. Additionally, ROC curve analysis evaluates model performance, balancing sensitivity and specificity. The objectives of the proposed study are:

- To evaluate and compare the diagnostic performance of various ML models for accurate monkeypox detection.
- To assess the impact of key clinical and epidemiological features on model interpretability and prediction accuracy.
- To validate the stability and robustness of the models through cross- validation and performance metrics for consistent detection across datasets.

2. Related Works

A range of studies and publications specializing in ML and DL, particularly in healthcare projects, have been consulted for the purpose of this research.

1.1. Deep Learning and Machine Learning

(Raha D. R., et al., 2024) demonstrated that the support vector machine (SVM) achieved the highest performance in terms of the area under the precision-recall curve (AUC-PR), with a value of 79.67%. This model also reported a recall of 88.35%, indicating its ability to correctly identify most of the true positive cases. The precision, at 72.86%, reflects the accuracy of the positive predictions, while the F2-score, which balances recall and precision with more emphasis on recall, was 84.53%. (Haripriya & Inbarani, 2022) utilized the Gradient Boosting algorithm in their study to predict outcomes in a given dataset. Among several algorithms tested, Gradient Boosting provided the best performance, yielding the highest accuracy of 71%. (Raha D. R., et al., 2024) proposed an attention enhanced MobileNetV2 model that consistently outperformed baseline models. The model achieved impressive accuracies of 92.28% on the extended MSID dataset, 98.19% on the original MSID dataset, and 93.33% on the Monkeypox Skin Lesion Dataset (MSLD). (Nayak , et al., 2023) evaluated multiple deep learning models for diagnosing monkeypox. Among these, ResNet18 demonstrated superior performance with an accuracy of 99.49%. In addition, the other modified deep learning models reached validation accuracies above 95%. (Azar, et al., 2023) introduced a DenseNet201-based architecture that achieved the highest performance among the tested models. In a two-class scenario, it reached an accuracy of 97.63%, an F1-score of 90.51%, and an AUC of 94.27%. For a four-class scenario, the model achieved an accuracy of 95.18%, an F1-score of 89.61%, and an AUC of 92.06%.

1.2. Explainable AI

(Akın, Gurkan , Budak, & Karataş, 2022) demonstrated that the MobileNet V2 model outperformed other methods in classifying skin images, achieving an impressive accuracy of 98.25%, a sensitivity of 96.55%, a specificity of 100.00%, and an F1-Score of 98.25%. The study used 572 images (Monkeypox and Normal) and found MobileNet V2 to be the best among 12 CNN models for accurate classification. (Siddick, et al., 2024) implemented a comprehensive monkeypox diagnostic system utilizing deep learning architectures such as the Vision Transformer, MobileNetV2, EfficientNetV2, ResNet-50, and a hybrid model, all trained on a meticulously curated dataset. EfficientNetV2 achieved training, validation, and testing accuracies of 99.94%, 97.80%, and 97.67% respectively, while the Vision Transformer reached 100% training, 87.51% validation, and 99.41% testing accuracy.

3. Methodology

The proposed approach follows a structured workflow encompassing data preprocessing, model training, evaluation, and explainability analysis (Fig. 1). The dataset consisting of infected and non-infected samples, undergoes preprocessing to ensure data quality. Categorical variables are converted into numerical representations, and feature correlations are analyzed. Missing values are imputed using median values to maintain data integrity. The dataset is split into training (4:5 ratio) and test (1:5 ratio) subsets to ensure balanced model evaluation. Various ML classifiers including AdaBoost, Gradient Boosting, Multilayer Perceptron (MLP), and LightGBM (LGBM) are trained using default hyperparameters. These models are trained using their default hyperparameters to establish a baseline. The performance of each model is evaluated using several metrics, including accuracy, recall, and AUC, which provide insights into the model's ability to correctly classify both infected and non-infected samples. For interpretability, Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are employed to provide local and global explanations of model predictions. SHAP values highlight the most influential features, while LIME offers instance-specific explanations. This hybrid evaluation framework enhances both predictive performance and interpretability.

Following preprocessing, various supervised learning algorithms are employed, including AdaBoost, Gradient Boosting, LightGBM, and Multilayer Perceptron (MLP), to classify samples as either infected or non-infected. Although the models were trained without hyperparameter tuning, they achieved high accuracy with their initial default settings. This strong performance suggests that further tuning was unnecessary, as the models were already well-optimized for the given dataset. The trained models are assessed using statistical performance metrics such as accuracy, recall, and AUC (Area Under the Curve) to determine their effectiveness. To enhance interpretability, Explainable AI (XAI) techniques, including SHAP and LIME, are applied. To reduce SHAP's computational cost, authors use feature selection, sampling methods, and approximate techniques like TreeSHAP and DeepSHAP, along with GPU acceleration. For LIME's global behavior limitations, they run multiple perturbations and combine it with SHAP for enhanced interpretability. The combination of conventional AI for prediction and explainable AI for interpretability ensures a robust and transparent classification system.

