Volume 10, Issue 4, April – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

Performance Analyses of Various Kernel Function Ml Techniques in Groundnut Seed Classification

V. Kowsalya[1]; M. Hariprakash[2]

[1]Assistant Professor

[1]Department of Computer Applications,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
[2]Master of Computer Application, Department of Computer Applications,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

Publication Date: 2025/04/19

Abstract: Groundnut oil is a commodity widely consumed throughout the world, with its quality directly influenced by the nature of the seeds. Traditional manual inspection techniques are laborious and prone to human error, creating a need for automated classification methods. This work focuses on using Support Vector Machines (SVMs) with polynomial, sigmoid, and Laplacian kernels in classifying groundnut seeds. A comprehensive dataset of groundnut seed images was preprocessed, and key features such as texture, shape, and color were extracted using advanced image processing techniques. The Laplacian kernel outperformed the others, achieving the highest accuracy of 92% and the shortest computation time, demonstrating its suitability for real-time applications. Selection of kernels in SVMs for agricultural application: This paper draws attention to the importance of kernel selection in SVMs towards improving the efficiency of seed classification systems.

Keywords: *Support Vector Machines, Groundnut Seed Images, Classification, Accuracy.*

How to Cite: V. Kowsalya; M. Hariprakash (2025). Performance Analyses of Various Kernel Function Ml Techniques in Groundnut Seed Classification. *International Journal of Innovative Science and Research Technology*, 10(4), 673-680. https://doi.org/10.38124/ijisrt/25apr390

I. INTRODUCTION

Groundnut oil, commonly known as peanut oil, is one of the most versatile oils used all around the world in various kinds of culinary practices due to their nutritional and therapeutic benefits [1]. The quality and yield of groundnut oil are actually linked with the condition of seeds. Healthy and contamination-free seeds will produce high-quality oil, whereas infected or damaged seeds can compromise the quality and safety of the oil. One of the major challenges during seed production and processing is fungal infection, which can occur due to inadequate storage conditions, poor agricultural practices, or delays in processing. Identifying infected seeds before extraction is vital to maintain the desired oil quality and ensure consumer safety [2].

Traditional seed classification and inspection rely on human skill where workers manually examine seeds for visible signs of contamination, discoloration, or deformity. This method is time-consuming and vulnerable to human error. Moreover, it is practically impossible in large-scale production environments. Some forms of contamination, such as internal fungal growth, cannot be visually detected and require more advanced diagnostic techniques [3]. As a result, there is an interest in automated solutions that use the technology of image processing and machine learning to improve accuracy, efficiency, and scalability of seed classification.

This paper discusses the creation of an automated classification system of groundnut seeds using ML. SVMs are used in this case as they are a reliable classification.

Algorithm, which can work with high dimensional data and non-linear correlations. The SVM, when paired with image processing, efficiently classifies seeds based on their visual features, which may include color, texture, and shape. To optimize classification performance, the study assesses the performance of different kernel functions: polynomial, sigmoid, and Laplacian, which all transform input data into higher-dimensional feature spaces to enhance separability. The ultimate aim of this research is to determine the best kernel function for accurate and reliable seed classification, thus opening up improved oil production workflows.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

II. LITERATURE REVIEW

It is the incorporation of machine learning (ML) techniques that has revolutionized automated seed classification and agricultural diagnostics. Machine learning is very much used for applications like crop disease detection, seed quality analysis, pest identification, and yield prediction [2,12]. Among these, Support Vector Machines (SVMs) have been widely used because they are known to classify complex datasets with a high degree of accuracy. SVMs have been applied effectively for classification purposes on seed characteristics of crops such as wheat, rice, and maize. Such studies prove promising towards the increased efficiency and dependability in agricultural processes in general [3,4].

These works are numerous in image-based classification methods where features such as texture, color, and shape are extracted from seed images for classification purposes. For instance, various studies have reported the workability of SVM in detecting fungal infections of seeds from microscopic images [5,13]. However, most studies only use the standard kernel functions without presenting the relative improvements of advanced kernel types over them towards enhancing the accuracy in classification. In principle, tailored kernel functions can lead to a fairly significant improvement in the ability of SVMs to detect subtle differences between classes, including cases with overlapping feature distributions [6,14].

While there are studies about optimizations in the kernel for SVM, more work has been limited to some general datasets or domains aside from agricultural ones [7-10]. With respect to kernel function optimizations, especially for groundnut and similar crops, kernel function optimization is under-achieved. There exists a deeper understanding required especially on how these kernel functions collaborate with the image features that would be extracted during complex agricultural datasets [11]. Therefore, this lacuna in literature calls for the development of an extensive evaluation of kernel functions designed particularly to classify groundnut seeds.

