International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Region Based Segmentation with Enhanced Adaptive Histogram Equalization Model with Definite Feature Set for Sugarcane Leaf Disease Classification

A. Vivek Reddy[1], R. Thiruvengatanadhan[2], M. Srinivas[3], P. Dhanalakshmi[4]

Submitted: 28/06/2023 **Revised**: 07/08/2023 **Accepted**: 26/08/2023

Abstract: Visual identification of plant diseases is a time-consuming process that yields inaccurate results and is only feasible in small settings. Instead, an autonomous detection method would require less time and manpower while also improving accuracy. Brown and yellow spots, late and early scorch, and other fungal, viral, and bacterial diseases are only a few of the more common plant ailments. Manually detecting the disease as well as the type of disease requires analyzing the color degradation in a diseased leaf or plant. This research will automate the human-performed step of disease identification and instill the methods by which humans recognize diseases from healthy plants. The proposed model after enhancing the image quality, features is extracted and relevant features are selected. The proposed model uses Enhanced K Nearest Neighbor (EKNN) model for accurate classification of disease and non disease leaves. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization based Image Quality Enhancement model with Definite Feature Set model using EKNN (RbS-EAHE-EKNN-LDC) for Leaf Disease Classification is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for accurate disease or non disease classification. The proposed model is contrasted with the state of the art models and the results represent that the proposed model performance is enhanced.

***Keywords**: Sugarcane Leaf, Image Processing, Segmentation, Histogram Equalization, Leaf Features, Feature Set, Classification, Disease Detection, Quality Enhancemen.*

1. Introduction

Agriculture is the main source of income for most people in India. Agricultural land accounts for more than half of India's total land area [1]. It is safe to argue that the growth of India's economy is directly tied to the success of its agricultural sectors [2]. Therefore, intensive care is required for plants and crops. Infectious diseases are a major cause of crop failure. Sugarcane is a crucial crop for many nations, including India [3]. Despite the difficulties, India's sugarcane industry continues to be an important part of the country's economic foundation. Most sugarcane diseases are caused by fungi and manifest as spots and stripes on the leaves [4]. These spots hinder the plant's ability to produce food via photosynthesis, which drastically reduces its growth and productivity. In the most extreme cases of infection, the leaf gets completely covered in spots. Eventually, the plant shrivels up and dies. The disease spreads by the airborne spores. A sugarcane farm may suffer catastrophic losses if a widespread infection were not promptly addressed. Taking the right steps to prevent sugarcane diseases is essential for keeping costs down [5].

Research into machine vision and Artificial Intelligence (AI) is being conducted at the moment in order to realize the goal of intelligent farming. In order to effectively manage crop diseases, early detection of symptoms is essential [6]. Some diseases have preventative or corrective treatments that can be implemented if the symptoms are caught early. The use of image processing software has advanced greatly in the realm of agricultural study. An automated approach based on image processing and pattern recognition has been developed to detect and categorize sugarcane fungal diseases [7]. Disease spots come in a wide variety, and the novice could mistake one for another if they aren't carefully identified. To make matters worse, if the wrong place is targeted with fungicide, the plant will not be protected and the disease will have more time to spread, resulting in a significant financial loss [8].

Diseases in plants can be defined as any disruption in a plant's normal physiological processes, whether caused

[1]Research Scholar, Department of CSE,Annamalai University Annamalai Nagar, Tamilnadu, 608002, India.
Email Id: ambativivekreddy@gmail.com
[2]Department of CSE, Annamalai University Annamalai Nagar, Tamilnadu, 608002, India.
Email Id: thiruvengatanadhan01@gmail.com
[3]Department of CSE, St.mary's Group Of Institutions - Hyderabad, Deshmukhi, Telangana,508284, India.
Email Id:sreenivasmeher@gmail.com
[4]Department of CSE, Annamalai University Annamalai Nagar, Tamilnadu, 608002, India.
Email Id: abidhana01@gmail.com

by biotic or abiotic factors, that threatens the plant's capacity for normal growth and reproduction [9]. Temperature extremes, too little or too much water, the wrong soil pH, a lack of necessary minerals, and other environmental factors all contribute to the spread of plant diseases. Infectious plant diseases and non-infectious plant diseases are the two broad categories into which plant diseases can be placed [10]. Bacteria, fungi, and viruses are the root of many infectious disorders [11]. Those that don't spread disease are brought on by unfavourable climatic conditions. The sugarcane leaf with diseases sample images are shown in Figure 1.