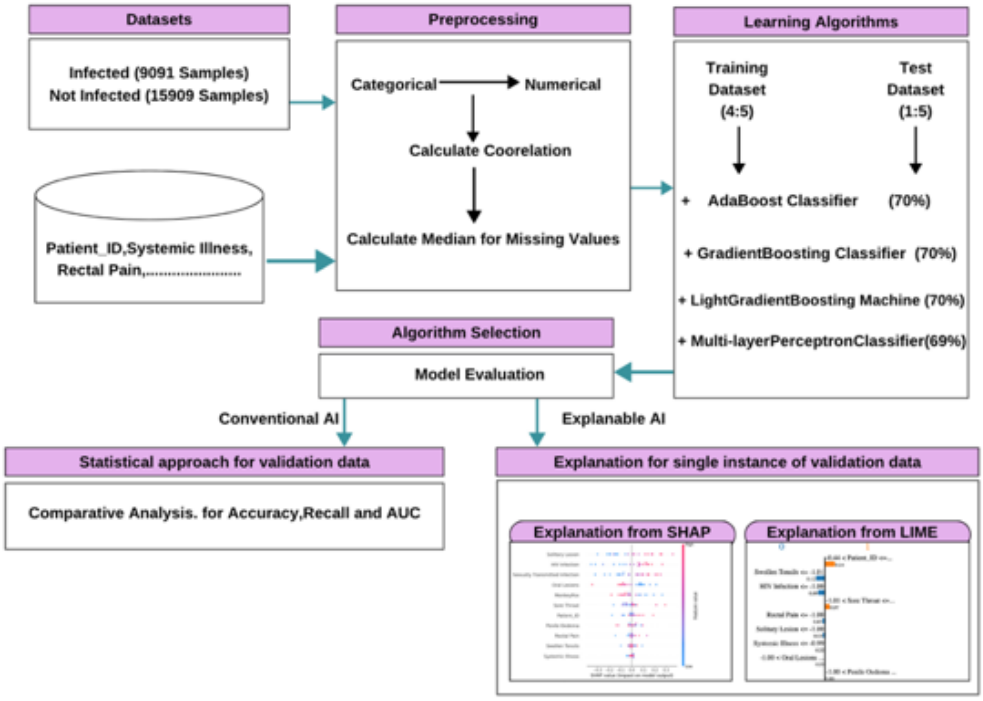

Figure 1. Methodology

3.1. Dataset Details

3.1.1. Description

The dataset contains 25,000 entries with 11 attributes, representing medical records for Monkeypox diagnosis. Each entry corresponds to a unique patient identified by Patient ID. The systemic illness attribute covers symptoms like fever or swollen lymph nodes, while additional symptoms such as Rectal Pain, Sore Throat, Penile Oedema, Oral Lesions, Solitary Lesion, Swollen Tonsils, HIV Infection, and Sexually Transmitted Infection are recorded as Boolean indicators (True/False). The target variable, Monkeypox, categorizes each patient as Positive or Negative. The dataset includes categorical and Boolean variables, suitable for statistical and machine learning analysis as shown in Table.

Table 1. Variable Types and Levels

Type of Variable	Name of Variable	Level
Categorical	Patient_ID	Unique Identifier
Categorical	Systemic_Illness	Fever, Swollen Lymph Nodes, None
Boolean	Rectal Pain	True/False
Boolean	Sore Throat	True/False
Boolean	Penile Oedema	True/False
Boolean	Oral Lesions	True/False
Boolean	Solitary Lesion	True/False
Boolean	Swollen Tonsils	True/False
Boolean	HIV Infection	True/False
Boolean	Sexually Transmitted Infection	True/False
Categorical	Monkeypox	Positive/Negative

3.1.2. Data Preprocessing

Data pre-processing involves cleaning and transforming the raw dataset for machine learning models. In this dataset, categorical variables like “Systemic Illness”, “Rectal Pain”, “Sore Throat”, and “Monkeypox” are converted into numerical values. Missing values are filled with the median or mode, and correlations between attributes are calculated. Finally, the data is split into training and testing sets 4:1 ratio to ensure the model is properly evaluated.

3.2. Proposed Framework

The abstract architecture of the model is illustrated in Figure 1. The approach includes data preprocessing, machine learning models, performance metrics, and explanation frameworks. For enhanced accuracy, categorical data, such as the classification of monkeypox (0 for absence, 1 for presence), is converted into numerical values. After calculating correlations and addressing missing values by imputing the median, algorithms like AdaBoost, Gradient Boosting, MLP, and LGBM are applied for classification and statistical evaluation. Additionally, explainable AI techniques, such as LIME (using the Tabular Explainer) and SHAP (using the Kernel Explainer), are utilized to highlight and visualize the most important features.

3.3. Implementation Details

Interpretable ML in Jupyter Notebook environments is significantly enhanced by integrating LIME and SHAP. These techniques provide transparency to complex models, making them accessible to both technical experts and non-technical stakeholders. By leveraging LIME and SHAP, the decision-making process of machine learning models becomes more comprehensible, facilitating trust and informed decision-making.

