



Early Detection of Grasserie Disease in Silkworms Using Computer Vision and Machine Learning

Sania Thomas¹, Binson V A^{2*}, Sini Rahuman³, Sivakumar K S⁴

¹Department of Computer Science and Engineering, Saintgits College of Engineering, Kerala, India

²Department of Electronics Engineering, Saintgits College of Engineering, Kerala, India

³Faculty of Foundation (Mathematics), Bahrain Polytechnic, Isa Town, Kingdom of Bahrain

⁴Department of Instrumentation & Control Engineering, NSS College of Engineering, Kerala, India

¹sania.thomas@saintgits.org, ²binson.va@saintgits.org, ³sini.rahuman@polytechnic.bh, ⁴kssivakumar86@gmail.com

Abstract

One of the major challenges associated with the sericulture industry is silkworm diseases, as they are very difficult to detect in the early stages. Timely identification of infected silkworms is essential to curb the spread of disease and reduce economic damage. This study focuses on diagnosing Grasserie disease, a highly contagious condition that can devastate silkworm populations, leading to substantial financial losses for farmers. To address the shortcomings of expert manual inspections, this study employed camera-captured images of silkworms for automated disease detection. A newly compiled dataset, consisting of 668 healthy silkworms and 574 infected with Grasserie disease are used for this study. The dataset is analyzed with machine learning techniques for image analysis, features are extracted from the pre-processed images using combining Histogram of Oriented Gradients (HOG) and the higher dimensional features are reduced with Kernel Principal Component Analysis (KPCA), and classification using supervised models. The results highlight the effectiveness of this approach in differentiating healthy silkworms from diseased ones. The machine learning model HOG integrated with KPCA and Decision Trees (DT) achieved strong performance, with accuracy, recall, and precision scores of 94.28%, 94.56%, and 92.48%, respectively. While these outcomes are encouraging, further research is needed to develop a practical IoT-based tool that enables sericulture farmers to quickly detect infections and take preventive measures, minimizing unexpected losses. This study marks a crucial advancement in silkworm disease detection, offering a pathway toward greater sustainability and economic stability in the sericulture sector.

Keywords: grasserie; histogram oriented gradient; machine learning; sericulture; silkworm

How to Cite: Sania Thomas, Binson V A, Sini Rahuman, and Sivakumar K S, "Early Detection of Grasserie Disease in Silkworms Using Computer Vision and Machine Learning", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 5, pp. 1255 - 1264, Oct. 2025.

Permalink/DOL: <https://doi.org/10.29207/resti.v9i5.6705>

Received: June 2, 2025

Accepted: September 20, 2025

Available Online: October 27, 2025

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

Sericulture, which refers to the cultivation of silkworms for silk production, has a rich history extending back thousands of years. Among the different species of insects that produce silk, the silkworm, scientifically known as *Bombyx mori*, has been the primary subject of sericulture due to its remarkable capacity for generating high-quality silk [1]. Sericulture, which involves cultivating silk and rearing silkworms, has played an important role in the economic and social development of many regions worldwide. South India, in particular, is recognized for its deep-rooted tradition in sericulture, benefiting from a favorable climate and environment that support the flourishing of the *Bombyx mori* species. The life cycle of silkworms is a complex

and captivating process, consisting of several key stages, each essential for silk production [2], [3]. A thorough understanding of these stages is critical for the successful practice of sericulture. The life cycle of silkworms generally includes four stages: eggs, larvae (caterpillars), pupae enclosed within cocoons, and adult moths [4], [5]. Despite the cultural and economic significance of sericulture in South India, the industry has faced significant challenges, particularly the spread of grasserie disease in silkworms. Recently, the rise of machine learning technologies has created advancements that offer earlier recognition and better handling of grasserie disease, bringing transformative changes to the sericulture sector [6], [7]. This research paper delves into the application of machine learning techniques for the effective detection of grasserie

disease in *Bombyx mori* silkworms, shedding light on its potential to safeguard the silk production process and contribute to the sustainable growth of South Indian sericulture.

A major problem with the sericulture industry is the devastating impact of grasserie disease, which severely affects silkworm populations [8]. This disease is triggered by the *Bombyx mori* nuclear polyhedrosis virus (BmNPV) and is highly contagious, posing a serious risk to silkworm farming [8]-[10]. Infected larvae suffer from symptoms, including lethargy, reduced appetite, sluggish movement, and difficulty spinning proper cocoons [11]-[14]. These symptoms typically emerge 5 to 7 days after infection. During the early stages, the silkworm's skin develops a shiny, delicate texture, and the larvae struggle with the moulting process. The virus spreads quickly through rearing facilities, resulting in significant financial losses. Infected silkworms also produce weak and poor-quality cocoons due to abnormal spinning patterns. Early detection is therefore crucial for disease control and to minimize economic harm. Emerging machine learning techniques offer promising solutions to address this threat and protect silk production.

Currently, the detection of grasserie disease in silkworms relies heavily on traditional methods, such as manual inspections and visual assessments by skilled sericulturists [15], [16]. These are time-consuming, labor-intensive, and subjective, often leading to inconsistent diagnoses and delayed detection. Sericulturists observe behavioral and physical changes such as decreased appetite, unusual spinning, and lethargy during silkworm development [17]. Microscopic analysis of gut tissues and excreta can help confirm the presence of the virus [18]. Despite their widespread use, these conventional diagnostic methods are not always reliable, particularly when dealing with early infections or subtle disease symptoms. Hence, researchers are increasingly focusing on advanced tools, including molecular techniques and machine artificial intelligence (AI) methodologies, to enable faster and more accurate detection, ensuring timely interventions and better disease management.