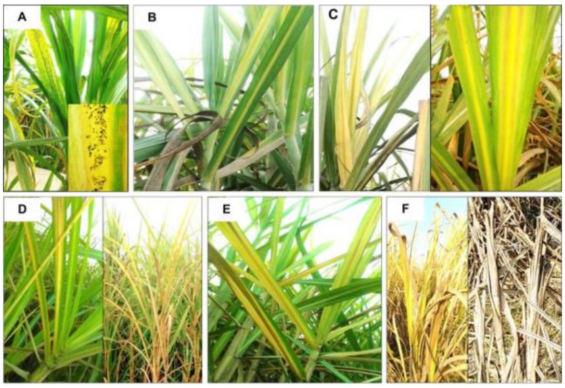

Fig 1: Sugarcane Leaf Diseases

Automatic disease detection in sugarcane, which can boost yield and quality, relies heavily on computer methods [12]. The diseases that affected sugarcane plants were identified by extracting features from Region of Interest (RoI) [13] and analyzing the same using computational methods, which required the use of image processing methods such as pre processing for noise removal [14], segmentation for the identification of the region of interest, features extraction, and classification [15]. Images of sugarcane leaves were often obtained in the field with sensors and cameras. Preprocessing procedures for photographs improve and clean them up [16]. Segmentation algorithms in images locate the infected area from which features can be extracted [17]. Diseases affecting sugarcane plants are now properly identified by categorization methods. In this research, initially denoising is applied on the images and then region based segmentation model is considered for extracting the features from a specific region of a leaf for accurate disease detection [18].

Leaf images can be segmented into many regions or objects by using either a similarity or a discontinuity criterion. Both the exterior and the interior of a leaf image can be used to define a region with the same level of precision. The goal of region-based segmentation is to separate a picture into groups of pixels that have comparable features. The program looks for clusters of pixels, or regions, starting with a given initial pixel set. After locating the initial pixel location, the algorithm can expand a region by adding new pixels or contract it by merging it with neighbouring regions. For region-based segmentation to be effective, all pixels must be assigned to an area, and all points inside a region must be connected in a certain way, which also specifies that the regions cannot overlap with one another that are in same grey level range.

After the denoising process on all images is completed, then these denoised images are segmented into regions using region based segmentation model for accurate feature extraction. To perform denoising operation, the proposed model used mean filtering. The mean filter is a spatial filter that works by swapping the central value of a sliding window for its average mean value. Although square is the most common shape for the window, or kernel, it can be any size or shape. As its name implies, the mean filter takes the average of neighbouring pixels to replace their individual values.

Leaf Images can have their contrast improved with the use of a technique called Adaptive Histogram Equalization (AHE). One way in which adaptive histogram equalization differs from conventional methods is that it can be used to boost contrast in certain regions [19]. In order to rebalance the brightness of an image, AHE generates many histograms, each corresponding to a different region. Common histogram equalization applies the same change to every pixel in the image [20]. This method is effective when the image's pixel values are uniformly distributed. There will be insufficient contrast improvement, however, if the image has areas that are noticeably brighter or darker than the rest of the image [21]. To remedy this, EAHE applies a transformation function drawn from the surrounding area to each pixel. The basic idea is to alter each pixel using the histogram of the square that contains it.

Before performing the segmentation process, image denoising model is applied on the considered images for removing the noise and enhancing the image quality. The goal of image denoising is to recover an original image from one that has been corrupted by noise [22]. However, noise, edge, and smoothness are high frequency characteristics, making it challenging to identify them during denoising and potentially leading to a loss of detail in the denoised images. In order to produce high-quality photos, removing unwanted noise from them is a necessary step [23]. However, this process often leaves behind potentially useful information.

The enhanced image is provided as input data that is turned into a condensed set of features [24]. Feature extraction is the process of converting raw data into a usable set of characteristics. Careful feature selection ensures that the extracted features will draw out the correct information from the given input data [25]. The process of feature extraction follows after segmentation [26]. The features of a leaf image that can aid in precise classification can be extracted using the feature extraction method [27]. Various characteristics, such as their coherence, intensity, homogeneity, etc., are examined. KNN is an image classification approach that makes use of the closest distance between the training dataset and the testing dataset to determine an image's classification. Finding the right value for k might be a challenge when using KNN. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization based Image Quality Enhancement model with Definite Feature Set model using EKNN for Leaf Disease Classification is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for accurate disease or non disease classification.