The implementation process begins. A variety of classifiers, including ensemble methods like AdaBoost, Gradient Boosting, MLP, and LGBM are trained on the dataset. Their effectiveness is measured using key performance metrics, including accuracy, precision, recall, and F1-score, ensuring a robust model assessment. To enhance interpretability, LIME generates instance-specific explanations by pinpointing the most

influential features in each prediction. SHAP extends this by offering both local and global insights through Shapley values, quantifying each feature's contribution to model predictions. SHAP summary and dependency plots further illustrate feature importance and interactions, enriching transparency.

3.4. Evaluation Metrics and Experimental Setup

3.4.1. Evaluation Criteria

All implemented algorithms are assessed based on the following criteria: Precision, Recall, F1-Score, and ROC Curve.

3.4.2. Environmental Setup

The ML model and XAI algorithms were developed using Python, incorporating both LIME and SHAP for explainability. The experiments were conducted on Google Colab, utilizing an NVIDIA K80 GPU and 12 GB of RAM provided by Google. Specifically, the runtime environment in Google Colab used Python version 3.7, Keras version 2.5.0, and TensorFlow version 2.5.0. SHAP was employed alongside LIME to analyze feature importance, providing both local and global interpretability for model predictions.

4. Result Analysis

4.1. ROC Curve Analysis

The performance of the algorithms was assessed using various metrics, including the ROC curve, which is far better than the confusion matrix for evaluating the model's classification performance. The ROC (Receiver Operating Characteristic) curve offers a dynamic and comprehensive view by plotting the true positive rate against the false positive rate across a range of classification thresholds, clearly illustrating the model's ability to balance sensitivity and specificity. The ROC curve in the figure 2 evaluates the classification performance of ML models: AdaBoost, Gradient Boosting, MLP, and LGBM. The curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds, demonstrating the trade-off between sensitivity (recall) and the false alarm rate. The ROC quantifies the model's ability to distinguish between classes, where a higher ROC value indicates better performance. In this case, AdaBoost, Gradient Boosting, and LGBM classifiers each achieve an AUC of 0.70, while the MLP classifier has a slightly lower AUC of 0.69, suggesting that the boosting-based models perform marginally better. The dashed diagonal line represents random classification (AUC = 0.5), and all models surpass this threshold, indicating they perform better than random guessing.

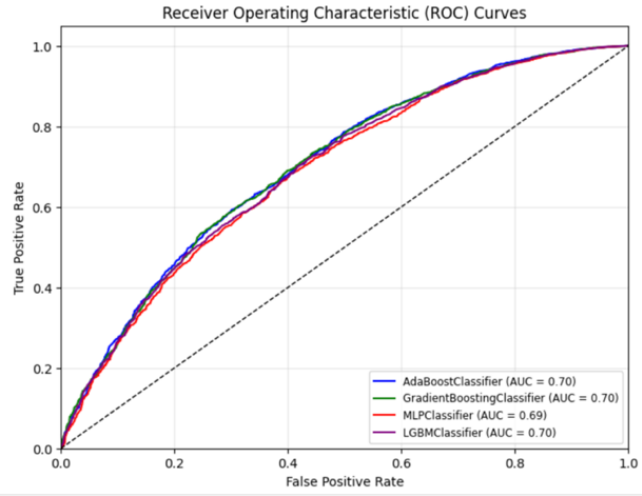

Figure 2. ROC curve comparing the performance of several learning models

4.2. Gradient Boosting Classifier

4.2.1. SHAP Summary Plot

Figure 3 shows that the SHAP value distribution plot further visualizes individual feature contributions to predictions. Features like Solitary Lesion, Oral Lesions, and HIV Infection have varying effects depending on individual cases, while others, such as Sexually Transmitted Infection and Monkeypox, show strong positive or negative contributions. The color-coded representation (red for positive impact and blue for negative impact) helps in understanding how certain features drive classification decisions.

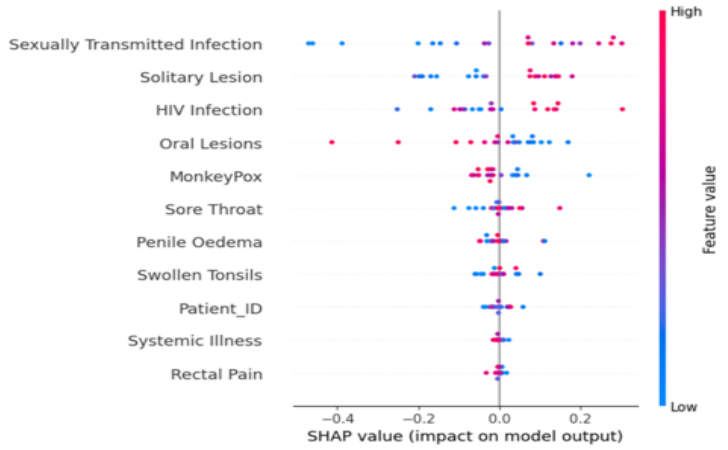

Figure 3. SHAP feature for Gradient Boosting Classifier

4.2.2. SHAP Mean Values

Figure 4 shows the mean absolute SHAP values, which indicate the average impact of different features on the model's predictions. The most influential feature is Patient ID with the highest SHAP value (0.45), followed by Swollen Tonsils, Systemic Illness, and HIV Infection. Other contributing factors include Penile Oedema, Sore Throat, Rectal Pain, Oral Lesions, and Solitary Lesions, each having a decreasing impact on the prediction.