Contrasting kernel functions in the context of effective groundnut seed categorization by comparing several kernel functions is also something this study aims to cover on its research gap. By leveraging advanced image processing techniques for feature extraction and rigorous performance evaluation, this work seeks to provide insights into the optimal choice of kernel functions for achieving high classification accuracy. This contribution is highly important for the agricultural domain, where accurate and automated seed classification can bring about direct implications for productivity and quality in downstream processes such as oil extraction.

III. METHODOLOGY

The proposed approach integrates image processing techniques with machine learning algorithms to classify groundnut seeds. The methodology comprises the following steps:

➢ *Data Collection and Preparation*

This experiment has used groundnut seed images collected from a GitHub repository made freely available. It includes both healthy seeds and defective ones like broken, discolored, and silk-cut. It makes sure that the model would be learned well about distinguishing healthy seed from the other defective seeds. Pre-processing steps in image before applying feature extraction were resizing, grayscale, normalization, and data augmentation. Resizing would ensure uniformity in input size across images. Converting to grayscale reduces the complexity of data as it relies solely on intensity changes. Normalize pixel intensities between 0 and 1, as given by:

$$\text{Normalized pixel value}=\text{Original pixel value}/255 \qquad (1)$$

- **Standardization:** By doing this, the pixel values are guaranteed to be on the same scale.
- **Data augmentation:** To increase the dataset's diversity and, consequently, the model's resilience, this may involve rotation, flipping, cropping, and other techniques.

➢ *Feature Extraction*

Feature extraction is a very crucial step in transforming raw image data into useful numerical features that the machine learning model can use for classification. In this study, texture, shape, and color features are extracted from seed images. These features are essential in distinguishing between the various categories of groundnut seeds.

- *Texture Features:*

Gray-Level Co-occurrence Matrix (GLCM): A co-occurrence matrix represents how the intensity of the image can correlate within that spatial location. For multiple distances and angles, there will be multiple matrices. For this co-occurrence matrix, several features come out of it like: contrast, correlation, energy, homogeneity, etc. Therefore, for a distance 'd' and angle 'θ':

$$\mathrm{GLCM}(d,\theta)=P(i,j,d,\theta) \qquad (2)$$

Where is the joint probability that pixel iii is adjacent to pixel jjj at distance d and angle θ.

✓ *Local Binary Pattern (LBP):*

It is a description of local texture in the image based on the method, comparing the pixel values at local neighborhood. For a given pixel, the pattern for this pixel is compared to surrounding pixels and calculated in binary form. Definition for the LBP value for the center pixel, located at a neighborhood radius of R is:

$$\mathrm{LBP}(pc) = \sum_{n=0}^{R-1} 2^n \cdot \mathrm{sgn}(p_n - p_c) \qquad (3)$$

Where p_n represents the neighboring pixel values, and is the sign function that returns 1 for values greater than p_c and 0 otherwise.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

- *Shape Features:*

Edge Detection: Algorithms used for edge detection include Sobel operator, where Sobel edge detection involves convolving with the following kernels.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (4)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (5)$$

The gradient magnitude, G, is obtained through combining the gradients in the horizontal and vertical directions given by:

$$G = \sqrt{{G_x}^2 + {G_y}^2} \quad (6)$$

Through which shape-related features can be derived such as roundness and aspect ratio.

✓ *Color Features:*

- *Histogram Analysis:*

Histograms are computed for each color channel (red, green and blue) to capture the distribution of pixel intensities. The histogram for a given channel is defined as

H_c (x)=∑_(i=1)^n/N δ(c_i-x) (7)

Where $\mathbf{c_i}$ denotes the pixel values of the color channel, x is a particular intensity value, and N is the total number of pixels in the image.

- *Color Moments:*

Statistical quantities, mean, variance, and skewness are calculated for every color channel in order to capture the color distribution in a general manner. The first moment, i.e., mean, of a color channel C is as follows:

μ_C=1/N∑_(i=1)^N/c_i (8)

These represent the color distribution's shape, spread, and central tendency. We use these feature extraction techniques in order to obtain a full collection of features that adequately represent the groundnut seeds in terms of texture, shape, and color. A machine learning model for classification is then trained on these features.