2. Literature Survey

Image classification methods are widely employed for agricultural data analysis, however they fall short when it comes to reliably recognising unhealthy regions on individual plant leaves that correspond to distinct disease kinds. Phan et al. [2] applied Simple Linear Iterative Clustering (SLIC) segmentation to produce super-pixels, clusters of pixels that reflect regions of interest on maize leaves, from pictures in the PlantVillage and CD&S datasets. Plant diseases are a major challenge to agricultural productivity. One of the most pressing questions in crop disease segmentation is how to accurately represent the disease outside appearance while also preserving all of the colour and texture data from the affected area. Yuan et al. [3] offered a spatial pyramid-oriented encoder-decoder cascade convolution neural network-based for crop disease leaf segmentation, which aims to solve the issue of low segmentation accuracy with classic convolution neural network-based approaches in the picture of the crop diseased leaves. The network is made up of two sub-networks: one for detecting diseases in certain regions, and another for segmenting those diseases. An example of a network that combines cascading convolution neural networks with spatial pyramids is the region disease detection network. This technique establishes a link between the simple to complicated structures of three-level convolution neural network models. The many stages of a neural network allow for the extraction of a variety of properties from the leaves of crops with disease. Screening pictures is the final step in the detection of crop leaf diseases. Each layer of the network additionally features a space pyramid pooling layer. As the size of the input model can be varied using this pooling method, it is more generalizable. The Encoder-Decoder architecture is the basis for the region segmentation network.

Grape leaf diseases are particularly harmful because of how rapidly they may spread. This emphasises the urgency of early detection and intervention. In order to detect, identify, and categorise three common illnesses in grape leaves, an unique Enhanced VGG16 model is proposed by Mousavi et al. [4]. Powdery mildew, anthracnose, and downy mildew are all examples of such diseases. The information is collected mechanically with the aid of photographs taken by a quadcopter in a vineyard. Diseases can be detected by analysing photographs sent from the quadcopter by the main central system. To combat these three diseases, the system communicates with a hexacopter to spray pesticide in the affected areas of the grapevine garden.

Many significant grain-producing countries halted exports after the worldwide spread of COVID-19. Fear for the future of the world's food supply has been sparked by this. How to boost grain production is a major issue for all countries. To detect illnesses in tomato leaves, Zhou et al. [5] developed a reorganized residual dense network. Improved computation accuracy, reduced training time, and faster data and gradient distribution are just some of the benefits of this hybrid deep learning model. Since the RDN model's first use was in picture super resolution, altering the properties and hyper parameters of the input images is required to adapt the network framework for classification tasks.

Since there is a great deal of variation among photos of leaf diseases within a given class but only a moderate amount of variation across classes, it is crucial to accurately portray the characteristics of the region in which the target is located. Wu et al. [6] proposed using an attention network to classify diseases at a finer level of detail. Classification Model makes use of the attention process to better identify objects. Since the Reconstruction-Generation Model was introduced into the training process, the Classification Model must now focus less on global traits and more on local ones when trying to identify distinctions. The approach of generalization ability improvement improves the identification accuracy much farther than the conventional classification network. Low cost terminal real-time detection of peaches and tomato leaf diseases is possible with this technology, and less memory is required.

Image contrast isn't the only issue that has seen widespread application of HE methods; they've also been used to address the aforementioned issues. However, the returned photos are often afflicted by unwanted artefacts, strange appearances, and undesirable washed-out effects due to these methods. Fawzi et al. [8] proposed a novel method, Adaptive Clip Limit Tile Size Histogram Equalization, to address this issue (ACLTSHE). Initial optimal values for clip limit (CL) and tile size minimum/maximum are assigned by the ACLTSHE. Causes of obscured features and dimmed images include inexperienced users, low-quality devices, improper lighting conditions, and unfavourable environmental situations throughout the photography process. HE approaches have been widely employed to overcome these issues and enhance image contrast. In many cases, however, the techniques' final outputs look artificial, whether because of oversaturation or distracting artefacts. For the purpose of improving contrast, Majeed et al. [9] offered a novel method called adaptive entropy index histogram equalization (AEIHE), which falls into the local sub-class of HE-based methods. In order to better emphasize specific features within the image, AEIHE first creates three separate images.