In their research, Nagashetti et al. attained an impressive accuracy of 97 % in differentiating between 670 healthy silkworms and 527 silkworms infected with Grasserie disease, using deep learning techniques [5]. Similarly, Zhen et al. used 600 images of silkworms affected by nuclear polyhedrosis virus and white Muscardine disease to assess the performance of the Attention-Concatenation Dense Convolutional Neural Network (AC-DenseNet). They achieved a notable peak accuracy of 85.6%, highlighting the effectiveness of AC-DenseNet in differentiating between these two diseases in silkworms [6]. Shilpashree et al. employed Gray-level Co-occurrence Matrix (GLCM) texture features to classify healthy and diseased silkworms. Their use of the Random Forest algorithm yielded a recall of 59%, precision of 61%, an F1-score of 85%,

and an overall accuracy of 59% [7]. Additionally, Xia et al. developed a silkworm disease recognition application using the Local Binary Pattern (LBP) feature extraction method combined with an SVM-based classifier, achieving an accuracy of 73.8% [19]. The system operates on a client/server network structure, enabling users to capture real-time images of diseased silkworms via a mobile application and upload them directly to a cloud for analysis.

In this study, we propose an artificial intelligence-based method for the detection of BmNPV infection. HOG feature extraction technique is combined with various dimensionality reduction techniques such KPCA, Multidimensional Scaling (MDS), and Isomap. To classify the extracted features, we employ KNN and decision tree algorithm, and we assess its performance using standard evaluation metrics. The robustness of the model is validated using a 10-fold cross-validation approach. Our proposed HOG-KPCA-DT method demonstrates superior classification accuracy in comparison with other dimensionality reduction methods, achieving high sensitivity and specificity in detecting Grasserie disease. This work significantly contributes to the advancement of automated disease detection in sericulture by providing a reliable and efficient method for identifying Grasserie disease in silkworms.

2. Methods

The process of detecting Bm NPV infection in *Bombyx mori* silkworms using image analysis starts with sample collection. A mobile camera is used to take the picture of the silkworm samples and the captured pictures are then saved to a personal computer for analysis. The next step involves preprocessing the images to remove any noise or irrelevant components. After preprocessing, key features are collected from silkworm image data. To optimize the data for analysis, dimensionality reduction methods are employed. The optimized feature set is then used as input for the machine learning algorithms to detect the presence of the grasserie disease. The complete methodology is visually depicted in Figure 1.

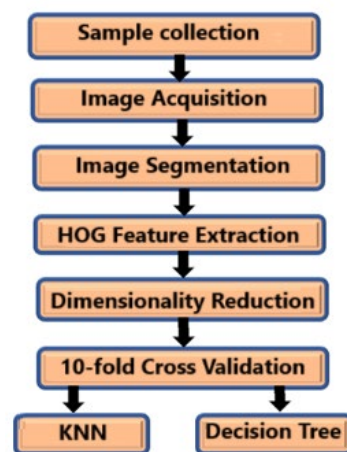


Figure 1. Workflow of the process

2.1. Sample Collection and Image Acquisition

This research utilized a double hybrid variety of silkworm called the FC1 x FC2 double hybrid. In southern India, CSR (Central Sericultural Research) hybrids FC1 and FC2 are frequently used to produce these double hybrids, which are valued for their robustness, higher yield, and adaptability to various climatic conditions [20]-[22]. FC1 and FC2 are hybrid breeds crossing CSR6 with CSR26 and CSR2 with CSR27 respectively. The study collected samples from farmers in Namakkal district, located in the Tamil Nadu state of India. The silkworm samples are collected from the farmers authorized by the sericulture department. A total of 1,242 samples were gathered for the research, comprising 668 healthy silkworms and 574 silkworms infected with Grasserie disease. Figure 2 illustrates an image of the silkworms from the farm. The initial classification of healthy and diseased silkworms was done in collaboration with experienced farmers and confirmed it using microscopic examination.

Starting on the fifth day after infection, an expert carefully examined each infected silkworm individually, selecting the diseased samples. The image capture process was completed in a single day, using a Samsung Galaxy S20 FE smartphone with a 12-megapixel camera. A consistent lighting setup with a LED Macro arm light, were used to provide uniform illumination and minimize shadows or glare. The phone was positioned 30 cm above the silkworms, mounted on a tripod to ensure that the camera lens was directed downward while allowing the silkworms to remain in their natural position. All images were taken from the same angle and under identical ambient conditions to avoid variation in lighting and perspective.



Figure 2. Image of silkworms

To preserve the silkworms' natural body shape when resizing the images, the aspect ratio of the camera screen was set to 1:1. The silkworms were placed against a black background. The black background was used in this study to ensure consistent image acquisition, enhance contrast, and minimize noise during preprocessing and feature extraction, which is particularly important when working with a limited dataset. Figures 3a and 3b display images of a silkworm

affected with Grasserie disease and a healthy silkworm, respectively. Following image acquisition, a cropping process was applied to resize the images to 640×480 pixels, focusing on the region of interest. Subsequently, the cocoon was isolated from the background through binary segmentation, effectively removing noise. The resulting segmented image served as the basis for subsequent feature extraction.



Figure 3. (a) Diseased silkworm (b) Healthy silkworm.

2.2. Feature Extraction

Feature extraction involves obtaining important information from silkworm images to generate a set of unique attributes that capture the visual characteristics of the silkworms. These features can include aspects such as texture, color, shape, and other visual traits. The extraction of these features is crucial to our methodology, as it enables the transformation of complex image data into a format that AI models can efficiently analyze [23], [24]. The effectiveness of our disease detection model largely relies on the selection and quality of these features, which allow for accurate differentiation between healthy and diseased silkworms, thereby contributing to improvements in silkworm health management within sericulture. HOG feature descriptor method is employed in this work as our feature extraction technique.

HOG is a key method used in feature extraction, particularly for solving object detection problems in computer vision. First introduced by Robert K. McConnell [25], this technique gained wide recognition when researchers Navneet Dalal and Bill Triggs demonstrated its effectiveness in achieving high accuracy for human detection [26]. HOG's strength in object detection lies in its ability to describe objects by leveraging the gradient magnitude and orientation. In silkworm image-based studies HOG feature extraction technique has given better accuracy results. He et al. effectively used HOG technique integrated with MLP, SVM and random forest for species and sex classification of silkworms [3]. They have attained an impressive accuracy of 99 % in both the cases of species and sex classification. In disease detection applications

HOG has given good results. S Bakheet achieved 98% sensitivity, 96% specificity, and 97% accuracy in skin cancer detection using HOG for feature extraction and SVM classifier [27]. Admass et al. utilized HOG and Convolutional Neural Networks (CNN) for detecting the mango leaf diseases and achieving an accuracy above 98% [28]. HOG method is also widely adapted in disease specific application in both human, plant and animal diseases.

