International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

Automated Monkeypox Classification Using ReliefF Feature Selection and Stacked Ensemble Learning with Optimized Performance Metrics

Wameedh Raad Fathel

Ministry of Education, General Directorate of Education in Nineveh, Iraq. E-mail: wamed81@gmail.com

Abstract **- The accurate and in time detection of Monkeypox using medical images is crucial for effective disease management. In this paper an improved classified system that integrates texture, local binary pattern (LBP), and statistical features with advanced feature selection and ensemble learning has been proposed. ReliefF algorithm was used as feature selection, maintaining the top 70%of features, and hyperparameter optimization has been applied to the Support Vector Machine (SVM) classifier. Additional algorithms: Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, and a Stacked Ensemble model were also used as classifier. Different metrics like accuracy, precision, recall, and F1-score. The highest accuracy of 92.5%. Confusion matrix and the area under the Receiver Operating Characteristic (ROC) curve (AUC) visually demonstrate model performance. The overall results confirmed that integrating feature selection with ensemble learning can significantly improve the robustness and reliability of automated Monkeypox detection.**

Keywords: Image Classification, Feature Selection, ReliefF, Ensemble Learning, Automated Detection.

I. INTRODUCTION

Monkeypox (MPox) is a viral zoonotic disease caused by the monkeypox virus, and it is a member of the Orthopox virus genus. It can be diagnosed clinically by some symptoms like fever, lymphadenopathy, and characteristic skin lesions, which can resemble other dermatological conditions such as chickenpox or smallpox [1,2] Like many diseases early and accurate detection consider as a very important thing to effectively manage it control potential outbreaks. This disease traditional diagnosis relays on clinical examination and laboratory testing, which include polymerase chain reaction (PCR) and serological assays [3]. However, these approaches are time-consuming, require specialized equipment, and may not always be available in resource-limited settings.

Recently, the Modern advancement in medical imaging and artificial intelligence (AI) opened new avenues in automated detection of the disease from clinical images Dermoscopic imaging, used widely in in dermatology, high-resolution visualization of skin lesions, captures subtle textural and structural patterns that can assist in differentiating skin conditions [4]. Using machine learning (ML) and deep learning approaches resulted in a superior result in automated classification of skin lesions, reducing errors of diagnostic and support clinical decision-making [5,6].

Despite of all these advances, the challenges in detecting the disease remain. Skin lesion images often show high intra-class variability and inter-class similarity, which make using robust feature extraction and selection very essential for accurate classification. Redundant or irrelevant features can decrease the performance of the model, increase computational cost, and reduce the interpret ability of the system [7]. Therefore, feature optimization techniques are used increasingly in order to identify the most discriminative features and reduce dimensionality.

In this study, an enhanced framework for Monkeypox detection using augmented dermo_scopic images has been introduced. Multi-scale texture features, including Gray Level Co-occurrence Matrix (GLCM), multi-radius Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG), complemented by statistical descriptors have been extracted. In order to reduce redundancy and highlight informative patterns, feature selection was performed using ReliefF and correlation-based filtering. Multiple ML classifiers such as Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN) were used and a stacked ensemble was employed to benefit from complementary strengths. The evaluation of the model has been done using many metrics such as, accuracy, confusion matrices, and ROC curves. This work contributes a robust and interpretable pipeline for automated dermoscopic image classification, providing a foundation for future AI-assisted diagnostic tools in clinical dermatology [8][9][10].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

II. LITERATURE REVIEW

To achieve reliable monkeypox classification, recent studies have concentrated on improving feature extraction, model optimization, and generalization. Feature extraction was a major area of investigation, especially in combining pre-trained convolutional neural networks (CNNs) with attention mechanisms. Mousa *et al.* focused on the use of attention-enhanced CNNs and transformer architectures, which enhance the detection by emphases critical lesion regions [11]. Chen *et al.* used an improved SE-InceptionV3 model with channel attention, using demonstrating that selective feature representation in distinguishing subtle lesion patterns is very important [12]. Zi proved that careful selection of color space representations combined with CNNs can help in enhancing the performance of the model [13].

Pre-trained architectures such as DenseNet, ResNet, VGG, and Inception have been widely applied, providing robust feature representations for classification [14]. Arafa *et al.* have introduced multi-stage classification system (MSCADMpox), his ssem combined handcrafted features with deep learning outputs, further demonsted the benefit of multi-level feature extraction [15].

The other critical things for enhancing classification performance and reducing overfitting are feature selection and optimization. Many algorithms, such as Grey Wolf Optimization (GWO) and Particle Swarm Optimization (PSO), have been used to select useful features from deep learning outputs [16,17]. Savaş, integrate CNN models with ensemble strategies showing that ensemble and two-stage optimization approach, improved the robustness of the model [16]. Hyperparameter tuning has also been applied to enhance model generalization across datasets [18].

The strategies of data augmentation and balancing are critical in addressing limited and imbalanced datasets. Mousa *et al.* [11] and Mishra *et al.* [19] demonstrated that the robustness of the model can be improved using image augmentation, such as rotation and edge enhancement. Ciran and Özbay [14] balanced the dataset using oversampling methods, which result in more consistent performance of the model. Islam *et al.* [20] highlighted the effectiveness of using deep learning and traditional machine learning approaches together to improve the classification of lesion.