Inadequate results are achieved by current image enhancement techniques because of the difficulty in enhancing both global and local image contrast at the same time. He et al. [11] suggested a solution based on histogram equalisation to fix this problem, and the author termed it RG-CACHE. It adjusts its level of brightness enhancement to meet the specific needs of the data at hand, so enhancing the readability of finer details without sacrificing the overall contrast. In order to achieve discriminative histogram equalisation, RG-CACHE uses density estimation to take into account the spatial information offered by the image environment. The author suggested defining spatial information using picture reflectance calculated with edge-preserving smoothing to lessen the impact of nonuniform illumination. Adaptive backdrop brightness adjustments and the exposure of previously concealed but informative image elements are two areas where RG-CACHE shines.

3. Proposed Model

Maintaining a healthy sugarcane crop necessitates attentive observation, particularly in regards to controlling the spread of diseases that have the potential to severely reduce output and lengthen the recovery period after harvest. Leaves and stems are the most common sites where pests and diseases can be found on a plant. Successful plant cultivation relies heavily on accurate leaf identification and the detection of plant diseases. Farmers use manual detection to spot signs of disease, a task that is extremely time-consuming and requires constant attention to the crop. It's important for farmers to consult with specialists in case they have questions. Whenever it is impractical to bring in other experts for a diagnosis.

An image of a disease-affected leaf is acquired by taking its RGB colour profile into account. The image contrast is improved further by EAHE. Pre-processing occurs after the image has been inserted. EAHE is used in preprocessing to lower the noise value and increase contrast in the image. The cumulative distribution function is used to spread the image's intensity in EAHE. The image's equalized intensity is then obtained via the remapping function. Image files containing photos of sugarcane leaf diseases are collected under controlled conditions. The diseased leaf has been laid out in plain view on a white background. Light sources are angled at 45 degrees on either side of the leaf to prevent shadows and provide uniform illumination throughout the surface. The leaf is magnified so that the photo shows simply the leaf on a white background. The diseased sugarcane leaf image is shown in Figure 2.

Fig 2: Diseased Leaf

The goal of image preprocessing is to either reveal previously hidden information or draw attention to specific aspects of an image. These are largely subjective operations used to modify an image so as to best exploit the human visual system's psychovisual properties. Specifically, median filtering and EAHE methods were implemented. After applying EAHE, segmenting an image is a crucial procedure because it allows to isolate the areas of interest within it; these areas of interest should not overlap, and they should also conform to certain consistency constraints. Segmentation refers to the process of separating an image into its individual features. The task at hand will determine how finely users need to subdivide. Segmentation should end after the objects of interest or regions of interest in an application have been identified. It takes multiple methods to get a good enough segmentation for recognition to be possible.

K-nearest neighbor algorithms (K-NN) are used in disease recognition to classify leaves based on the instances that are closest to them in the feature space. Unlike other instance-based learning methods, K-NN waits to perform the function and calculation until classification. For this reason, the K-nearest neighbour approach dominates all other machine learning methods. The object is given to the group to which it belongs most frequently among its K closest neighbours. If the integer K equals 1, the item is placed in the same class as its immediate neighbor. To determine the nature of the disease, this research employs a method involving four successive phases. Preprocessing, leaf segmentation, feature extraction, and classification are the four steps involved. Denoising technique is applied on the images and then noise is removed for generating clear images. Image segmentation is applied on denoised images to separate the damaged or diseased parts of the leaf from the healthy ones. To solve classification and regression problems, enhanced KNN technique is employed that contains weights, which is a directed, supervised, and advanced machine learning approach. Feature Extraction is a dimensionality reduction technique that helps express the interesting features of an image in a small feature vector. When image sizes get too big, and simplified feature representations are needed for fast image matching and retrieval, this method is quite useful.