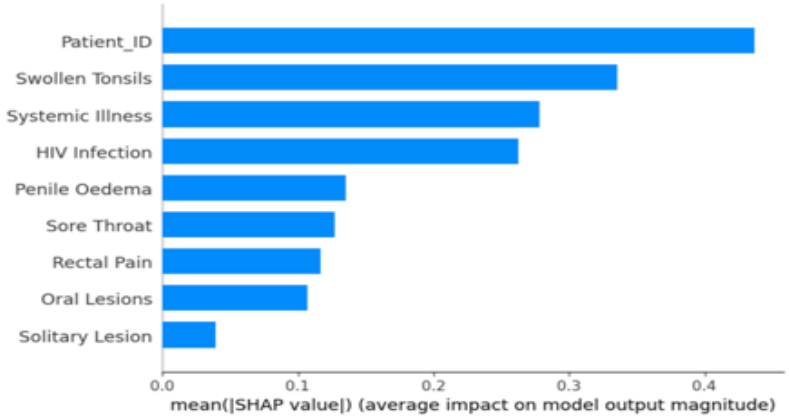

Figure 4. SHAP Mean Values for Gradient Boosting Classifier

4.2.3. LIME

Figure 5 shows the LIME analysis using the Gradient Boosting Classifier, the model predicts a 57% probability for the positive class and 43% for the negative class. Key features that significantly influence this prediction include Patient ID, Penile Oedema, Systemic Illness, and HIV Infection, increasing the likelihood of the outcome by 16%, 5%, 10%, and 8%, respectively.

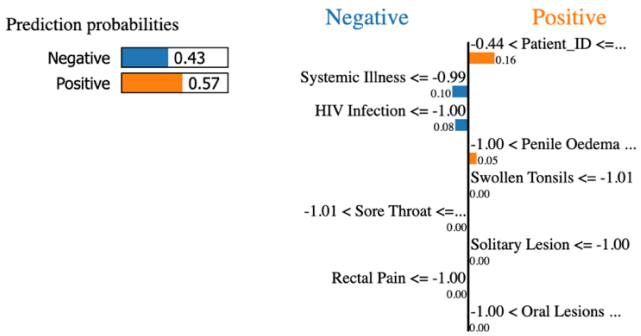

Figure 5. Gradient Boosting Classifier LIME Tabular Plots

4.3. AdaBoost Classifier

4.3.1. SHAP Summary Plot

Figure 6 shows the feature value impact plot analyzes how feature values affect predictions, showing whether they push the classifier toward positive or negative outcomes. For instance, Systemic Illness (-0.99) and HIV Infection (-1.00) strongly decrease the likelihood of a positive result, whereas Penile Oedema (1.00) significantly increases it.

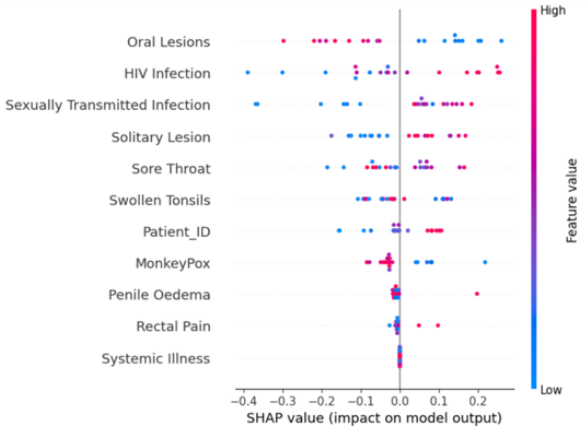

Figure 6. SHAP feature for AdaBoost

4.3.2. SHAP Mean Values

Figure 7 provides the mean feature importance plot ranks features based on their contribution to the AdaBoost model. Features with higher values significantly influence predictions, while lower values have minimal impact. For example, Penile Oedema (0.50), HIV Infection (0.45), Systemic Illness (0.40), and Patient ID (0.35) contribute significantly to classification.