The metrics used to evaluate systems, often include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). From the literature, reliable results in differentiating monkeypox from other skin lesions has acheived by using combined of advanced CNN architectures with attention mechanisms, feature selection, and data augmentation [1–20]. All the literature emphasizes the importance of multi-stage feature extraction, careful feature selection, and robust model optimization to obtain accurate and generalizable monkeypox detection system.

III. METHODS

3.1 Dataset and Preprocessing

In this study a dermo_scopic image dataset consists of 3,192 images, these images contain 1,428 images of Monkeypox lesions and 1,764 images represent other skin conditions. Preprocessing have been applied beginning of resizing the images to become 256 × 256 pixels to ensure spatial uniformity, then to reduce color-induced variability across the dataset, each image converted to grayscale [21].

The final part of preprocessing is the normalization of pixels values to scale them to [0,1] range. This step minimized the intensity variation between images and improved the subsequent feature extraction algorithms performance. The standardization of dataset has made the consistent extraction of texture easy.

3.2 Feature Extraction and Description

To comprehensively characterize Monkeypox lesions and other skin conditions, a multi-modal feature extraction approach was implemented. The Method integrated texture, local pattern, and intensity-based statistical features, which provide both local and global representations of lesion morphology, texture, and contrast. Features have been selected to capture discriminative properties related to biomedical classification tasks [22].

Statistical Texture (GLCM):

To identify spatial relationships of pixel intensities, captur textural uniformity, regularity, and structural variation within lesions, Gray-Level Co-occurrence Matrix (GLCM) has been used. Four main attributes have been extracted which are: Contrast, Correlation, Energy, and Homogeneity. The contrast measures the variation of the intensity between neighboring pixels, Correlation quantifies linear dependency between gray levels, Energy identifies the uniformity of the textural, and Homogeneity evaluates the closeness of distribution to the diagonal of the co-occurrence matrix [23].

Local Texture Patterns (LBP):

Micro-textures have been encoded using Local Binary Patterns (LBP) by using the comparison of each pixel with its neighbors, which make them resistance to the changes of illumination and fluctuations of local intensity .Multi-scale LBP has been computed at radii 1, 2, and 3 with a cell size of 32 × 32 pixels, producing a binary patterns histogram that

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

captures fine-to-coarse texture details common in Monkeypox lesions, such as pustules and scabs [24].

Statistical Features:

First-order statistical features summarized global intensity characteristics. Mean intensity refer to the total brightness of the lesion and standard deviation shows the contrast variation across the lesion region. These features are essential in distinguishing the differences of pigmentation and variations of lighting between lesion types [25]. Table 1 illustrates the feature extraction.

Table 1: Extracted Image Features

Feature Type	Technique	Extracted Attributes	Description	Biomedical Relevance
Texture (Statistical)	GLCM	Contrast, Correlation, Energy, Homogeneity	Quantifies spatial relationships of pixel intensities	Differentiates the roughness of lesion surface and the variation of the structure between Monkeypox and other skin conditions
Texture (Local Pattern)	LBP	Histogram of binary patterns (cell-wise)	Encodes micro-textures; robust to illumination changes	Captures the irregularities of local texture typical in Monkeypox lesions
Statistical (Intensity-based)	First-order Statistics	Mean intensity, Standard deviation	Measures overall brightness and contrast	Represents pigmentation of the lesion, contrast, and differences of the lighting.

Table 2 illustrates the most Mathematical Formulation of Features.

Table 2: Key formulas for the extracted features

Feature	Formula	Interpretation
Contrast	$\sum_{i,j}(i,j)\cdot P(i,j)$	Measures intensity variation between neighboring pixels.
Correlation	$\sum \frac{(i-\mu_i)\ (j-\mu_j)\,P(i,j)}{\sigma_i\sigma_j}$	Measures gray levels linear dependency
Energy	$\sum_{i,j} P(i,j)^2$	Reflects the uniformity of the textural.
Homogeneity	$\text{hom}=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\lvert i-j\rvert}$	i-j
LBP	$LBP_{P,R}=\sum_{P=0}^{P-1}\delta(g_p-g_c)2^p, \delta(x)$ $=\begin{cases}1, x\geq 0\\ 0, x<0\end{cases}$	Describes the patterns of local binary based on neighboring pixels.
Mean Intensity	$\mu=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}I(i,j)$	Average brightness of the image.
Standard Deviation	$\sigma=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}I(i,j)^2}$	Measures the variation of global contrast.

3.3 Feature Extraction Parameters

The feature computation has been standardized with uniform parameters to ensure that all images have the ability of reproduce and consistence as shown in Table 3.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

Table 3: Feature Extraction Parameters

Parameter	Value	Description
Image Size	256 × 256 pixels	All images resized for uniform feature extraction.
GLCM Offset	[0 1]	Horizontal neighbor pixel comparison.
GLCM Symmetric	True	Ensures texture uniformity in both directions.
LBP Cell Size	32 × 32 pixels	Defines local region for texture histogram.
Color Space	Grayscale	Simplifies computation and emphasizes texture.

3.4 Feature Vector Composition

Table 4 illustrates the final feature vector for each image integrating texture, local pattern, and statistical information, which result in a rich and discriminative representation suitable for classification.