Image contrast can be enhanced with the help of EAHE, a pre-processing technique. Redistributing the image's brightness values are accomplished by computing multiple histograms, each corresponding to a different region of the image based on the segmentation performed. Sub-histogram normalization was followed by EAHE, and then the entire image was standardized with improved quality. The image has been improved with EAHE without the drawbacks of over-enhancement, noise amplification, or intensity saturation. Intensities can be more fairly spread across the histogram's intensity range with this tweak. Because of this, places with low local contrast can improve. This is achieved by the use of EAHE, which effectively disperses the densely packed intensity values utilized to reduce visual contrast. The histogram of an image is a frequency distribution of sugar leaf grayscale values that reveals how often each value appears in the image. In an image of size 1024*1024 8-bit considered image, the abscissa values are between 0 and 255, and there are a total of 1024*1024. The images are processed in a window of 224*224 scale for processing of features. The proposed model considers the features as radius, mean, perimeter, entropy, histogram, EGB, edges, region of interest, shape, texture mean, perimeter mean. The radius represent the image pixel radius. The radius of a image is the shortest distance between two points on its perimeter and its centre. The perimeter of a image shape is a numerical value representing the whole circumference of that image. The amount of bits used to encode an image is known as its discrete entropy. A greater entropy number indicates a more refined image.

A region of interest (ROI) refers to a specific area of a picture that needs special treatment in terms of filtering or other processing. An ROI can be represented in the form of a binary mask picture. Texture metrics are calculated in image processing to quantify how an image appears to the human eye. The texture of an image tells us how the colors and brightnesses are distributed across the image or within a defined portion of the image.

The radius of an image is calculated as

$$Radius[M] = \sum_{i=1}^{M} \frac{ImageCircumference}{2*\pi} \tag{1}$$

The mean pixel intensity is calculated as

$$Mean\_Intensity[M] = \sum_{i=1}^{M} \frac{R*H}{G} \tag{2}$$

Here R is the image pixel intersected area, H is the non intersected area and G is the total area of intersected shapes.

The pixel intensity standard deviation is calculated as

$$Std[M] = \sum_{i=1}^{M} \frac{\sum(R*H*H)-\tau*\tau*G}{G} \tag{3}$$

Here τ is the mean value of intensity levels.

The EKNN algorithm was utilized to create this particular type of categorization system by allocating weights. Because it can be used to solve both classification and regression problems, the EKNN algorithm fits the definition of a supervised machine learning algorithm. Starting from the premise that "similar items are close together," this program finds clusters of related data. The EKNN algorithm begins by loading the data and setting K to the desired number of neighbours. Next, for each instance in the data, it calculates the distance between the irrelevant feature and the relevant one. The gathered information is then sorted from least to greatest size according to the distances, and the labels of the first k entries are taken. The proposed model framework is shown in Figure 3.

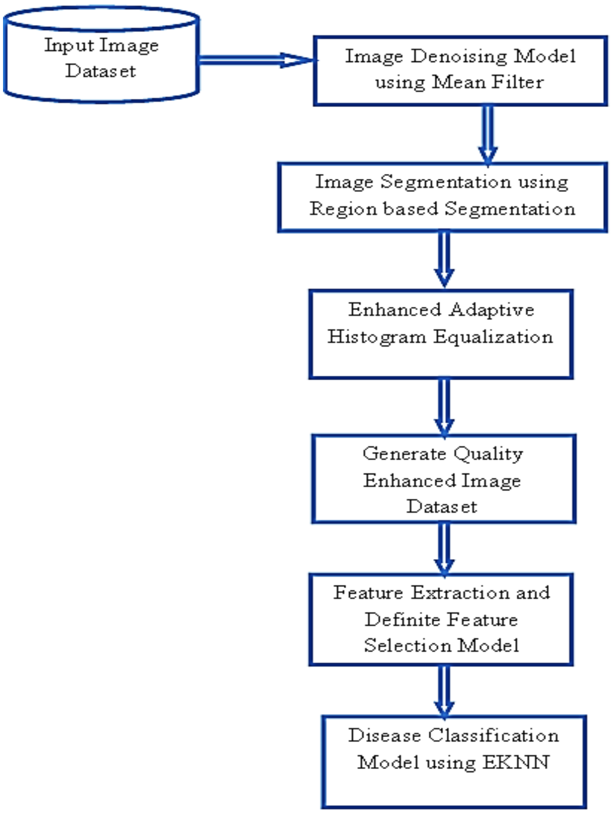

Fig 3: Proposed Model Framework

Image denoising technologies, formerly a useful add-on to computer-aided analysis, have become essential with the proliferation of leaf images under less-than-ideal settings. The technique of removing noise from an image in order to read it is a pressing issue nowadays. Noise in an image can be eliminated by the denoising process. Denoising techniques are employed in leaf images to get rid of noise. The denoising of images was used to remove noise from sugarcane leaves and identify disease plants. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization (RbS-EAHE) based Image Quality Enhancement model is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for disease detection. The features are extracted from the enhanced images and the feature selection is performed that generates the definite feature set using correlation model. The EKNN model is applied by allocating weights to the definite feature set that classifies the leaves as disease and non disease leaves. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization based Image Quality Enhancement model with Definite Feature Set model using EKNN (RbS-EAHE-EKNN) for Leaf Disease Classification is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for accurate disease or non disease classification. The proposed model initially performs image denoising to remove the noise completely and segmentation using region based segmentation technique is applied on the denoised image set. The proposed model considers the extracted features and performs 2 level classification as disease or non disease. Class_1 represents the diseased leaves and Class_2 represents the Non diseases l eaves.