The HOG feature extraction process consists of several essential steps. To start, the original image, shown in Figures 4.a and 5.a with dimensions of 640×480 , is converted to grayscale and resized to 128×64 , as illustrated in Figures 4.b and 5.b, to enhance the accuracy. Following this, the gradients of the resized image are computed using Formulas 1 and 2.

$$G_x(r, c) = I(r, c + 1) - I(r, c - 1) \quad (1)$$

$$G_y(r, c) = I(r - 1, c) - I(r + 1, c) \quad (2)$$

Here in x-direction and y-direction gradients are denoted as G_x and G_y respectively. c and r refers to the columns and rows and, I stands for the intensity value of a pixel.

Once G_x and G_y are calculated, the magnitude and orientation are obtained using Formulas 3 and 4. These are visualized in Figures 4 and 5.

$$\text{Magnitude}(\mu) = \sqrt{G_x^2 + G_y^2} \quad (3)$$

$$\text{Angle}(\theta) = \left| \tan^{-1} \left(\frac{G_y}{G_x} \right) \right| \quad (4)$$

The following step consists of splitting the magnitude and angle arrays into 8×8 blocks. Next step is to divide the gradient angle into nine bins, each with a step size ($\Delta\theta$) of 20 degrees. Bin boundaries $[\Delta\theta.j, \Delta\theta.(j+1)]$ are defined for each bin, with the center of each bin, represented as c_j , calculated using equation 5. For each angle value (θ), the magnitude values are assigned to the corresponding bins based on Formulas 6, 7, and 8.

$$c_j = \Delta\theta \cdot (j + 0.5) \quad (5)$$

$$j = \left\lfloor \frac{\theta}{\Delta\theta} - \frac{1}{2} \right\rfloor \quad (6)$$

$$V_{j+1} = \mu \cdot \left[\frac{\theta}{\Delta\theta} - \frac{1}{2} \right] \quad (7)$$

$$V_{j+1} = \mu \cdot \left[\frac{\theta - c_j}{\Delta\theta} \right] \quad (8)$$

The HOG feature descriptor histogram is calculated for 8×8 sections of the image. Each cell within these sections holds 9×1 values. Next, four 9×1 cells are combined to form new 2×2 blocks, resulting in 36 features per block. The values within these feature blocks are then normalized using the L2 norm, as defined by Formula 9, with a small constant (ϵ) added to prevent division by zero error.

$$f_b = \frac{f_b}{\sqrt{\|f_b\|^2 + \epsilon}} \quad (9)$$

In the end, the final feature vector for an image consists of $7 \times 15 \times 36$, resulting in a total of 3780 features. The conclusive depiction of the HOG feature descriptor is given in Figures 4e and 5e.

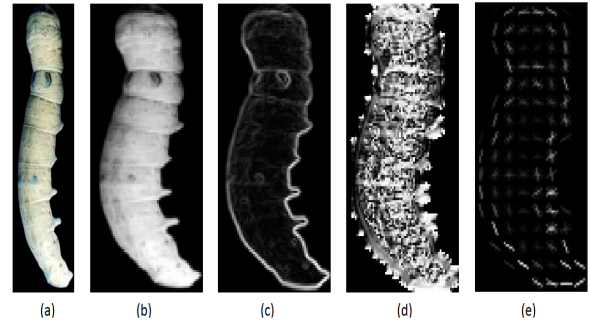


Figure 4. HOG feature descriptor of infected silkworm a) Original Image b) resized image c) magnitude visualization d) orientation visualization e) HOG feature descriptor.

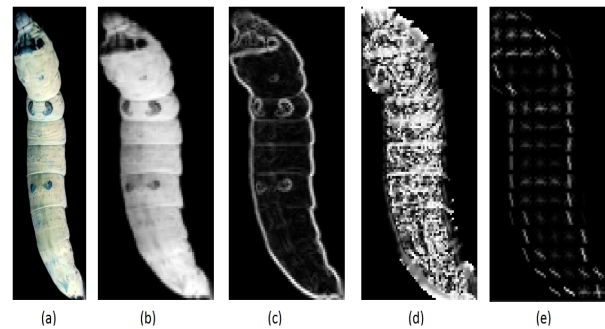


Figure 5. HOG feature descriptor of healthy silkworm a) Original Image b) resized image c) magnitude visualization d) orientation visualization e) HOG feature descriptor.

2.3 Dimensionality Reduction

After extracting HOG features, the next crucial step is dimensionality reduction. This process is essential if the feature extraction process is giving a high dimensional feature vector. The feature vectors generated from the HOG descriptors are often high-dimensional, with numerous components. While this wealth of features can be beneficial, it can also lead to increased computational demands and potentially introduce noise. In dimensionality reduction technique, the algorithm selects best features and reduces the overall data dimensions. This strategy streamlines the training process and reduces the likelihood of overfitting, leading to enhanced model accuracy and reliability in distinguishing between healthy and diseased silkworms. In this work 3 non-linear dimensionality reduction methods, KPCA, multidimensional scaling (MDS) and Isomap were studied and compared.

KPCA, MDS, and Isomap are non-linear dimensionality reduction methods that help address the complexities of high-dimensional data by preserving important features, relationships, or distances in a lower-dimensional space. These techniques have a wide range of applications in various domains where capturing non-linear structures is crucial for data analysis and interpretation.