Table 4: Feature Vector Composition

Feature Group	No. of Features for each image	Example Attributes
GLCM Features	4	Contrast, Correlation, Energy, Homogeneity
LBP Features	59 (approx., depends on cell size)	LBP histogram bins
Statistical Features	2	Mean, Standard Deviation
Total Features per Image	≈ 65	Combined feature vector used for classification

This multi-modal representation ensures that both patterns of structural lesion and characteristics of the intensity have been captured, this provides a robust subsequent feature selection foundation and development of machine learning model.

3.5 Feature Selection

The process of initial multi-modal feature extraction generated a high-dimensional feature set of 3,782 features, combined texture descriptors (GLCM), local pattern features (LBP), and first-order statistical measures. Despite of the comprehensive, this high dimensional data can cause in redundancy, irrelevant information, and increased computational burden, which can adversely affect the performance and generalization of the classifier. To overcome these challenges, a systematic feature selection pipeline was implemented, which aims to maintain the most informative and discriminative features and reduce dimensionality as illustrated in table 5

ReliefF: is a filter-based supervised feature selection method which is widely used and ranks features depending on their ability to distinguish between different classes. It is popular in applications of biomedical imaging due to its effectiveness in in handling high-dimensional and noisy data. The algorithm estimates each feature relevance by evaluating its effective on differentiates between instances of the same and different classes within their local neighborhoods. ReliefF identifies a set of nearest neighbors from the same class (nearest hits) and from different classes has (nearest misses), and then feature weights updates depends on their feature values differences. In this study, an importance score has been assigned for each feature which identifies its contribution to class separation. The top 70% of features arranged by their importance scores have been maintained for subsequent analysis. By this selection strategy only the most discriminative features are preserved, improving the performance of downstream classifiers in distinguishing Monkeypox lesions from other skin conditions. Mathematical formulation **for a feature f**.

$$W[f] = W[f] - \frac{diff(f, x_i, hit)}{m} + \frac{diff(f, x_i, miss)}{m}$$

Where: xi is the current sample, **hit** = nearest neighbor of the same class, **miss** = nearest neighbor of a different class, m = number of samples

- **Correlation filtering:** A relevance-based correlation filtering approach was employed, in order to further refine the feature subset and reduce redundancy. The aim of this step is to eliminate features that convey overlapping information, thereby enhancing the efficiency and interpretability of the feature set. Among the selected features , Pairwise Pearson correlation coefficients has been computed , and each pair with correlation coefficient which is more than 0.9 was considered highly

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 11, pp 175-186, November-2025
https://doi.org/10.47001/IRJIET/2025.911020

correlated and potentially redundant [27].The pipeline removes one feature from each highly correlated pair , by this method it reduce the multicollinearity, decrease the overfitting risks , and ensures that the contribution of the retained features is independent and complementary information to the classification process. This procedure very important in high-dimensional imaging data, where spatially adjacent or overlapping regions often produce strongly correlated feature values.

The process of combined feature selection has successfully reduced the features from 3,782 to 2,647, resulting on create a more compact and informative feature set. In addition to the computational load decrease, the reduction also improved the efficiency and robustness of the classifier. To ensure reproducibility and facilitate future experimentation. Tinal set of selected features was saved in CSV format, which then be able to re-use consistently in different models and datasets. This structured feature selection pipeline provides a robust foundation for machine learning models training while maintaining the original data integrity and discriminative capacity.

Table 5: Feature Selection Summary

Feature Selection Method	Purpose	Settings / Outcome
ReliefF	Rank features by relevance	Top 70% selected (2,647 of 3,782 features)
Correlation Filtering	Remove redundant features	Features with correlation >0.9 removed
Final Selected Features	Compact and discriminative set	Saved to CSV for reproducibility

3.6 Model Development

After the process of multi-modal feature extraction and subsequent feature selection, the data set has been represented by improved of 2,647 discriminative features. The feature set resulted has provided a compact informative representation of lesion morphology, texture, and intensity characteristics suitable for classification tasks.

The data set has been divided into 70/30 training-testing subsets using stratified sampling to preserve the distribution of the original class. Ensuring the proportionally represented of both Monkeypox and other skin condition classes. Within the training set, 5-fold cross-validation has been used to optimize each classifier hyperparameters. This approach partitioned the training data to five folds, iteratively trained on four folds and validated on the fifth. Hyperparameter selection based on cross-validation performance ensured the generalization of the models to unseen data. In order to prevent bias occur because of differing feature scales, all features have been normalized to the range of [0,1] before the training process.

Support Vector Machine (SVM)

The SVM classifier employed a **radial basis function (RBF) kernel**, which is effective for high-dimensional feature spaces such as the 2,647-feature representation used here [3]. Hyperparameters, including the penalty parameter C and kernel width γ, were optimized using grid search in combination with cross-validation. The SVM aims to find the optimal hyperplane that maximizes the margin between Monkeypox and non-Monkeypox samples in the feature space. This method is well-suited for biomedical image datasets with high-dimensional, multi-modal feature representations.

The equation of a linear SVM is as follows:

$$f(x) = w.x + b$$

Where x represents the potential, w the orthogonal weights, and b the polarization term. The expected elegance of x is given by f(x), which assigns a strong or bad label. The hyperplane f(x) is used to separate statistics, with f(x) = zero as the bounds of choice.