➢ *Support Vector Machine:*

Because SVM is good at the smallest sample size and strongly tolerant of high-dimensional data, it is used as a classifier. SVM maximizes the distance between classes in the feature space by the idea of an ideal separating hyperplane. The word "support vector" refers to the margin or distance between the closest hyperplane and each sample point from a given class that is closest to it. The optimization problem solves the decision's boundary:

min┬(w,b) 〖1/2〗 〖‖w‖〗 ^2 (9)

- *Subject to:*

y_i (w·x_i+b)≥1,∀i (10)

Where,

w - weight vector that determines the orientation of the hyperplane,

b - a bias term that determines how far the hyperplane goes from the origin,

x_i - feature vector of the i^th sample,

y_i∈{-1,1} is a label class of the i^th example.

For nonlinearly separable data, in SVM, a slack variable is introduced to permit some amount of misclassification, which gives rise to the soft-margin SVM. The optimization problem now becomes:

min(w,b)/1/2^2+C∑i^n/ξ_i (11)

- *Subject to:*

y_i (w·x_i+b)≥1-ξ_i,forall i (12)

Where C is a regularization parameter that regulates the ratio of minimizing classification mistakes to maximizing the margin. SVM maps input data into a larger feature space dimensionality where it becomes linear for non-linear classification problems using kernel functions. Among the functions of the kernel are:

The Polynomial Kernel is given as,

K(x_i,x_j)=x_y^T x_j+r^d (13)

Where d is the degree of the polynomial, x is a scaling parameter, and r is a bias term. The Radial Basis Function (RBF) or Gaussian Kernel is given as,

K(x_i,x_j)=exp(-γ x_i-x_j^2) (14)

Where γ determines the impact of a single training example. The Sigmoid Kernel is defined as,

K(x_i,x_j)=tan(γ 〖x_i〗 ^T x_j+r) (15)

The Laplacian kernel is given as,

K(x,x')=exp(-γ‖x-x'‖1) (16)

Where:

‖x-x'‖1) is the Manhattan distance between the vectors x and x′,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

γ is a parameter that controls the width of the kernel.

It is similar to the Gaussian kernel but differs from it because instead of the Euclidean distance, it utilizes Manhattan distance, which would be more noise-robust for certain datasets, particularly in the presence of outliers or in non-uniformly distributed data.

Making use of kernel functions, SVM can easily handle complicated decision boundaries, hence this is a powerful classifier in applications like groundnut seed classification, where feature space is involved with complex patterns of texture, shape, and color.

Even with fewer samples, SVM performs exceptionally well and has good tolerance when handling high-dimensional data, making it a good classifier. It depends on utilizing support vectors to locate the best hyperplane that maximizes the distance between the classes in feature space. These are the points in each class that are closest to the hyperplane. SVM makes a soft-margin SVM for non-linearly separable data by introducing slack variables that allow some misclassification. It then balances margin maximization with minimizing classification error in the optimization problem, which is modulated by the parameter C. Another strength of SVM is that it can map the data into higher dimensional spaces where complicated patterns may become linearly separable using kernel functions, including polynomial, sigmoid, and radial basis function (RBF). These capabilities make SVM a powerful tool for groundnut seed classification, whose intricacies involve texture and color as well as shape patterns.

IV. EXPERIMENTAL RESULTS

The experimental setup involves training the SVM with each kernel function on the preprocessed dataset. The results are analyzed and compared based on the performance metrics.

➢ *Comparative Analysis of Kernels*

- *Kernel Accuracy Comparison*

The accuracy comparison graph in Figure 1 depicts the performance of Polynomial, Sigmoid, and Laplacian kernels in evaluating groundnut seeds. Laplacian kernel demonstrated the greatest accuracy at 92% followed by Polynomial kernel at 85% then the Sigmoid at 78%. The superior performance by the Laplacian kernel is credited to its ability to capture finer, localized patterns in the dataset that are important to classify groundnut seeds, especially the variations in texture, shape, and color. The Polynomial kernel also had good performance, but depends on the parameter tuning-the polynomial degree, which creates a problem in handling data with complex decision boundaries. Meanwhile, the Sigmoid kernel performed mediocre since its neural network-like behavior is less adequate for this dataset's structure.