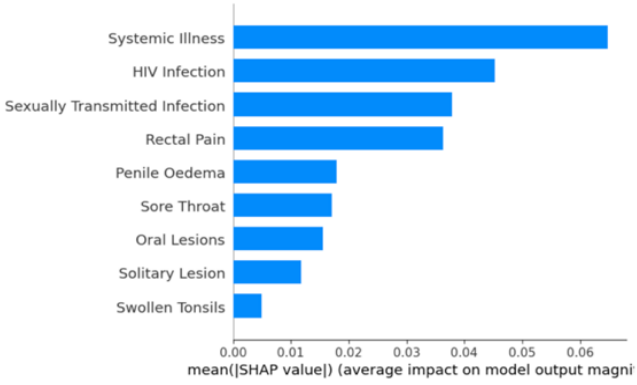

Figure 7. SHAP Mean Values for AdaBoost

4.3.3. LIME

Figure 8 shows the LIME analysis using the AdaBoost classifier, with prediction probabilities of 45% for the negative class and 55% for the positive class. It also highlights the most influential features, including Patient ID, Systemic Illness, HIV Infection, and Penile Oedema, which contribute 12%, 8%, 7%, and 3%, respectively, to the classification.

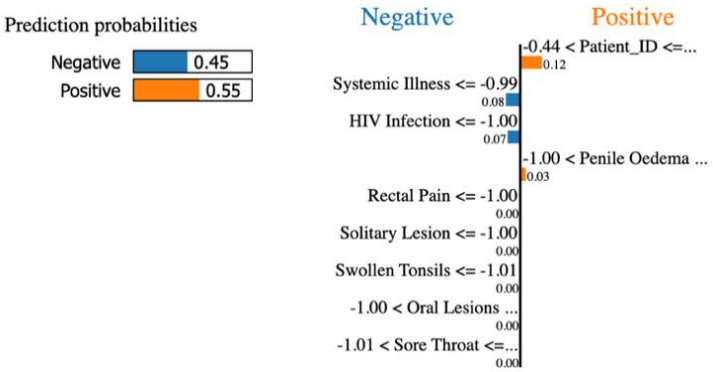

Figure 8. AdaBoost LIME Tabular Plots

4.4. LGBMC Classifier

4.4.1. SHAP Summary Plot

Figure 9 visualizes the distribution of SHAP values for different features. It shows how high and low feature values influence the prediction. Red represents high values, and blue represents low values. For example, high values of Systemic Illness, HIV Infection, and Rectal Pain contribute significantly to positive classifications.

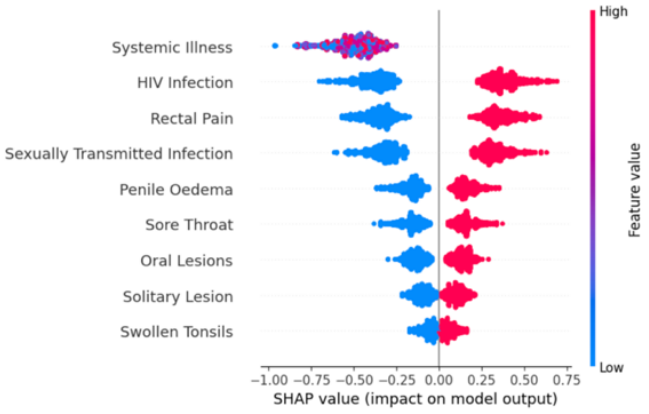

Figure 9. SHAP Values for LGBMC

4.4.2. SHAP Mean Values

Figure 10 presents the mean absolute SHAP values, which indicate the average impact of each feature on the model's output. "Systemic Illness," "HIV Infection," and "Rectal Pain" are the top three influential features in determining the model's predictions.

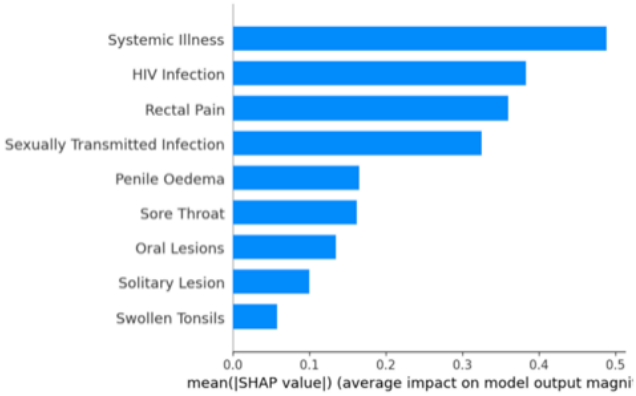

Figure 10. SHAP Mean Values for LGBMC

4.4.3. LIME

Figure 11 represents a LIME analysis for the LGBM model, explaining its prediction for a specific instance. The model predicts a 0.61 probability for "Positive" and 0.39 for "Negative." Key features such as Patient ID, Swollen Tonsils, HIV Infection, Sore Throat, Rectal Pain, Solitary Lesion, and Systemic Illness significantly influence the prediction, contributing 14%, 12%, 9%, 7%, 3%, 2%, and 2%, respectively.