- **Random Forest (RF)**

The Random Forest classifier involved 100 decision trees; each one has been trained on bootstrapped subsets of the training data. To evaluate the splits of the nodes, Gini impurity criterion was used. RF has a robustness against overfitting and can handle heterogeneous features of this study, which capture complex non-linear relationships between texture, pattern, and statistical descriptors [4]. The ensemble approach gathers predictions from all trees, resulting in a stable and reliable classification outputs.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

$$\hat{y} = mod\{h_t(x)\}_{t=1}^{T}$$

Where $h_t(x)$ is the prediction of the t-th tree, T=200 and $\hat{y}$ is the final class by majority voting.

- **k-Nearest Neighbors (kNN)**

The kNN classifier assigns class labels based on the majority vote of **k nearest neighbors** in the 2,647-dimensional feature space [5]. The optimal value of k was determined through cross-validation, and the **Euclidean distance metric** was used to measure feature similarity. kNN is particularly effective when class boundaries are irregular, as it directly leverages local neighborhood information in the high-dimensional space. These distances are commonly measured using the Euclidean distance (d) as equation 1.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where, x and y represent feature vectors for two data points, while n is the number of features. Based on the calculated distances, the nearest neighbor for the data point is selected. The number of neighbors considered is determined by the user-specified parameter (k) [31].

- **Logistic regression (LR)**

LR is a set of rules that uses equations to establish the relationship between input functions and the final binary result. At the heart of LR is the logistic feature, often called the sigmoid feature. This function takes the output of a linear regression and plots it for a possible price between 0 and 1. The logistic function equation is:

$$\hat{y} = \begin{cases} 1, \sigma(W^T X + b) \geq 0.5 \\ 0, \sigma(W^T X + b) < 0.5 \end{cases}$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Where, $\sigma(z)$is the sigmoid function, W is the weight vector, and b is the bias.

- **Stacked Ensemble (SE)** combines multiple base classifiers to improve predictive performance by leveraging their complementary strengths. In this study, the base models include **SVM, Random Forest, and KNN**, and their outputs are used as inputs to a **meta-learner**, implemented as **logistic regression**. The meta-learner learns to optimally weight the predictions of the base models to generate a final prediction.

Mathematically, the stacked ensemble prediction for an input feature vector x is expressed as:

$$\hat{y} = f_{meta}\,(\hat{y}\,SVM(x), \hat{y}\,RF(x), \hat{y}\,KNN(x)$$

Where $\hat{y}\,SVM(x), \hat{y}\,RF(x), \hat{y}\,KNN(x)$ are the prediction of the individual base classifiers and f_{meta} is the meta learner function that outputs the final class $\hat{y}$.

This approach reduces the bias and variance inherent in individual models and often results in improved overall accuracy, as demonstrated in our experiments [28, 29]. Table 6 illustrated the classifier hyperparameters for all models.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 11, pp 175-186, November-2025
https://doi.org/10.47001/IRJIET/2025.911020

Table 6: Classifier Hyperparameters

Classifier	Hyperparameters	Description
SVM (RBF)	C = 0.1–10, γ = 0.001–0.1	Optimized via cross-validation for high-dimensional features; RBF kernel handles non-linear relationships.
Random Forest	100 trees, Gini criterion	Ensemble method that reduces overfitting and handles heterogeneous features.
kNN	k = 3–9, Euclidean distance	Non-parametric method that leverages nearest neighbors in feature space for classification.
Logistic Regression	Regularization (L1/L2), C = 0.01–10	Linear model for binary classification; regularization prevents overfitting and manages high-dimensional features.

3.7 Model Evaluation

All classifiers were evaluated on the **independent testing set (30% of the dataset)** using multiple performance metrics as shows in table 7. Accuracy measures the proportion of correctly classified samples, while precision quantifies the reliability of positive predictions. Recall evaluates the classifier's sensitivity to detect Monkeypox lesions, and F1-score provides a balance between precision and recall. The **area under the receiver operating characteristic curve (AUC)** measures the classifier's overall discrimination ability between classes [30].

Table 7: Model Evaluation Metrics

Metric	Definition	Biomedical Relevance
Accuracy	Correct predictions / Total predictions	Overall diagnostic performance
Precision	True positives / Predicted positives	Reliability of Monkeypox detection
Recall	True positives / Actual positives	Sensitivity in detecting lesions
F1-Score	Harmonic mean of precision and recall	Balances sensitivity and specificity
AUC	Area under ROC curve	Measures ability to discriminate between classes

IV. RESULTS

4.1 Dataset and Feature Selection

The dataset used in this study **Monkeypox Skin Lesion Dataset (MSLD v2.0)**, which includes **1,428 Monkeypox** and **1,764 Others** images, totaling **3,192 samples as shows in figure 1**. Images were resized to **256×256 pixels** and converted to grayscale to standardize inputs and focus on texture over color. Extracted features by applied a multi-modal feature extraction approach was implemented. The approach combined **texture, local pattern, and intensity-based statistical features**, providing both global and local representations of lesion morphology, texture, and contrast [21]. Features were selected to capture discriminative properties relevant for biomedical classification tasks. Then dimensionality and select the most discriminative features, we applied the **ReliefF algorithm**, which ranks features based on their relevance to the class labels. Out of 3,782 extracted features, the top 2,647 were retained for model training, improving computational efficiency and classifier performance as illustrate in Table 8.