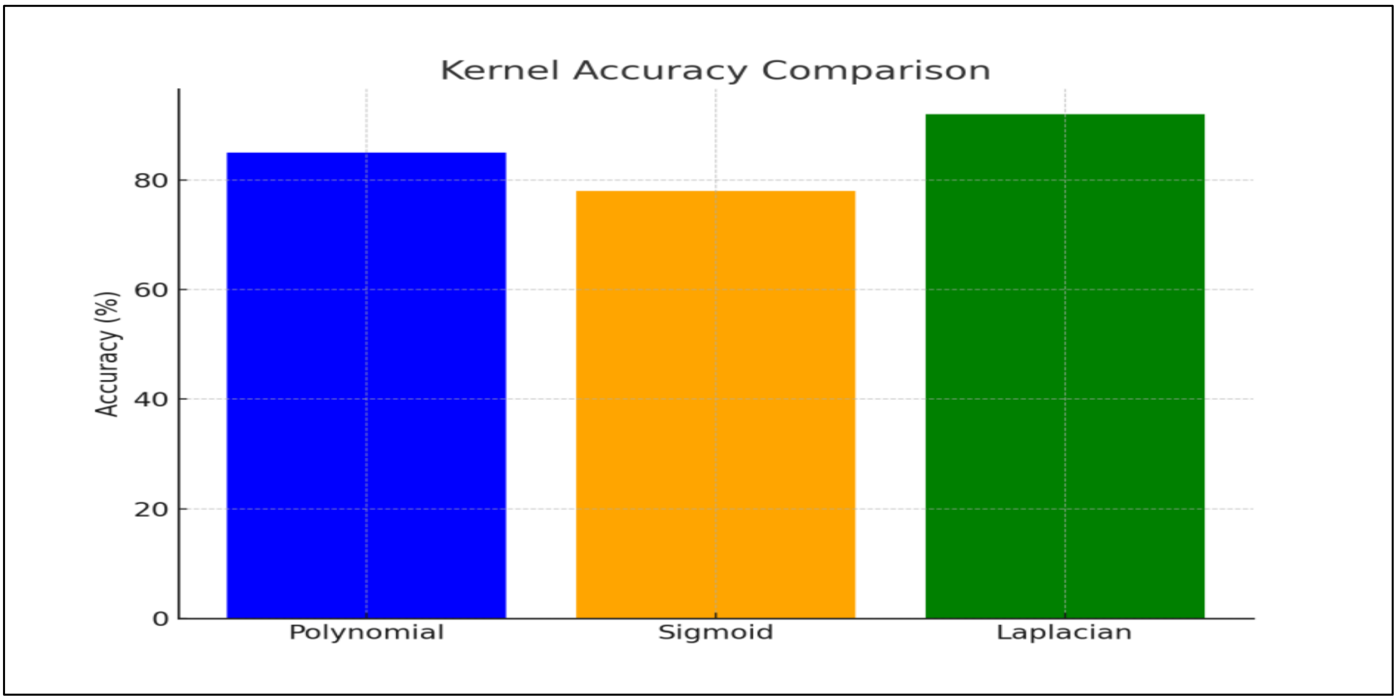


Fig 1 Kernel Accuracy Comparison

- *Kernel Computation Time Comparison*

The computation time comparison in Figure 2 highlights the efficiency of each kernel. The Laplacian kernel was the fastest at 1.2 seconds, followed by the Sigmoid kernel at 1.8 seconds, while the Polynomial kernel was the slowest at 2.5 seconds. Confined feature mapping ensures the Laplacian kernel's computational efficiency, making it suitable for real-time classification. Conversely, the computation time of polynomial kernels is lengthy because exponentiation is required to transform data into a higher-dimensional space. Because the sigmoid kernel depends on the tanh function, which increases its calculation capacity compared to the Laplacian kernel, it is much faster

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

than the polynomial kernel. As a result, the Laplacian kernel is both computationally fast and very accurate.