KPCA is an advanced version of Principal Component Analysis (PCA) developed to handle non-linear relationships within data. By utilizing a kernel trick, this method transforms the data into a higher-dimensional feature space allowing linear PCA to be applied in this transformed space. This technique enables the capture of intricate, non-linear interactions between data points [29]. The process of KPCA begins with the selection of a kernel function (e.g., polynomial or RBF), which transforms the original data into an expanded feature space. After this process linear PCA is performed in this space to reduce dimensionality. The principal components obtained are subsequently mapped back to the original space. KPCA finds applications in various domains, including image processing, food quality assessment, and medical diagnostics, where it is essential to extract non-linear patterns from complex, high-dimensional data [30]-[32]. Thomas and Thomas effectively applied KPCA for silkworm sex discrimination from cocoon images and attained an accuracy of 70 % [1]. Jain et al used KPCA dimensionality reduction technique in histopathological images to discriminate controls from patients with lung cancer with a better accuracy of 97 % [31].

MDS preserves the pairwise distances or dissimilarities between datapoints. The main aim is to develop a lower-dimensional representation of the data while maintaining relative distances [30]. MDS starts by constructing a dissimilarity matrix based on the pairwise dissimilarities between data points. It then attempts to determine a lower-dimensional space that most accurately reflects these dissimilarities. Various optimization techniques can be used to achieve this. It helps reveal underlying structures and relationships in data by compressing it into a lower-dimensional form while retaining the original dissimilarities. Thomas and Thomas utilized MDS integrated with ensemble machine learning methods in classification of male and female silkworm pupa from cocoon images and attained an accuracy of 96 % [9]. Iqtidar et al. employed MDS feature extraction technique to classify different coronary artery diseases from Phonocardiogram (PCG) signal analysis and obtained an accuracy score of 98 % with 10-fold cross validation [33].

Isomap is a non-linear method that aims to preserve geodesic distances or the intrinsic geometry of the data. It connects data points in a graph, considering their shortest path distances in a manifold [31]. Isomap starts by constructing a neighborhood graph, connecting data points based on their nearness. It then computes geodesic distances (shortest path distances) along this graph. Finally, it applies classical MDS to create a lower-dimensional representation that preserves the geodesic distances. It is particularly effective when data resides on a non-linear manifold and euclidean distances fail to capture the true relationships between data points. Gunasundari and Thiagarajan employed Isomap and classical machine learning techniques for skin cancer classification from images and obtained an accuracy of 99% [34].

2.4 Classification

Two classification algorithms, Decision Tree (DT) and K Nearest Neighbors (KNN) are used in this work.

A decision tree is a supervised machine learning classification algorithm where labelled data is used for the classification. It is designed to model data through a tree-like structure of decisions based on feature values. This approach helps in predicting the target variable by partitioning the feature space into regions that are as uniform as possible in terms of their class labels [30]. In a decision tree, the process begins at the root node, which contains the complete dataset. The data is then repeatedly partitioned according to different feature values. Internal nodes apply a decision criterion based on a chosen feature, with branches denoting the corresponding outcomes. The leaf nodes, located at the ends of the branches, represent the final output, providing either the predicted class label or a regression value.

To determine the optimal splits at each node, decision trees utilize criteria such as Gini impurity, entropy, information gain, and chi-square. Gini impurity measures the purity of a node and is calculated as Formula 10.

$$Gini(t) = 1 - \sum_{i=1}^c P_i^2 \quad (10)$$

The entropy and information gain can be calculated as Formulas 11 and 12.

$$Entropy = - \sum_{i=1}^c P_i \log_2 P_i \quad (11)$$

P_i is the probability of class i . c is the total number of classes.

$$Information\ Gain = Entropy - [(weighted_average \times entropy(each_attribute))] \quad (12)$$

The decision tree is constructed by starting with the entire dataset at the root node and then applying the selected splitting criteria to partition the data recursively into subsets. This process continues until a stopping criterion is met, such as reaching a maximum depth or having a minimum number of samples in a leaf node. To prevent overfitting, the tree can be pruned, which involves removing nodes or subtrees that contribute little to the model's predictive power. Pruning methods can include pre-pruning, where the tree is stopped from growing based on certain conditions, or post-pruning, where a fully grown tree is trimmed based on validation performance or complexity considerations. In practice, decision trees are used in various applications due to their interpretability and ability to handle non-linear relationships [31].

KNN is a supervised learning classification method which relies on a similarity measure to make classifications. In the KNN algorithm, classification is determined by a majority vote of the nearest

neighboring points, determined through this similarity measure [33], [34].

Given a dataset where X denotes the set of observed features and Y represents the corresponding class labels, the KNN algorithm functions as a classification technique that predicts the class membership of a new observation based on similarity. For a chosen positive integer K , the method locates the K training samples that are most similar (or nearest) to the test point x . It then estimates the probability that x belongs to a particular class j by evaluating the proportion of those K neighbors that fall into class j , as expressed in Formula 13. The test observation is ultimately assigned to the class with the highest estimated probability.

$$(Y = j|X = x) = \frac{1}{K} \sum_{i \in A} I(Y^{(i)} = j) \quad (13)$$

Here, X represents the matrix of input features, while Y indicates the corresponding class labels. The symbol j refers to a particular class under consideration, and K denotes the number of nearest neighbors used in the classification process. The set A contains the K data points that are closest to the test sample, and I is an indicator function used to determine whether a given data point belongs to class j . If I equals 1, the data point is classified as belonging to class j ; if I equals 0, it does not belong to that class. It is important to note that choosing a smaller value of K generally results in lower bias but higher variance, whereas using a larger K reduces variance but introduces higher bias in the classification performance [35], [36].

Decision tree and KNN algorithms have given impressive accuracy results in silkworm-based studies. In silkworm sex classification studies these algorithms have provided more than 95 % accuracy when integrated with HOG feature extraction method [1], [21]. In disease diagnosis problems when integrated with feature extraction and dimensionality reduction methods, KNN and decision tree classification models gives good classification results [9], [23], [36]. Both are proven techniques in disease diagnosis specific applications.