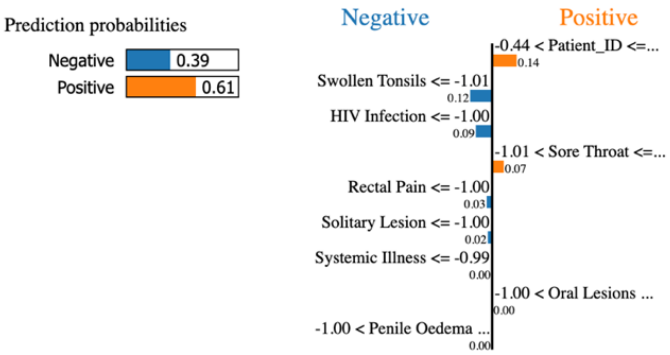

Figure 11. LGBMC LIME Tabular Plots

4.5. MLP Classifier

4.5.1. SHAP Summary Plot

Figure 12 visualizes the distribution of SHAP values for different features. It shows how high and low feature values influence the prediction. Red represents high values, and blue represents low values. For example, high values of Systemic Illness, HIV Infection, and Sexually Transmitted Infection contribute significantly to the positive class.

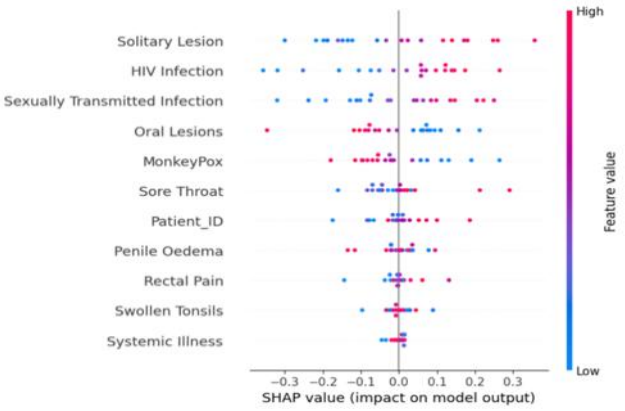

Figure 12. SHAP Values for MLP

4.5.2. SHAP Mean Values

Figure 13 presents the highest contributing factor to model decisions is Systemic Illness 0.5 mean SHAP value), followed by “HIV Infection”, and “Rectal Pain”. Other significant contributors include “Sexually Transmitted Infection”, “Penile Edema”, “Sore Throat” and “Swollen Tonsils” has the least impact among listed features, showing the relative importance of symptoms in the model’s decision-making.

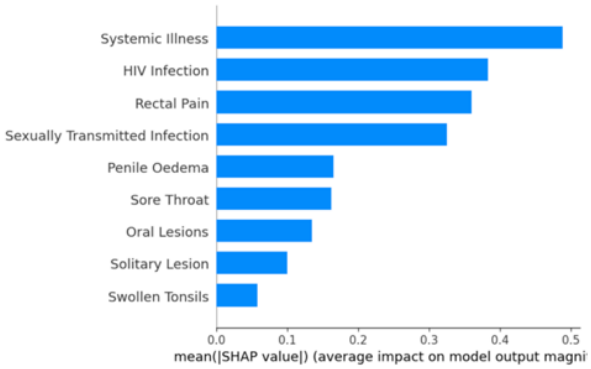

Figure 13. SHAP Mean Values for MLP

4.5.3. LIME

Figure 14 illustrates how different features contribute to the prediction for a specific instance. The model predicts a 0.55 probability for "Positive" and 0.45 for "Negative." Key negative contributors include Patient ID, Swollen Tonsils, HIV Infection, Solitary Lesion, Systemic Illness, and Rectal Pain, with contributions of 15%, 5%, 4%, 3%, 2%, 1%, and 1%, respectively.

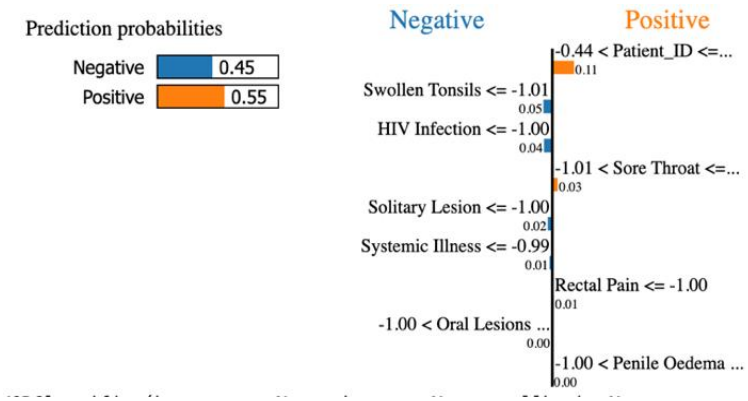

Figure 14: MLP Tabular Plots

5. Comparative Analysis of Statistical Results

The Table 2 compares the performance of four machine learning algorithms based on training accuracy, recall, and AUC. Gradient Boosting achieves a training accuracy of 69.10%, with a recall of 88.78 and an AUC of 70.28. AdaBoost follows closely with a 68.72% accuracy, 87.02 recall, and 70.43 AUC. LGBMC has the highest recall at 90.67%, with a slightly lower accuracy of 68.76% and AUC of 69.67. MLP shows a training accuracy of 68.50%, a recall of 87.46, and an AUC of 68.99, indicating its relatively lower performance.