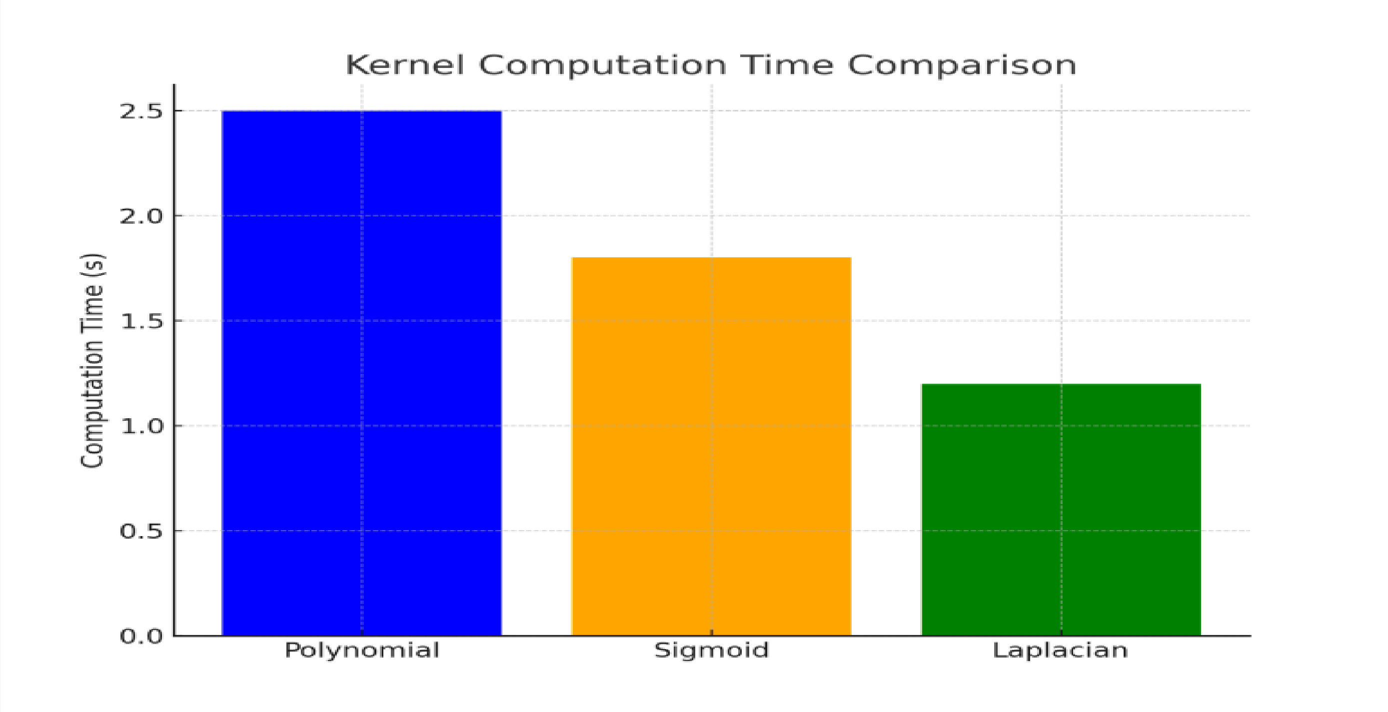
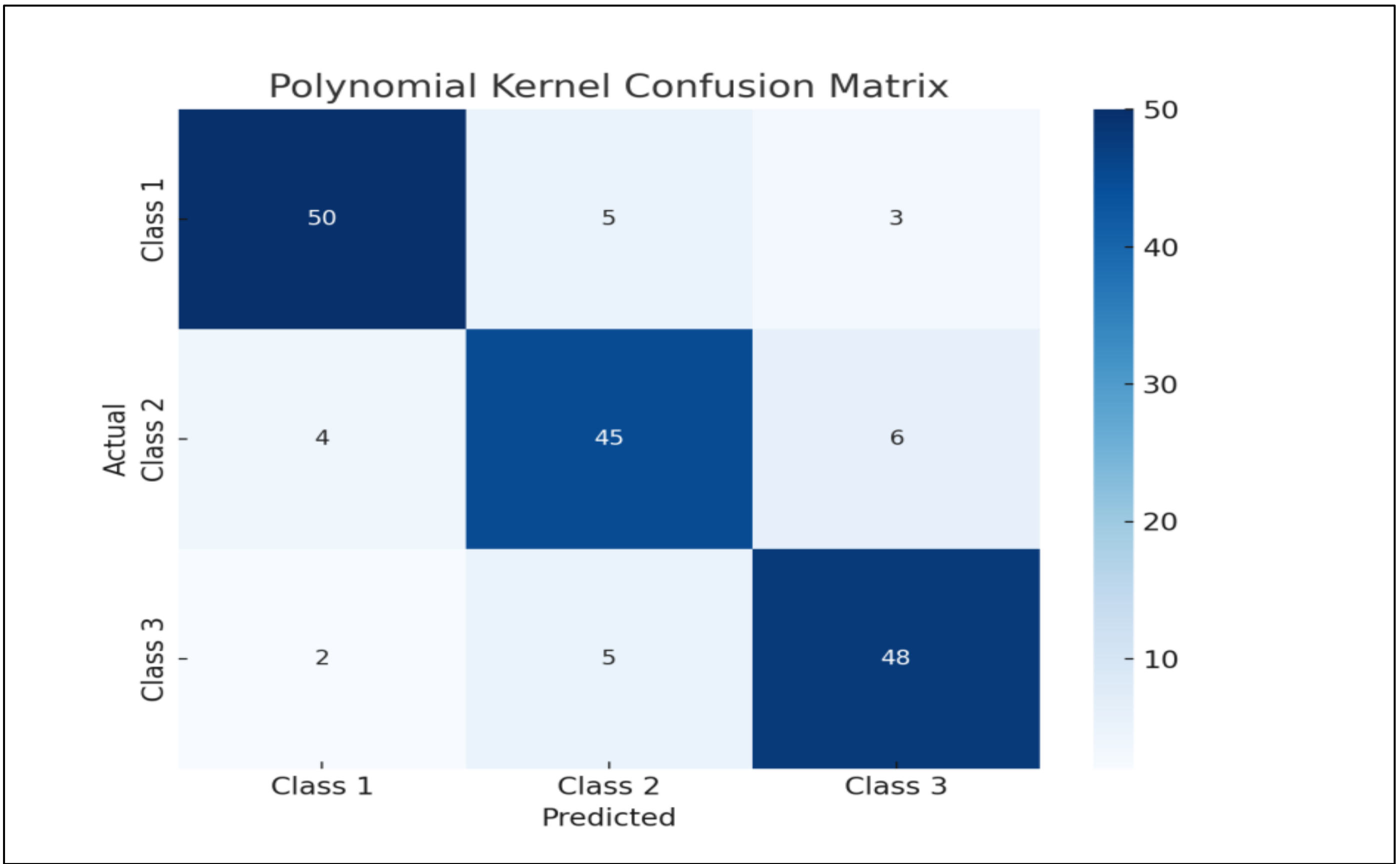