3. Results and Discussions

3.1 Performance Evaluation Metrics

To examine the efficiency of our trained machine learning model, we use a variety of metrics. For assessing predictions on new data or images, we utilize a confusion matrix as a visual method for measuring the model's effectiveness. The confusion matrix includes four essential elements: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP and TN represent cases where the model's predictions match the actual outcomes. FP are instances where the model incorrectly predicts a positive outcome, while FN are cases where the model incorrectly predicts a negative outcome. From this confusion matrix, we can calculate key performance metrics which are given in Formulas 14 to 18.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (14)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (15)$$

$$\text{Recall or Sensitivity} = \frac{TP}{TP+FN} \quad (16)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (17)$$

$$f1 \text{ score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (18)$$

Furthermore, ROC curve is used to conduct an additional performance assessment. From this curve, AUC is computed, which serves as one of the important measure of the model's overall effectiveness. The ROC curve demonstrates how well the model distinguishes between different classes, and a higher AUC score indicates stronger predictive capability and improved class discrimination. ROC-AUC is a visual method that shows the accuracy of a machine learning model.

3.2 Disease Detection Using Decision Tree

In this study, we utilized the cross-validation method, k-fold cross-validation for evaluating the model's performance. The classifier performance is analysed using 10-fold cross-validation technique. In this method, the sample is divided into ten equal parts and in the first iteration nine folds are used for training and the tenth fold is used for testing. This process is repeated ten times such that each fold will be used as test set once. The average accuracy obtained in this method is the accuracy score of a machine learning model. In disease specific applications employing supervised classification algorithms, 10-fold method have given better classification results.

The evaluation of our proposed decision tree classifier, which utilizes HOG feature descriptors along with three different dimensionality reduction techniques, is summarized in Table 1.

Table 1. Classification results of DT

Performance Metrics	Classifier		
	HOG-KPCA-DT	HOG-MDS-DT	HOG-Isomap-DT
Accuracy (%)	94.28	91.06	91.87
Recall (%)	94.56	91.46	92.16
Specificity (%)	93.86	90.72	91.62
Precision (%)	92.48	89.44	90.43
f1 score (%)	93.45	90.44	91.29

Among these approaches, the HOG-KPCA-DT classifier stands out by delivering the best classification performance compared to both HOG-MDS-DT and HOG-Isomap-DT. This is evident from its high sensitivity of 94.56% and an accuracy of 94.28%. Additionally, the classifier exhibits impressive precision, F1 score, and specificity, with values of 92.48%, 93.45%, and 93.86%, respectively. Figure 6 illustrates the ROC curve of the performance of the HOG-KPCA-DT model. The mean AUC score, shown in Figure 6, is 0.95, highlighting the model's robust classification capability. The HOG-MDS-DT model

demonstrates the lowest performance among the evaluated methods, with an accuracy of only 91.06%. In addition to accuracy, the other performance metrics such as precision, recall, and F1-score are also comparatively lower for this model. This suggests that the combination of HOG feature extraction with MDS for dimensionality reduction and DT classification may not be as effective in capturing the discriminative features of the data when compared to the other approaches considered in the study.

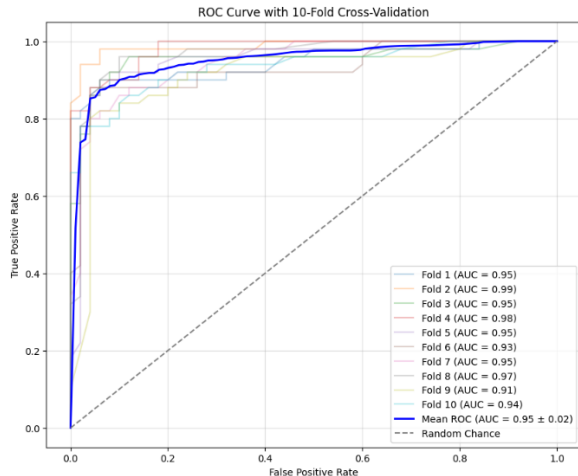


Figure 6. ROC curve for HOG-KPCA-DT

3.3 Disease Detection Using KNN

In Table 2, we detail the classification assessment of our proposed KNN classifier for disease detection, using HOG feature descriptors and three dimensionality reduction techniques: KPCA, MDA, and Isomap. Among these different approaches, the HOG-KPCA-KNN classifier performed well, displaying superior classification performance in comparison to HOG-MDS-KNN and HOG-Isomap-KNN.

Table 2. Classification results of KNN

Performance Metrics	Classifier		
	HOG-KPCA-KNN	HOG-MDS-KNN	HOG-Isomap-KNN
Accuracy (%)	92.59	90.34	91.22
Recall (%)	92.86	90.77	91.29
Specificity (%)	92.37	89.97	91.17
Precision (%)	91.27	88.61	89.88
f1 score (%)	92.06	89.67	90.58

This superiority is evidenced by its remarkable sensitivity of 92.37% and an impressive accuracy of 92.59%. Moreover, it exhibits noteworthy precision, F1 score, and specificity, with values reaching 91.27%, 92.06%, and 92.37%, respectively. HOG-MDS-KNN have given the lowest score for all the performance parameters, with the accuracy of 90.34% and recall of 90.77%. In Figure 7, we present an ROC curve, offering a overview of the classification performance achieved by the HOG-KPCA-KNN model. Importantly, the mean AUC score, as indicated in Figure 7, is computed at 0.93, further confirming the model's robust classification capabilities.