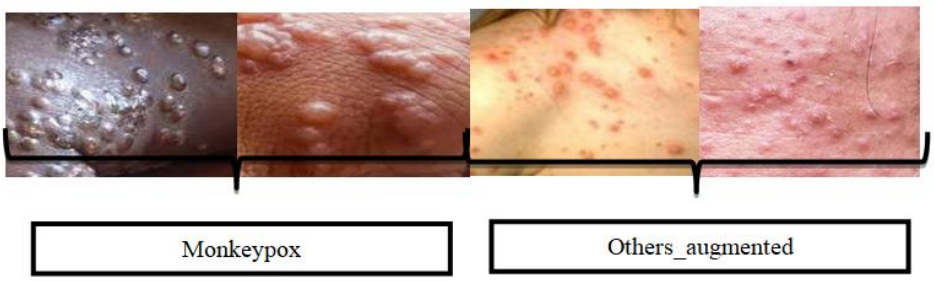


Figure1: Illustrate samples of data set

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

Table 8: Dataset Summary and Feature Selection

Dataset	Count	Percentage
Monkeypox_augmented	1428	44.74%
Others_augmented	1764	55.26%
Total Samples	3192	100%
Features Extracted	3782	-
Features Selected (ReliefF)	2647	70%

The dataset showed a moderate class imbalance, with higher of samples number in the Others_augmented category. To address the redundancy of the feature and reduce complexity of the computation operation, the feature selection process maintained the top 70% of the most relevant features. This strategy preserved the discriminative information which necessary for classification accuracy while enhancing the efficiency and generalization of the model.

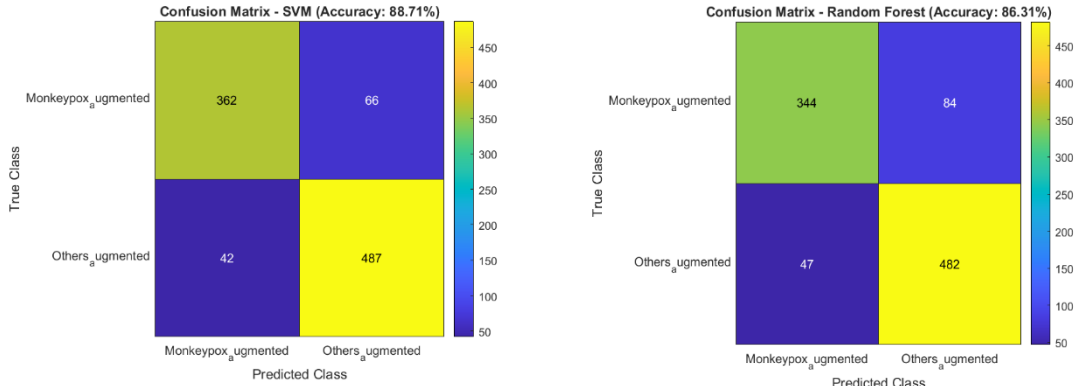
4.2 Classification Performance

Table 9 illustrates the evaluating results of four classifiers, RF, KNN, and LR used for training on the selected features. The evaluation metrics include accuracy, precision, recall, and F1-score.

Table 9: Performance Metrics of All Classifiers

Classifier	Accuracy	Monkeypox F1-score	Others F1-score	Precision (Monkeypox / Others)	Recall (Monkeypox / Others)
SVM	0.8871	0.8702	0.9002	0.8458 / 0.9206	0.8960 / 0.8807
RF	0.8631	0.8400	0.8804	0.8037 / 0.9112	0.8798 / 0.8516
KNN	0.8558	0.8253	0.8772	0.7617 / 0.9319	0.9006 / 0.8286
LR	0.8265	0.8005	0.8466	0.7780 / 0.8658	0.8243 / 0.8282

Among all models used in evaluation , SVM classifier has obtained the highest performance, with an overall accuracy of 88.71%, exhibiting high recall for the Monkeypox_augmented class (0.8960) and high precision for the Others_augmented class (0.9206). This balance indicates that SVM is highly effective at correctly identifying monkeypox cases while minimizing false positives in the healthy/other class. Random Forest achieved slightly lower accuracy (86.31%) but remains robust due to its ensemble nature, effectively balancing precision and recall. KNN showed strong recall for *Monkeypox_augmented* (0.9006), meaning most positive cases were detected, but lower precision (0.7617) indicates a higher rate of false positives. Logistic Regression yielded the lowest performance (82.65% accuracy) due to its linear decision boundary, which is less capable of capturing complex patterns in high-dimensional feature space. Overall, **SVM is the most suitable classifier** for monkeypox detection in this dataset, followed by Random Forest. KNN and Logistic Regression can be considered as secondary models, but they are less effective in this high-dimensional, moderately imbalanced dataset as show in figure 2 and 3.