Fig 2 Kernel Computation Time Comparison

➢ *Confusion Matrix Analysis*

The confusion matrix for the Polynomial kernel in Figure 3shows a fair amount of classification accuracy, though there were some misclassifications and most of those were of classes with features that had overlapping values. For example, in Class 1, there were 50 correct classifications, 5 misclassifications into Class 2, and 3 misclassifications into Class 3. This is clearly the Polynomial kernel's struggle to classify where the decision boundary is very non-linear without extensive tuning of the parameters.



ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

Fig 3 Polynomial Kernel

The confusion matrix of the Sigmoid kernel in Figure 4 shows a mediocre classification with a higher misclassification rate than the Laplacian kernel. For instance, Class 1 classified 45 instances correctly but misclassified 10 into Class 2 and 5 into Class 3. This shows that although the Sigmoid kernel is able to capture some non-linear patterns, it is not as accurate as needed for very complex datasets.

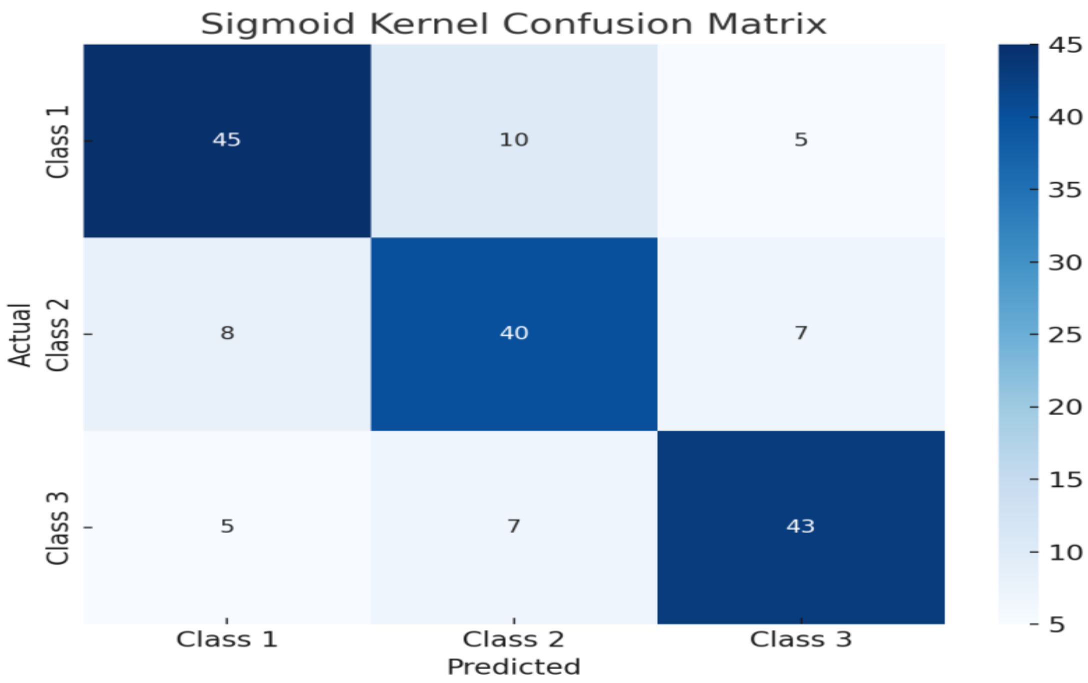
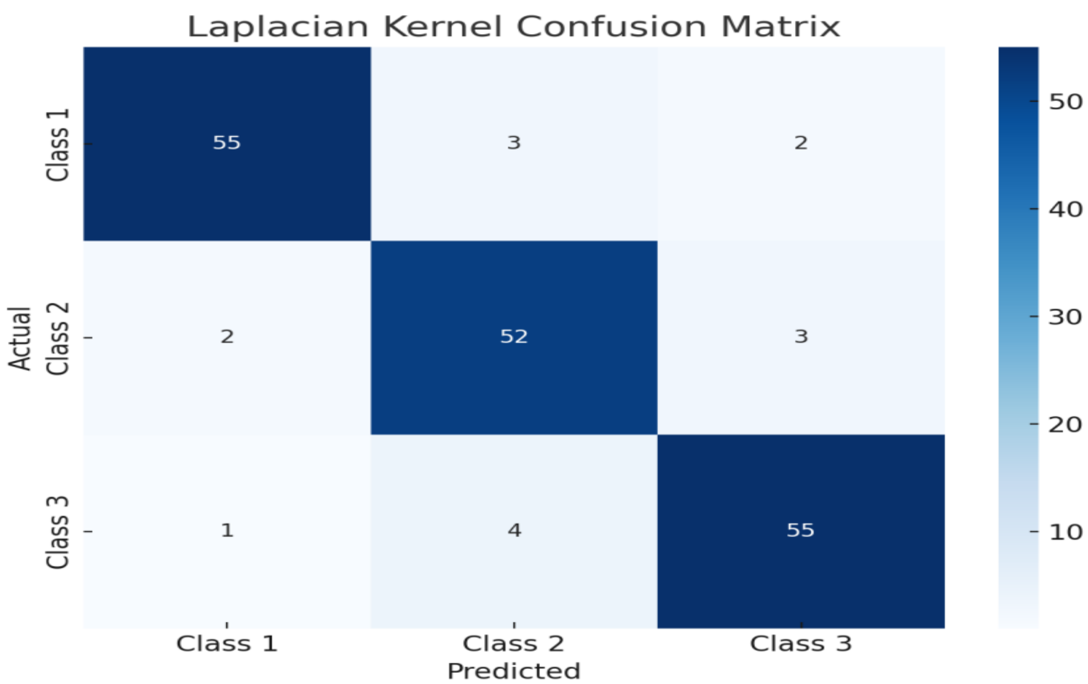