Input: Sugarcane Leaves Images {SLIset}

Output: Classification Set {Clset}

3.1 Denoising and Image Segmentation

Initially on the image dataset, denoising is applied to remove the noise completely. The noise in the images impact the disease prediction accuracy. To accurately detect the diseases, noise in the image is completely removed by applying the filtering techniques. An image can be broken down into its constituent pixels with the aid of a mask or a labeled image. If a picture is segmented beforehand, only the necessary sections of the image will need to be processed, which greatly improves processing speed. Each image from the sugarcane leaf image dataset is carefully chosen and processed for feature extraction. The image loading in the proposed model performs scaling that considers image of any size to a fixed size for processing. The original image and scaled image after performing image loading is shown in Figure 4.

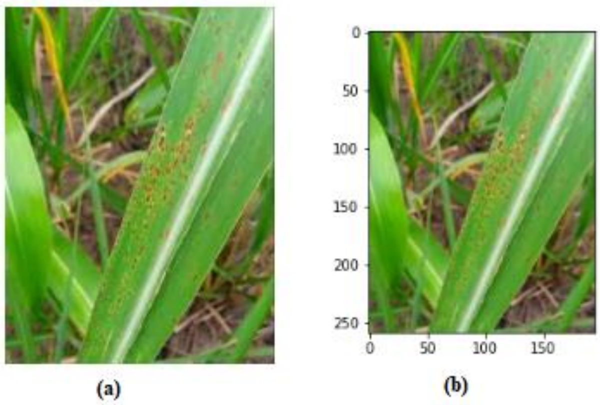

Fig 4: (a) Original Input Image (b) Image Scaling to fit the Window

The image loading and pixel analysis is performed as

$$LImg(SLIset[M]) = \sum_{L=1}^{M} getIMGattr(L) + size(L) + maxIntensity(L) \qquad (4)$$

Here getIMGattr model extracts a image from the dataset and the size is calculated and the maximum intensity is considered that considers the image with minimum intensity of a range.

In addition to aiding in the detection of leaf disease, image denoising is useful for elucidating anatomical features normally obscured in the sugarcane leaf. Using a standard database of leaf images, it can spot any deviations. When an image is noisy, denoising can help by eliminating the noise and revealing the original picture for performing segmentation. Denoising may cause some loss of detail in the original photos due to the fact that noise, edge, and texturing are all high frequency components. After denoising, image segmentation divides a whole picture into smaller, more manageable chunks, generally according to the properties of individual pixels.

The original image and denoised image after performing the denoising operation is shown in Figure 5 and the denoised image and the segmented images is shown in Figure 6.

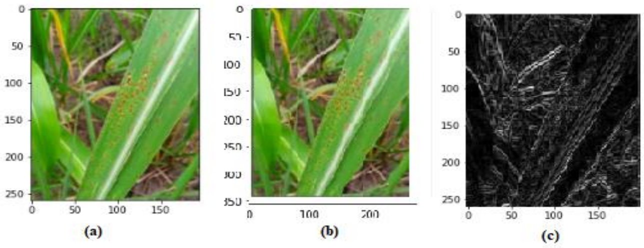

Fig 5: (a) Original Scaled Image (b) Denoised Image (c) Segmented Image

The denoising and image segmentation is performed as

$$Denoise(LImg[M]) = \prod_{L=1}^{M} mean\big(LImg(X,Y)\big) * \sum_{L=1}^{M} Stdcoeff\big(L(X,Y)\big) - \delta(LImg(L) \quad (5)$$

$$LSeg(LImg[M]) = \lim_{L \to SLIset}\left(maxInten\left(\sum_{i=1}^{M} LImg(i) + \frac{minInten(LImg(i))}{size(LImg)}\right) + \sqrt{\sum_{L=1}^{M} \frac{Pixattr(LImg(i)) - \delta(X,Y)}{(X+1,Y+1)}}\right. \quad (6)$$

Here X and Y are the pixel coordinates of the leaf image and the δ is the model that considers the poor intensity pixels that can be removed from the set for classification of leaf diseases.