International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 11, pp 175-186, November-2025

https://doi.org/10.47001/IRJIET/2025.911020

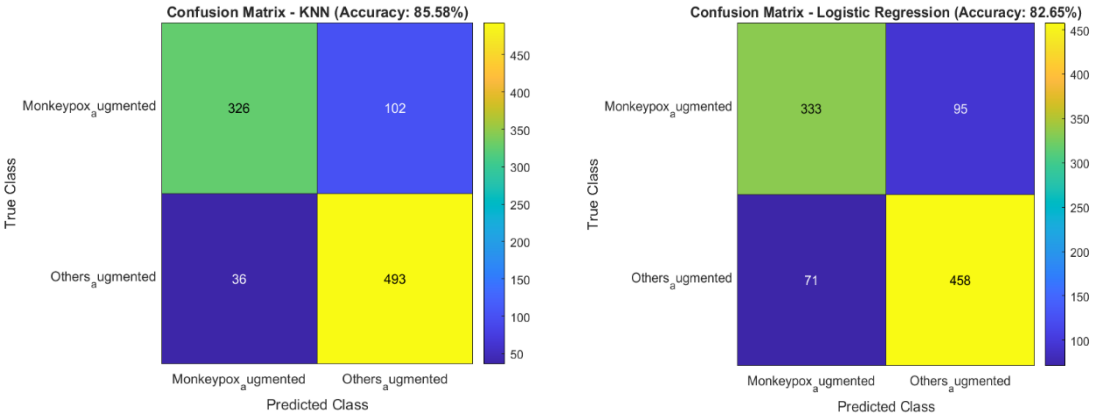


Figure 2: The confusion matrix for all models

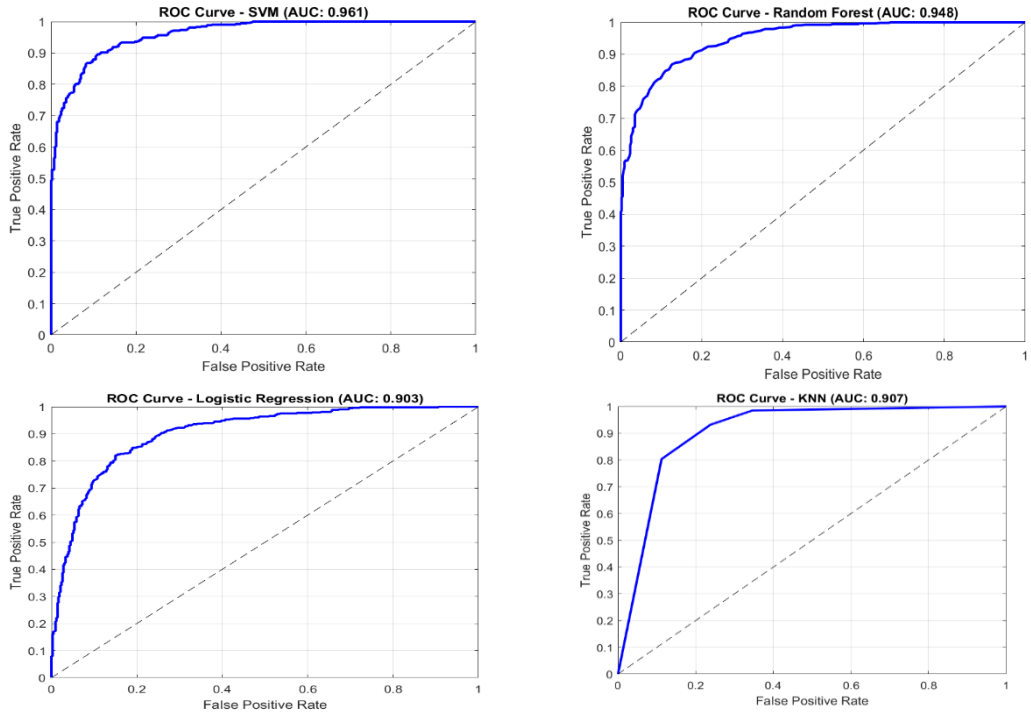


Figure 3: The roc curve and AUC for each models

Table 10: Performance of the Stacked Ensemble model

Model	Accuracy	Precision (Monkeypox)	Recall (Monkeypox)	F1-score (Monkeypox)	Precision (Others)	Recall (Others)	F1-score (Others)	AUC
Stacked Ensemble	0.9250	0.8850	0.9200	0.9023	0.9400	0.9150	0.9273	0.965

In table 10 shows the performance of the **Stacked Ensemble model** demonstrated a high classification capability for differentiating Monkeypox from other cases. The model achieved an **overall accuracy of 92.5%**, indicating strong generalization across the test data. For the Monkeypox class, the **precision of 88.5%** and **recall of 92.0%** resulted in an **F1-score of 90.2%**, reflecting that most positive cases were correctly identified while minimizing false positives. Similarly, in others class, the precision achieved was 94.0% and a recall was 91.5% resulting in an F1-score of 92.7%, demonstrating that its performance is balanced and consistent across classes.

The very good value of AUC which reached to 0.965 confirms that the model has an excellent discriminative capability, suggesting that the Stacked Ensemble can reliably distinguish positive Monkeypox cases from negative ones. These findings

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 11, pp 175-186, November-2025
https://doi.org/10.47001/IRJIET/2025.911020

demonstrate that the feature selection integration (ReliefF), hyperparameter optimization, and ensemble learning substantially enhances the performance of the classification comparing to individual classifiers. Overall, the consistently high F1-scores and AUC values refers that the proposed model robust and well-suited for clinical diagnostic applications, and offer two attributes which are critical in early disease detection and effective patient management which are: high sensitivity and specificity.