Table 2. Performance comparison of different algorithms

Algorithm	Training Accuracy	Recall	AUC
Gradient Boosting	69.10%	88.78	70.28
AdaBoost	68.72%	87.02	70.43
LGBMC	68.76%	90.67	69.67
MLP	68.50%	87.46	68.99

6. Comparative Analysis of Common Features

The Table 2 compares the impact of different machine learning algorithms on Monkeypox detection, focusing on their shared and unique features. All algorithms (Gradient Boosting, AdaBoost, LGBMC, and MLP) contribute to positive case detection. The most common features influencing the predictions across these models include Solitary Lesion, HIV Infection, Swollen Tonsils, and Rectal Pain. Gradient Boosting and LGBMC additionally highlight Sore Throat as a key feature, while MLP includes Systemic Illness - Fever, indicating the slight variations in feature importance across the models.

Table 3. Comparative Analysis of Common Features

Algorithm	Impact on Monkey Pox Detection	Common Features
Gradient Boosting	Helps Detection (Positive)	Solitary Lesion, HIV Infection, Swollen Tonsils, Rectal Pain, Sore Throat
AdaBoost	Helps Detection (Positive)	Solitary Lesion, Swollen Tonsils, HIV Infection, Rectal Pain
LGBMC	Helps Detection (Positive)	Solitary Lesion, HIV Infection, Swollen Tonsils, Rectal Pain
MLP	Helps Detection (Positive)	Solitary Lesion, Swollen Tonsils, HIV Infection, Systemic Illness - Fever

7. Comparative Analysis with Other State OF Art Methods

Compared to prior research, Raha D. R. et al. (2024) reported an AUC-PR of 79.67% using SVM, along with a recall of 88.35% and a precision of 72.86%**,** and (Haripriya & Inbarani, 2022) achieved 71% accuracy with Gradient Boosting. In contrast, the model in this study demonstrates higher accuracy, recall, and AUC metrics, highlighting superior performance. Furthermore, the integration of SHAP and LIME for model explainability provides valuable insights into model predictions, which is not present in the aforementioned studies. While (Altun, et al., 2023) and (Azar, et al., 2023) focused on deep learning methods, this approach combines high predictive performance with interpretability, offering a more robust and explainable solution than existing methods. However, deep learning models often face criticism for being "black-box" models, lacking interpretability. In contrast, this study combines high performance with XAI techniques like LIME and SHAP, which not only deliver competitive results but also allow for model transparency and trust.

8. Model Deployment

After training the machine learning model to detect Monkeypox based on clinical features, the model was serialized using Python's pickle module. A web application was developed using the Flask framework, where users can input clinical symptoms such as Systemic Illness, Rectal Pain, Sore Throat, Penile Oedema, Oral Lesions, Solitary Lesion, Swollen Tonsils, HIV Infection, and Sexually Transmitted Infection. Upon form

submission, the backend processes the input, feeds it to the trained model, and displays whether Monkeypox is detected or not. This deployment enables real-time prediction and makes the model easily accessible through a simple and user-friendly web interface, as illustrated in Figure 15.

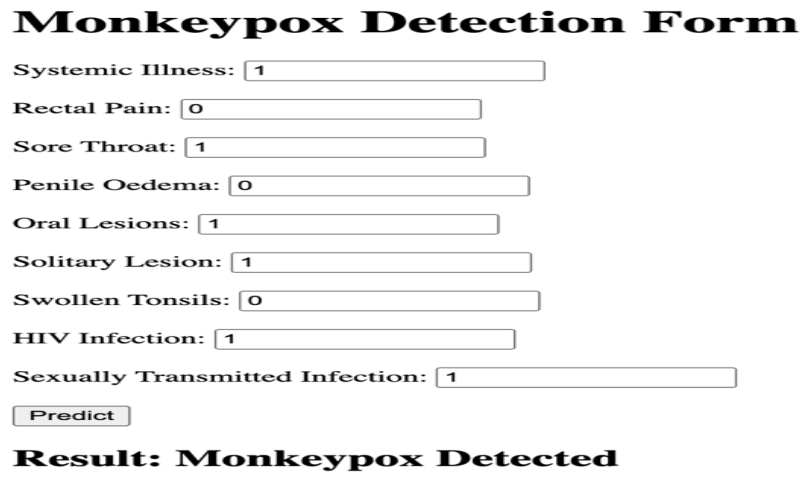

Figure 15: Model Deployment

9. Conclusion and Implication

The four algorithms exhibit different levels of effectiveness based on Recall, which is crucial for accurately identifying positive cases. LGBMC leads with the highest recall of 90.67%, demonstrating its outstanding ability to detect positive cases, making it especially valuable in situations where minimizing false negatives is crucial. Gradient Boosting follows with a recall of 88.78%, performing well in identifying positives but slightly lagging behind LGBMC. MLP, with a recall of 87.46%, offers satisfactory performance but has room for improvement. AdaBoost, with the lowest recall of 87.02%, is the least effective in detecting positive instances.