3.2 Enhanced Adaptive Histogram Equalization

Enhanced adaptive histogram equalization is a method of digital image processing for increasing contrast. Unlike traditional histogram equalization, the adaptive technique boosts contrast in certain areas. It does a histogram equalization calculation for each of the image's blocks. For this reason, EAHE generates a large number of histograms, each of which represents a different region of the image. It improves local contrast and edge definition across the entire image. The HE of individual picture regions is calculated. There is no loss of edge definition between visually distinct areas of the image. It provides a more noticeable contrast in that specific area.

Image contrast can be improved by enhanced adaptive histogram equalization, a method of digital picture processing. Contrast is increased in specific areas, setting the adaptive technique apart from traditional histogram equalization. With the use of image processing techniques like shifting color spaces, reversing images, dehazing, upping the saturation, a local contrast preserving methodology based on adaptive histogram equalization is developed. The suggested technique is meant to save local image features while enhancing contrast with EAHE. Specifically, the peak signal-to-noise ratio and the normalized absolute error are the image brightness metrics computed to reflect the efficacy of this technique. The EAHE model applied on the image enhances the smoothness and quality of the image as shown in Figure 6 and the gray level image is shown in Figure 7.

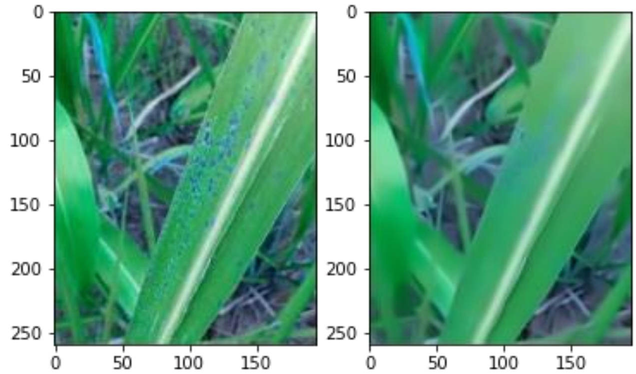

Fig 6: (a) Original Image (b) EAHE resultant Image

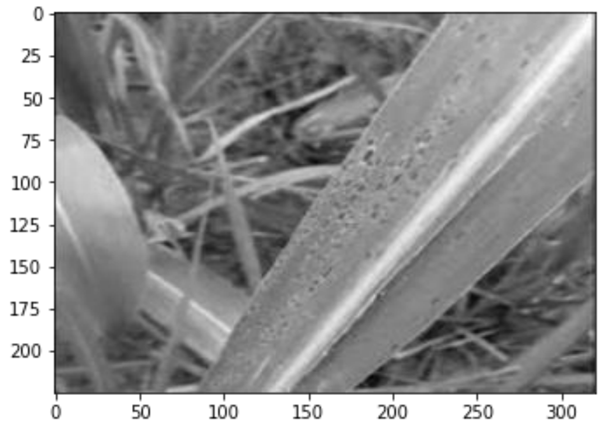

Fig 7: Grey Level Image

A grey level Image {GI} is considered from the set {SLIset} that calculates the number of grey occurrences with m probabilities as

$$Prob\big(GI(L)\big) = \sum_{L=1}^{M} \frac{\max(SLIset(L))}{M} \quad (7)$$

$$HE(Prob(L)) = \prod_{L=1}^{M} \frac{\sum_{L=1}^{M} Lseg_L}{maxrange(L(X,Y))} + \tau \quad (8)$$

Here τ is the adaptive threshold value considered for image quality enhancement and the logarithmic mapping is performed for accurate edge based extraction as

$$EAHE(HE(L)) = \sum_{L=1}^{M} L\binom{X}{Y} * \log\left(X^{L} + \beta\right) * \sum_{i=1} \delta^{2} + Th \quad (9)$$

Here β is the pixels with maximum grey level and with maximum intensity and Th is the threshold intensity used for image quality enhancement.