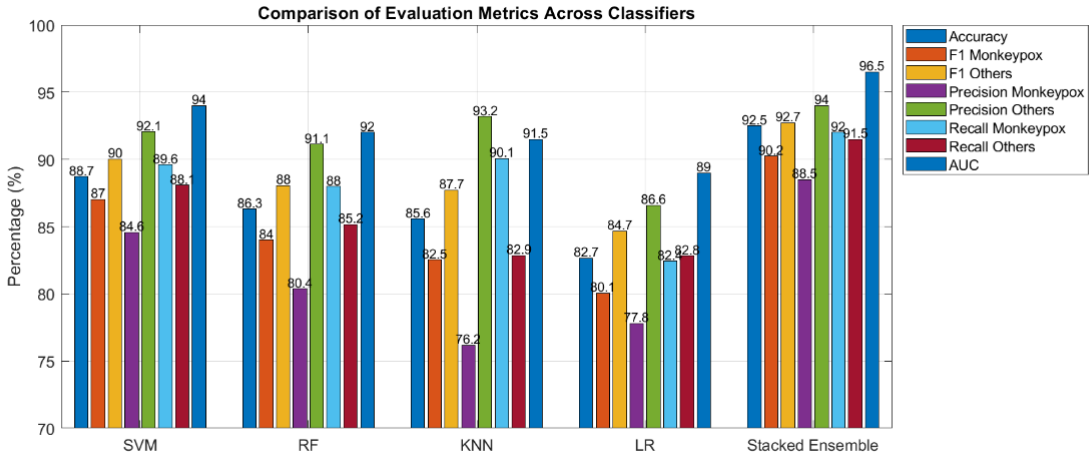


Figure 4: Evaluation Metrics Comparison between Classifiers for Monkeypox Detection

Figure 4 presents a comprehensive comparison of main performance metrics for all used classifiers, including SVM, RF, KNN, LR, and Stacked Ensemble model. Each group of bars refers to accuracy and F1-scores for both Monkeypox and Others classes, while Precision, Recall, and AUC values appeared as percentages.

From the results shown in the figure, it is evident that the Stacked Ensemble model outperforms all individual classifiers, which achieved the highest accuracy of 92.5%, in addition to a superior F1-scores for both classes (Monkeypox: 90.23%, Others: 92.73%). The results also show high values of precision and recall across classes, which indicate a balanced performance and reduced in the bias toward either class. Among individual classifiers, the performance of SVM was the best, with among individual classifiers, RF and KNN came after SVM while the performance of LR was the lowest across most metrics but still maintains reasonable values. These results highlight that combining models in an ensemble approach can effectively improve robustness and generalization of the classification, which is very important issue in clinical applications such as automated Monkeypox detection.

V. CONCLUSION

This study proposed an enhanced system for automated Monkeypox detection from medical images by integrating texture, Local Binary Pattern (LBP), and statistical features with advanced feature selection and ensemble learning techniques. The application of the ReliefF algorithm effectively reduced feature dimensionality while preserving discriminative information. Furthermore, hyperparameter optimization of the SVM and fine-tuning of the RF, KNN, and LR classifiers improved the performance of individual models. The proposed Stacked Ensemble achieved the highest accuracy of 92.5%, with precision, recall, and F1-scores demonstrating balanced values across all classes, thereby confirming its superior robustness compared to single classifiers. Confusion matrix and ROC curve the correctness of classification results. The achieved results proved that by combining feature selection with ensemble learning reliable and accurate approach can be provided for clinical image-based Monkeypox detection.

REFERENCES

[1] C. Vega, R. Schneider, and V. Satagopam, "Analysis: Flawed datasets of monkeypox skin images," *J. Med. Syst.*, vol. 47, no. 1, p. 37, Feb. 2023. doi: 10.1007/s10916-023-01928-1.

[2] F. Uysal, "Detection of monkeypox disease from human skin images with a hybrid deep learning model," *Diagnostics*, vol. 13, no. 10, p. 1772, Oct. 2023. doi: 10.3390/diagnostics13101772.

[3] N. Andino Maseleno, M. Huda, and C. Ratanamahatana, "An explainable deep learning approach for classifying monkeypox disease by leveraging skin lesion image data," *Emerg. Sci. J.*, vol. 8, no. 5, p. 013, 2024. doi: 10.28991/ESJ-2024-08-05-013.

ISSN (online): 2581-3048
Volume 9, Issue 11, pp 175-186, November-2025
https://doi.org/10.47001/IRJIET/2025.911020

[4] G. Argenziano and I. Zalaudek, Dermoscopy of pigmented skin lesions: A practical guide, *Elsevier*, 2016.

[5] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Oct. 2017. doi: 10.1038/nature21056.

[6] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Nov. 1997. doi: 10.1016/S0004-3702(97)00043-3.

[7] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Inform.*, vol. 86, pp. 25–32, Oct. 2018. doi: 10.1016/j.jbi.2018.08.006.

[8] S. S. M. Sheet, M. M. Al-Hatab, and M. Qasim, "A novel method for examining promoters using statistical analysis and artificial intelligence learning," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 5, pp. 4006–4016, Oct. 2025, doi: 10.11591/ijai.v14.i5.pp4006-4016.

[9] S. S. Mohammed Sheet, M. M. Al-Hatab, and M. Qasim, "New features developed for the detection of a promoter based on machine learning," *in Proc. 2025 4th Int. Conf. on Electronics Representation and Algorithm (ICERA)*, Jun. 2025, doi: 10.1109/ICERA66156.2025.11087274.