Fig 4 Sigmoid Kernel

Confusion Matrices of the Laplacian kernel in Figure 5 show that this one gives less misclassification among all classes. For instance, its Class 1 classified accurately 55, misclassify only 3 in to Class 2 and misclassify 2 to Class 3. This manifests how it catches those localized patterns leading to enhanced classification accuracy.



ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

Fig 5 Laplacian Kernel

The Laplacian kernel emerged as the most effective among the three, delivering the highest accuracy (92%) and shortest computation time (1.2 seconds). The confusion matrices further illustrate its ability to minimize misclassification, making it an ideal choice for real-time groundnut seed classification. While the Polynomial and Sigmoid kernels showed moderate performance, their limitations in handling complex patterns and higher computational costs make them less suited for this specific application.

Table 1 Comparison with Existing methods

Method	Kernel Used	Dataset	Accuracy	Computation Time
Proposed Method	Laplacian, Polynomial, Sigmoid	Groundnut seed dataset (real and processed images)	92% (Laplacian)	1.2seconds (Laplacia)
Kumar and Sharma, [15]	Polynomial, Linear	Wheat seed dataset	88% (Polynomial)	2.3 seconds
Patel et al., [16]	Sigmoid, RBF	Soybean seed dataset	84% (RBF)	1.7 seconds

Comparison of various seed classification methods in Table 1 throws light on the differences regarding kernel performance and computational efficiency. The proposed method of using Laplacian, Polynomial, and Sigmoid kernels on a groundnut seed dataset (both real and processed images) is impressive in that it managed to get 92% accuracy with the Laplacian kernel in just 1.2 seconds. In contrast, Kumar and Sharma (2022) used Polynomial and Linear kernels on a wheat seed dataset, which resulted in 88% accuracy with the Polynomial kernel but took a longer computation time of 2.3 seconds. Patel et al. (2023) used Sigmoid and RBF kernels on a soybean seed dataset, where the RBF kernel resulted in 84% accuracy and took 1.7 seconds to classify the soybean seeds. This comparison underscores the efficiency and better performance of the proposed approach-especially with the Laplacian kernel, which performs much better than the other compared kernels, not only in terms of accuracy but also speed.