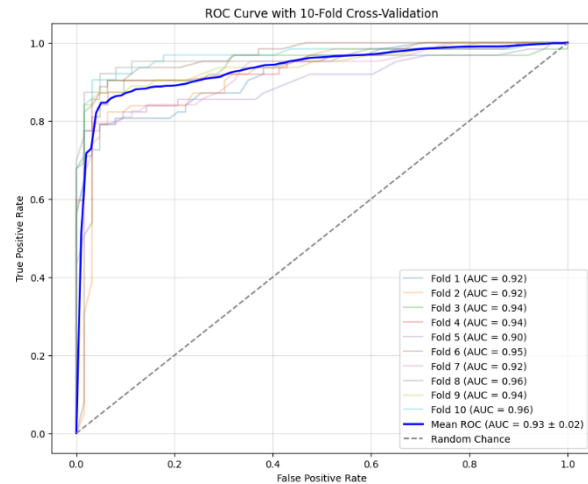


Figure 7. ROC curve for HOG-KPCA-KNN

3.4 Comparative Analysis

In this research, we have conducted an analysis of the HOG feature descriptor integrated with KPCA dimensionality reduction, which has yielded promising results. However, we have also explored the performance of four other classifier models for a comprehensive evaluation. The six classifier models under investigation are HOG-KPCA-DT, HOG-MDS-DT, HOG-Isomap-DT, HOG-KPCA-KNN, HOG-MDS-KNN, and HOG-Isomap-KNN. In Figures 8 and 9, we have presented bar diagrams that illustrate the evaluation metrics for all three DT models and KNN models, respectively.

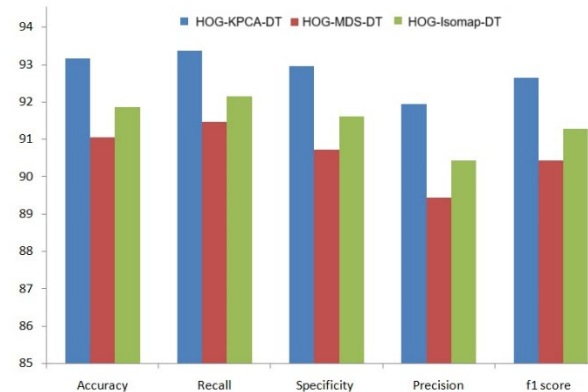


Figure 8. Comparison of evaluation metrics for DT models

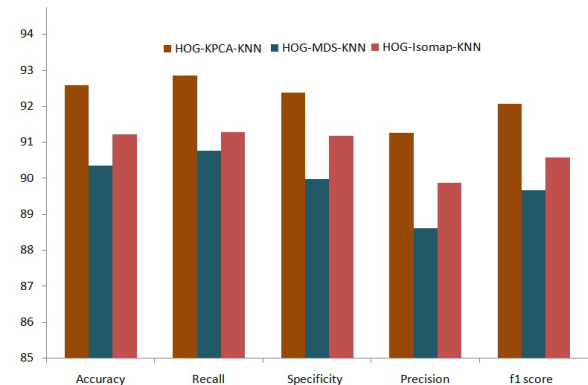


Figure 9. Comparison of evaluation metrics for KNN models

Notably, the evaluation metrics of HOG-KPCA-DT outperform the other three classifiers. Additionally, HOG-Isomap demonstrates superior performance in comparison to HOG-MDS across both classification schemes, achieving an accuracy of 91.87% with DT and 91.22% with KNN. These findings provide a comprehensive view of the classifiers' relative performance

3.5. Performance Evaluation Against Existing Methods

In comparison to prior research on Grasserie disease detection in silkworms, the proposed method demonstrates competitive and consistent performance. The comparative analysis is shown in table 3. Nagashetti et al. attained a slightly higher accuracy of 97% using deep learning methods without dimensionality reduction on a larger dataset of 670 healthy and 527 infected silkworms [5]. Suthar et al. achieved 94% accuracy using EfficientNet, a CNN, on a balanced dataset of 200 healthy and 200 infected silkworms [13]. Shilpashree et al., using GLCM-based texture features and the Random Forest model, reported a notably lower accuracy of 59%, indicating limitations in traditional feature-based approaches [7]. Zhen et al. employed an advanced AC-DenseNet model to classify multiple silkworm diseases, achieving 85.6% peak accuracy, which is notably lower than the result achieved in this study [6]. Our study is an effective and interpretable approach combining HOG for feature extraction, KPCA for dimensionality reduction, and DT for classification. This shows that classical machine learning methods, when combined with good feature extraction and dimensionality reduction, can give accurate results for early detection of Grasserie disease in silkworms.

Table 3. Comparative analysis against existing models

Study	Dataset Size (Healthy / Infected)	Method / Model Used	Accuracy (%)
Nagashetti et al. [5]	670 / 527	Deep Learning	97
Suthar et al. [13]	200 / 200	End-to-end learning and EfficientNet (CNN)	94
Shilpashree et al. [7]	245/247	GLCM texture features and Random Forest	59
Zhen et al. [6]	989/1045	AC-DenseNet (Attention-Concatenation CNN)	86
This work	668/574	HOG-KPCA-DT	94

The findings from our work demonstrate the effectiveness of combining HOG feature descriptors with advanced dimensionality reduction techniques and machine learning algorithms for automating the detection of Grasserie disease in silkworms. By evaluating different model configurations, the results indicate that certain models, such as HOG-KPCA-DT, offer significant improvements in classification performance. These outcomes contribute to the ongoing efforts in developing accurate, efficient, and scalable

disease detection methods in sericulture. The use of such automated systems has the potential to not only reduce the dependency on manual inspections but also to improve overall precision and accuracy of disease management, leading to more sustainable and productive sericulture practices. This study underscores the importance of the integration of AI and image analysis to address key challenges in silkworm health monitoring.

Although Deep learning approaches have given better results in several domains including disease diagnosis, we selected machine learning models in this work due to the relatively limited dataset size, need for interpretable features, and feasibility of deployment in resource-constrained sericulture settings. Traditional machine learning models gives good performance under these conditions and can be efficiently implemented in portable diagnostic systems. Deep learning approaches remain a promising future direction once larger annotated datasets become available. While deep learning studies have reported slightly higher accuracy, our machine learning methods have advantages in terms of computational efficiency and feasibility. The proposed machine learning model required only a few seconds of training on a standard CPU and provided near-instant inference, making it practical for deployment in sericulture farms where advanced hardware may not be available. In contrast, deep learning approaches often demand extensive training time, dedicated GPUs, and higher costs, which limit their scalability for real-world field applications. This highlights the contribution of our research in providing a lightweight, interpretable, and cost-effective solution for early Grasserie disease detection.