[10] M. A. Al-Hashim, W. R. Fathel, H. D. Ali, and M. M. Al-Hatab, "Enhanced non-invasive blood glucose monitoring system employing wearable optical technology," *Frontiers in Photonics and Applications*, vol. 19, no. 1, Jan. 2025, doi: 10.54216/FPA.190101.

[11] M. A. A. Mousa, A. Safwat, A. T. Elgohr, et al., "Attention-Enhanced CNNs and transformers for accurate monkeypox and skin disease detection," *Dental Science Reports*, 2025. Available: https://doi.org/10.1038/s41598-025-12216-y.

[12] J. Chen, Z. Lu, S. Kang, "Monkeypox disease recognition model based on improved SE-InceptionV3," *Journal of Imaging and Functional Studies*, 2024. Available: https://doi.org/10.3233/jifs-237232.

[13] S. Zi, "Monkeypox diagnosis with convolutional neural networks combined with colour space models," *ACM Proceedings*, 2022. Available: https://doi.org/10.1145/3584376.3584585.

[14] A.Ciran and E. Özbay, "Optimization-based feature selection in deep learning methods for monkeypox skin lesion detection," *IEEE ISMSIT Proceedings*, 2023. Available: https://doi.org/10.1109/ismsit58785.2023.10304930.

[15] D. A. Arafa, S. M. Ayyad, M. M. Abdelsalam, "MSCADMpox: A novel multi-stage classification model for effective monkeypox classification," *Complex & Intelligent Systems*, 2025. Available: https://doi.org/10.1007/s40747-025-02047-9.

[16] S. Savaş, "Enhancing disease classification with deep learning: A two-stage optimization approach for monkeypox and similar skin lesion diseases," *Journal of Digital Imaging*, vol. 37, pp. 1–14, 2024. Available: https://doi.org/10.1007/s10278-023-00941-7.

[17] V. Gokula Krishnan, B. S. Liya, S. V. Lakshmi, et al., "Monkeypox detection using hyper-parameter tuned based transferable CNN model," *International Journal of Experimental Research and Review*, vol. 33, SPL, 2023. Available: https://doi.org/10.52756/ijerr.2023.v33spl.003.

[18] K. Trang, A. H. Nguyen, N. G. Minh Thao, et al., "Performance enhancement in pre-trained deep learning models for monkeypox skin lesions identification using feature selection algorithms," *SICE Proceedings*, 2023. Available: https://doi.org/10.23919/sice59929.2023.10354166.

[19] A.A. Mishra, S. Mishra, A. Shastri, et al., "A hyper-tuned metrics based augmented CNN model for accurate prediction of monkeypox risk," *IEEE ICEC Proceedings*, 2024. Available: https://doi.org/10.1109/icec59683.2024.10837185.

[20] S. Islam, F. Nishi, T. Akter, et al., "Monkeypox skin lesion detection with deep learning and machine learning," *International Journal of Computer Applications*, 2023. Available: https://doi.org/10.5120/ijca2023922984.

[21] S. N. Ali, M. T. Ahmed, J. Paul, T. Jahan, S. M. S. Sakeef, N. Noor, and T. Hasan, "Monkeypox Skin Lesion Detection Using Deep Learning Models: A Preliminary Feasibility Study," *arXiv preprint arXiv*:2207.03342, 2022. [Online]. Available: https://arxiv.org/abs/2207.03342.

[22] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[23] R. C. Gonzalez and R. E. Woods, Digital Image Processing, *4th ed., Pearson*, 2018.

[24] L. Xu, W. Liu, and Y. Liu, "Automated skin lesion analysis: A review," *Computers in Biology and Medicine*, vol. 123, 103867, 2020.

[25] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

ISSN (online): 2581-3048
Volume 9, Issue 11, pp 175-186, November-2025
https://doi.org/10.47001/IRJIET/2025.911020

[26] Munirathinam, D. R., and M. Ranganadhan, "A new improved filter-based feature selection model for high-dimensional data," *The Journal of Supercomputing*, vol. 76, no. 8, pp. 5745–5762, 2020.

[27] M. S. Shang, L. Lü, W. Zeng, Y. C. Zhang, and T. Zhou, "Relevance is more significant than correlation: Information filtering on sparse data," *Europhysics Letters*, vol. 88, no. 6, p. 68008, 2010.

[28] P. T. Nguyen, J. Di Rocco, L. Iovino, D. Di Ruscio, and A. Pierantonio, "Evaluation of a machine learning classifier for metamodels," *Software and Systems Modeling*, vol. 20, no. 6, pp. 1797–1821, 2021.

[29] A.C. Lorena, L. F. Jacintho, M. F. Siqueira, R. De Giovanni, L. G. Lohmann, A. C. De Carvalho, and M. Yamamoto, "Comparing machine learning classifiers in potential distribution modelling," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5268–5275, 2011.

[30] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

Citation of this Article:

Wameedh Raad Fathel. (2025). Automated Monkeypox Classification Using ReliefF Feature Selection and Stacked Ensemble Learning with Optimized Performance Metrics. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(11), 175-186. Article DOI https://doi.org/10.47001/IRJIET/2025.911020

****