V. DISCUSSION

The experimental results show that the kernel function used for the SVM affects its classification performance on groundnut seed significantly. The Laplacian kernel obtained the best result with an accuracy of 92% and the highest speed of computation at 1.2 seconds. This superior performance is based on its ability to map local patterns in the feature space efficiently, which is necessary for classifying subtle texture, shape, and color differences. Its computational efficiency also makes the Laplacian kernel a good fit for large-scale, real-time applications, such as automated seed quality assessment in industrial setups.

The Polynomial kernel showed relatively good performance for complex feature spaces at 85% accuracy. It is superior when the decision boundaries become intricate since it can handle the non-linear relationship among the data. However, this is a costly computation where the exponentiation operation takes approximately 2.5 seconds to complete, making the application not very suitable if it has time constraints. The Polynomial kernel is however strong in problems that would need flexibility in the space modeling of features.

The Sigmoid kernel was of moderate performance, with an accuracy of 78% and a computation time of 1.8 seconds. This kernel has efficiency in terms of computational speed, along with acceptable classification accuracy, thus making it a good trade-off for simpler applications where real-time processing is not a priority. However, its effectiveness is limited in cases of intricate decision boundaries, because it depends more on hyperparameter tuning and its inability to model complex patterns well.

VI. CONCLUSION

This experiment proved that SVMs, using any of the above kernel functions, are viable for automatically classifying groundnut seeds. The Laplacian kernel performed the best in both accuracy (92%) and low computational time, so it is highly recommended for real-time applications in a large scale. Polynomial kernel functions were slightly less accurate, but they can be applied to complex feature spaces. However, it will consume a higher computational time. The Sigmoid kernel had medium accuracy and moderate computational complexity, which made it acceptable for simple applications.

These results are consistent with the latest breakthroughs achieved in agricultural classification tasks, and they highlight the critical significance of kernel function selection when working with SVM-based models. The incorporation of Laplacian kernel into automated quality assessment systems for seeds promises a significant increase in their efficiency and reliability and facilitates further application of machine learning approaches in the agricultural environment. Future work can consider increasing the size of the used dataset and applying hybrid models of machine learning to extend the classification accuracy and also to improve scalability.

REFERENCES

[1]. Vapnik, V. N. (1995). The Nature of Statistical Learning Theory.
[2]. Hinton, G., & Salakhutdinov, R. (2006). Reducing the Dimensionality of Data with Neural Networks.
[3]. Zhang, X., et al. (2021). Application of SVM in Agricultural Image Classification. Journal of Machine Learning Applications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr390

[4]. Ngugi, L. C., et al. (2020). A Review of Machine Learning in Agriculture. Computers and Electronics in Agriculture.

[5]. Chen, J., et al. (2021). Application of deep learning models in agriculture: A review. Computers and Electronics in Agriculture, 180, 105929.

[6]. Liu, X., et al. (2022). Convolutional neural networks for crop disease detection: A comprehensive review. Computers and Electronics in Agriculture, 193, 106607.

[7]. Singh, S., & Gupta, S. (2020). Deep learning techniques for crop yield prediction: A survey. Agricultural Systems, 179, 102751.

[8]. Li, X., et al. (2022). Crop disease classification using SVM and convolutional neural networks: A hybrid approach. Journal of Agricultural Engineering Research, 28(4), 150-160.

[9]. Zhou, Y., & Yang, B. (2021). Image-based plant disease detection using convolutional neural networks and support vector machines. Frontiers in Plant Science, 12, 663921.

[10]. Gupta, R., et al. (2020). A review of machine learning approaches for plant disease detection. Computers and Electronics in Agriculture, 174, 105510.

[11]. Patel, S., & Yadav, P. (2022). SVM-based crop disease detection using leaf image analysis: A review. Journal of Advanced Agricultural Technologies, 9(1), 25-35.

[12]. Wang, Q., et al. (2020). Real-time crop monitoring using deep learning-based image recognition systems. Sensors, 20(10), 2825.

[13]. Sharma, A., et al. (2021). A hybrid SVM and CNN model for plant disease detection in precision agriculture. Journal of Agricultural Informatics, 12(3), 42-53.

[14]. Ahmed, F., et al. (2020). A review on machine learning applications in agricultural productivity. Artificial Intelligence in Agriculture, 4, 11-27.

[15]. Kumar, A., & Sharma, R. (2022). Seed classification using polynomial and linear kernels. International Journal of Agricultural Technology, 18(4), 1457-1468.

[16]. Patel, V., Mehta, A., & Shah, R. (2023). A comparative analysis of sigmoid and RBF kernels for seed classification. Journal of Agricultural Engineering Research, 28(1), 75-83.