A limitation of this study is that images were acquired against a black background to maintain consistency and reduce noise, which may affect the generalizability of the model in real-world environments. Future work will address this by incorporating images with varied and natural backgrounds.

4. Conclusions

In this paper, we presented an automated method for the detection of Grasserie disease in silkworms using machine learning techniques combined with HOG for feature extraction and dimensionality reduction methods such as KPCA. The proposed system aims to address the disadvantages of traditional techniques that rely on manual inspection and microscopic examination, offering a faster, more accurate alternative. By leveraging image analysis and advanced classification techniques, we evaluated the performance of several models, including DT and KNN, with the HOG-KPCA-DT classifier achieving the best results in terms of classification accuracy and sensitivity.

The experimental results demonstrate that combining HOG with KPCA and machine learning classifiers significantly improves the detection of Grasserie disease in silkworms. The HOG-KPCA-DT model, in

particular, outperformed other configurations, emphasizing the critical role of choosing suitable feature extraction and dimensionality reduction methods. This automated approach reduces the need for manual labor in sericulture disease detection and offers a more scalable solution for larger sericulture operations. The promising results of this work underscore the possibilities of AI methods into the management of silkworm health. Future studies will enhance the dataset and incorporate additional disease types for a more comprehensive system, as well as exploring real-time detection capabilities through mobile applications. Ultimately, this research contributes to enhancing the efficiency and productivity of the sericulture industry by providing a reliable, automated disease detection tool.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- [1] S Thomas and J Thomas, "Nondestructive and cost-effective silkworm, *Bombyx mori* (Lepidoptera: Bombycidae) cocoon sex classification using machine learning". *Int. J. Trop. Insect Sci.*, vol.44, pp. 1125–1137, March 2024, doi: 10.1007/s42690-024-01207-7.
- [2] T Chauhan and M Tayal, "Mulberry sericulture. *Industrial entomology*, pp.197-263, Feb 2017, doi: 10.1007/978-981-10-3304-9_8.
- [3] H He, Zhu S, Shen L, Chang X, Wang Y, Zeng D, Xiong B, Dai F, Zhao T. Integrated Analysis of Machine Learning and Deep Learning in Silkworm Pupae (*Bombyx mori*) Species and Sex Identification. *Animals*. Vol. 13, no. 23, pp.3612, Nov 2023, doi: 10.3390/ani13233612.
- [4] S Thomas and J Thomas, "A review on existing methods and classification algorithms used for sex determination of silkworm in sericulture". *International Conference on Intelligent Systems Design and Applications*, Cham: Springer International Publishing, Dec 2020, doi: 10.1007/978-3-030-71187-0_52.
- [5] S Nagashetti, S Biradar, S Dambal, C Raghavendra and B Parameshachari, "Detection of disease in *Bombyx mori* silkworm by using image analysis approach. In *2021 IEEE Mysore Sub Section International Conference (MysuruCon)* (pp. 440-444). IEEE, Oct 2021, doi: 10.1109/MysuruCon52639.2021.9641676.
- [6] Y Zhen, X Dingyuan, T Liang and S Meining, 2020, "Attention-Concatenation Dense Convolutional Neural Network for Silkworm Disease Recognition". *5th International Conference on Smart Grid and Electrical Automation (ICSGEA)* (pp. 254-259). IEEE, Jun 2020, doi: 10.1109/ICSGEA51094.2020.00061.
- [7] P Shilpashree, A Karegowd, K Suresh and L Rani. "Classification of Healthy and Diseased Silkworms using Ensemble Learning and CNN". *International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)* (pp. 1-5). IEEE, Jul 2023, doi: 10.1109/ICSSES58299.2023.10200374.
- [8] M Hăbeanu, A Gheorghe and T Mihalcea. "Silkworm *Bombyx mori*—sustainability and economic opportunity, particularly for Romania". *Agriculture*, vol. 13, no.6, pp.1209, Jun 2023, doi: 10.3390/agriculture13061209.
- [9] S Thomas and J Thomas, "Non-destructive silkworm pupa gender classification using X-ray images using ensemble learning", *Artif Intell Agri*. vol. 6, pp. 100-10, Jan 2022, doi: 10.1016/j.aiia.2022.08.001.
- [10] X Wang, K Wu, H Pang, P Xu, M Li. And G Zhang. "Study on the role of cytc in response to BmNPV infection in silkworm, *Bombyx mori* (Lepidoptera)". *Int. J. Mol. Sci.*, vol. 20, no. 18), p.4325, Sep 2019, doi: 10.3390/ijms20184325.
- [11] A Silva, E Costa, S Reis, C Spencer, R Calhelha, S Miguel, M Ribeiro, L Barros, J Vaz and P Coutinho, P, "Silk sericin: A promising sustainable biomaterial for biomedical and pharmaceutical applications". *Polymers*, vol. 14, no. 22, p.4931, Nov 2022, doi: 10.3390/polym14224931.
- [12] M Gani, S Senger, S Lokanath, P Saini, K Bali, R Gupta, V Sivaprasad, J Jehle and J Wennmann. "Patterns in genotype composition of Indian isolates of the *Bombyx mori* nucleopolyhedrovirus and *Bombyx mori* bidensovirus". *Viruses*, vol. 13, no. 5, p.901, May 2021, doi: 10.3390/v13050901.
- [13] S R Suthar, S Devaraj, M R Maanasa, K V Vidya, N Siby, and B R Revanna, "Silk Shield: Empowering Sericulture with AI-Driven Disease Detection". *International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS) 2024* Nov 7 (pp. 1-6). IEEE. doi: 10.1109/CSITSS64042.2024.10816994
- [14] V A Binson and G Manju, "Automated Disease Detection in Silkworms Using Machine Learning Techniques". *ASSET*. Vol. 6, no. 4, pp. 02404015-02404015, Sep 2024, doi: 10.26877/asset.v6i4.965.
- [15] S Thomas and J Thomas, "Comparative Analysis of Non-Destructive Silkworm Cocoon Sex Classification using Machine Learning Models Based on X-Ray and Camera Images". *International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, (pp. 1-7). IEEE., Aug 2024, doi: 10.1109/IACIS61494.2024.10721914.
- [16] G Ganga, An introduction to sericulture". Oxford and IBH Publishing, Mar 2019.
- [17] A Kalagi, D Raghavan, C Raghavendra, S Bajannavar and V Bhavani. "Sericulture Technology Towards Sustainable Management". *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1-5). IEEE, Apr 2022, doi: 10.1109/ICDCECE53908.2022.9793139.
- [18] D Boucias and J Pendland. "Principles of insect pathology". Springer Science & Business Media, Dec 2012.
- [19] D Xia, Z Yu, A Cheng, L Tang and M Shi. "Development and Application of Silkworm Disease Recognition System Based on Mobile App". In *Image and Graphics: 10th International Conference, ICIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part II* 10 (pp. 471-482). Springer International Publishing, Aug 2019, doi: 10.1007/978-3-030-34110-7_39.
- [20] A Begum, B Rakesh, P Rao and M Mamatha. "Biochemical analysis of fat bodies of the popular silkworm breeds/hybrids and foundation crosses". *Indian J Anim Res*. vol. 48, no. 2, pp.134-138, May 2014, doi: 10.5958/j.0976-0555.48.2.029.
- [21] S Thomas and J Thomas, "An optimized method for mulberry silkworm, *Bombyx mori* (Bombycidae: Lepidoptera) sex classification using TLBPSSGA-RFEXGBoost". *Bio open*. vol. 13, no. 7, 060468, July 2024, doi: 10.1242/bio.060468.
- [22] L Satish, L Kusuma, A Shery, S Moorthy, G Manjunatha and V Sivaprasad, "Development of productive multi-viral disease-tolerant bivoltine silkworm breeds of *Bombyx mori* (Lepidoptera: Bombycidae)". *App. Entomo. Zool*. vol. 58, no. 1, pp.61-71, Feb 2021, doi: 10.1007/s13355-022-00803-8.
- [23] V A Binson, S Thomas, M Subramoniam, J Arun, S Naveen and S Madhu. "Review of Machine Learning Algorithms for Biomedical Applications" *Annal. Biomed. Eng*. vol. 52, no. 5, pp.1159-83, May 2024, doi: 10.1007/s10439-024-03459-3.
- [24] V A Binson, S Thomas, G Manju, KS SK. Precision crop prediction using IoT-enabled soil sensors and Machine Learning. *ITEGAM-JETIA*. Vol. 10, no.49, pp. 65-70, Sep 2024, doi: 10.5935/jetia.v10i49.1219.
- [25] R McConnell. "Method of and apparatus for pattern recognition (No. US 4567610). 1986.
- [26] N Dalal and b Triggs, "Histograms of oriented gradients for human detection". *IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee, Jun 2005, doi: 10.1109/CVPR.2005.177.
- [27] S Bakheet, "An SVM framework for malignant melanoma detection based on optimized HOG features". *Computation*. vol. 5, no. 1, pp. 4, Jan 2017, doi.org/10.3390/computation5010004