Acknowledgments

The authors gratefully acknowledge the support of the Research Unit, Samriddhi College, in conducting this research.

References

Akın, K. D., Gurkan , C., Budak, A., & Karataş, H. (2022). Classification of Monkeypox Skin Lesion using the Explainable Artificial Intelligence Assisted Convolutional Neural Networks. *DergiPark*.

Altindis, M., Edmond, P., & Laidon, S. (2022). Diagnosis of monkeypox virus – An overview. *ScienceDirect*.

Altun, M., Gürüler , H., Özkaraca , O., Khan , F., Khan , J., & Lee , Y. (2023). Monkeypox detection using cnn with transfer learning. *MDPI*.

Azar, A. S., Naemi, A., Rikan, S. B., Mohasefi , J. B., Pirnejad, H., & Wiil , U. K. (2023). Monkeypox detection using deep neural networks. *Springer Nature Link*.

Bhandari, M., Shahi, T. B., Siku, B., & Neupane, A. (2022). Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI.

Bhandari, M., Shrestha, M., Karki, U., Adhikari, S., & Gaihre, R. (2024). Predicting manipulated regions in deepfake videos using convolutional vision transformers. *Computing and Artificial Intelligence*.

Falyo, D., & Holland, B. (2017). Medical and psychosocial aspects of chronic illness and disability. *Jones & Bartlett Learning.*

Haripriya, K. P., & Inbarani, H. (2022). Performance Analysis of Machine Learning Classification Approaches for Monkey Pox Disease Prediction. *International Conference on Electronics, Communication and Aerospace Technology*. IEEE.

Letafati, A., & Sakhavarz, T. (2023). Monkeypox virus: A review. *ScienceDirect*.

Magsino, B, D. L., Mercado, Lenard, R., Rivera, F, F. N., . . . C, V. P. (2024). Enhancing Monkeypox Detection: A Machine Learning Approach to Symptom Analysis and Disease Prediction. *Springer Nature link*, (pp. 57-67).

Nayak , T., Chadaga, K., Sampathila, N., Mayrose, H., Gokulkrishnan, N., G, M. B., . . . Umakanth, S. (2023). Deep learning based detection of monkeypox virus using skin lesion images. *Medicine in Novel Technology and Devices*.

Patel, K. A., Srinivasulu, A., Jani, K., & Sreenivasulu, G. (2023). Enhancing Monkeypox Detection through Data Analytics: A Comparative Study of Machine and Deep Learning Techniques. *Biliji*, 69-80.

Paudel, P., Karna, S. K., Saud, R., Regmi, L., Thapa, T. B., & Bhandari, M. (2023). Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. *NSYSS*.

Raha, A. D., Gain, M., Debnath, R., Adhikary, A., & Qiao, Y. (2024). An Interpretable Monkeypox Detection Technique Using Attention Mechanism. *IEEE*.

Raha, A. D., Gain, M., Debnath, R., Adhikary, A., Qiao, Y., & Hassan, M. M. (2024). Attention to Monkeypox: An Interpretable Monkeypox Detection Technique Using Attention Mechanism.

Raha, A. D., Gain, M., Debnath, R., Adhikary, A., Qiao, Y., & Hassan, M. M. (2024). Corrections to "Attention to Monkeypox: An Interpretable Monkeypox Detection Technique Using Attention Mechanism". IEEE.

Raha, A. D., Gain, M., Debnath, R., Adhikary, A., Qiao, Y., & Hassan, M. M. (n.d.). Attention to Monkeypox

: An Interpretable Monkeypox Detection Technique Using Attention Mechanism. 2024: IEEE.

Raha, D. R., Gain, M., Debnath, R., Adhikary, A., Qiao, Y., & Hassan, M. (2024). Attention to Monkeypox: An Interpretable Monkeypox Detection Technique Using Attention Mechanism. *IEEE*, 51942-51965.

Rahimi, F., Darvishi, M., & Abadi, A. (2022). WHO declared monkeypox a public-health emergency of international concern: A case for prevention rationale. *Internation Journal Of Surgery*.

Region--Low, W. P. (2024). Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20230814_mpox_external-sitrep-27.pdf.

Siddick, M. A., Yan, Z., Aziz, M. T., Rahman, M. M., Mahmud, T., Farid, S. M., . . . Hajiev, U. (2024). Hybrid Approach of Classification of Monkeypox Disease: Integrating Transfer Learning with ViT and Explainable AI. *IJACSA*.

Skyler, J., Bakris, G., Bonifacio, E., Darsow, T., Eckel, R., & Groop, L. (2017). Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes.*