- [28] W S Admass, Y Y Munaye and G A Bogale, "Convolutional neural networks and histogram-oriented gradients: a hybrid approach for automatic mango disease detection and classification". *Int. J. Inform. Technol.*, vol. 16, no. 2, pp.817-829, Feb 2024, doi.org/10.1007/s41870-023-01605-z
- [29] F Anowar, S Sadaoui and B Selim, B, "Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne)". *Computer Science Review*, vol. 40, p.100378, May 2021, doi: 10.1016/j.cosrev.2021.100378.
- [30] R Oktafiani, A Hermawan, and D Avianto. Max Depth Impact on Heart Disease Classification: Decision Tree and Random Forest. *J. RESTI (Rekayasa Sistem Teknol. Inf.)* vol.8, no.1, pp.:160-168, Feb 2024, doi: 10.29207/resti.v8i1.5574.
- [31] D K Jain, K M Lakshmi, K P Varma, M Ramachandran, and S Bharati, "Lung Cancer Detection Based on Kernel PCA-Convolution Neural Network Feature Extraction and Classification by Fast Deep Belief Neural Network in Disease Management Using Multimedia Data Sources". *Comput. Intell. Neurosci.* May2022;doi: https://doi.org/10.1155/2022/3149406.
- [32] P N Andono, and S H Nugraini. Texture feature extraction in grape image classification using k-nearest neighbor. *J. RESTI (Rekayasa Sistem Teknol. Inf.)*. vol. 6, no.5, pp. 768-775, Oct 2022, doi: 10.29207/resti.v6i5.4137.
- [33] K Iqtidar, UQamar, S Aziz, and M U Khan, "Phonocardiogram signal analysis for classification of Coronary Artery Diseases using MFCC and 1D adaptive local ternary patterns", *Comput. Biol. Med.*, vol. 138, pp.104926, Nov 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104926.
- [33] E Laksono, A Basuki, and F Bachtiar. Optimization of k value in knn algorithm for spam and ham email classification. *J. RESTI (Rekayasa Sistem Teknol. Inf.)*. vol. 4, no. 2, pp. 377-383, Apr 2020, doi: 10.29207/resti.v4i2.1845.
- [34] B Gunasundari and R Thiagarajan, "Evaluation of high-dimensional data classification for skin malignancy detection using DL-based techniques", *Cancer Investigation*. Vol. 42, no. 5, pp. 365-389, May 2024, doi: https://doi.org/10.1080/07357907.2024.2345184
- [35] V. A. Binson, M. Subramoniam and Luke Mathew, " Prediction of lung cancer with a sensor array based e-nose system using machine learning methods.", *Microsys. Technol.*, vol. 30, pp. 1421-1434, Apr 2024, doi: 10.1007/s00542-024-05656-5.
- [36] S Zhang, X Li, M Zong, X Zhu, D Cheng. Learning k for knn classification. *ACM Trans. Intell. Syst. Technol.*, vol. 12, no.3, pp. 1-9, Jan 2017, doi: 10.1145/2990508.