3.3 Feature Extraction and Classification

The process of reducing loosely organized leaf data to a collection of quantifiable qualities that may be handled without sacrificing any of the initial data's context is known as feature extraction. The outcomes are far better than when machine learning is applied to raw data. Feature extraction helps to clean up a dataset by removing any extraneous data. Reduced data sets make it easier to develop models with less computing overhead, and they speed up the learning and generalization stages of machine learning. The process of selecting relevant aspects of a dataset for further processing and analysis is known as feature selection. The extracted features are radius, mean, texture, smoothness, entropy, perimeter, color, shape, contrast and similarity.

The feature extraction and selection of the most relevant features are performed as

$$FeatExtr(Lseg(L)) = \frac{maxInten(L(X,Y))}{\delta} + \left(\max(EAHE(X+1, Y+1)) - \frac{Min(EAHE(X,Y))}{\beta}\right)^{2} + [R - \delta] \quad (10)$$

$$Fset[M] = \frac{\sum_{L=1}^{M} \max(FeatExtr(L)) - \min(FeatExtr(L)}{corr(FeatExtr(L,L+1))} * \delta + \sqrt{\frac{maxattr(L+1)}{minattr(L)}} \quad (11)$$

Here R is the features extracted from the segmentation using the EAHE model that fully considers all the features with 8 dimensions like radius, mean, smoothness, perimeter, grey level, intensity level, entropy, dissimilarity.

Every instance must include the incorrect answer in order for the k-NN to be used to construct one of the most successful attribution systems. Inputs to the EKNN are the most common ones. Following data extraction, the nearest neighbors for each of the numerical values are taken into account when deciding on the nominal values. Ignoring the instances which include them is the simplest approach when working with data affected by a large technique, but mining findings show that this can be not always achievable.

3.4 Enhanced KNN Algorithm

It is vital to remember that these characteristics are interconnected and don't work separately, which leads to inaccurate forecasts even when estimations are used for calculating the wrong feature values. The proposed EKNN model process is shown in Figure 8.

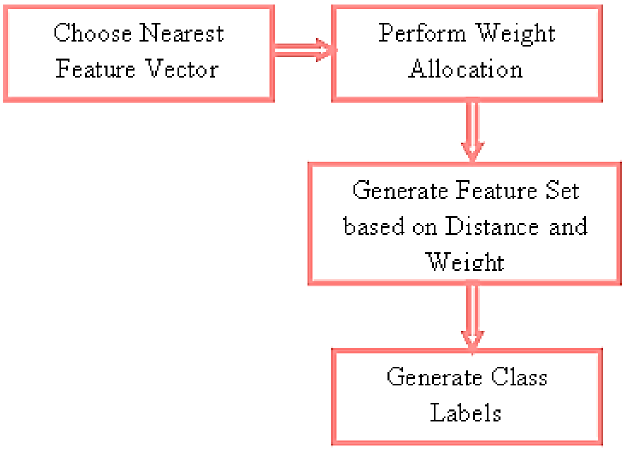

Fig 8: EKNN Workflow

To determine how dissimilar or similar a given instance is to a given training instance, EKNN employs the use of the Euclidean distance. All instance characteristics, relevant or not, are considered with the same weight. The EKNN model allocates varying weights to the features when determining how far apart two instances are. The EKNN process is performed as

Choose a nearest neighbor feature vector as

$$Fset[L] = \sum_{f=1}^{H} \frac{min(dist(f+1,f)}{size(Fset)} \quad (12)$$

Based on the nearest neighbor set considered in the Fset, the weight allocation is performed to the features that are nearer in range for training the model. The weight allocation is performed as

$$Wei(Fset[L]) = \sum_{f=1}^{H} \frac{maxVal(Fset(f)) - minVal(Fset(f+1)}{Ssize(Fset)+\beta} \quad (13)$$

From the weighted set calculated, perform weight wise sequencing of features that are relevant and in nearest distance. The weight based and distance based feature set is generated as

$$FeatVec[L] = \sum_{f=1}^{H} \frac{min(dist(f+1,f)}{size(Fset)} + \sum_{f=1}^{H} \frac{maxVal(Fset(f)) + mindist(f)}{K} \quad (14)$$

Here K is the feature vector size that is updated with the features newly considered based on low correlation and in minimum distance.

Use the next weighted center as the position in the interval represented by a threshold number K that is performed as

$$featVec(f < K < H) \qquad (15)$$

The feature vector is used to train the model that performs classification of leaf disease as diseased or normal, that is performed as

$$Class[L] = \sum_{L=1}^{M} \frac{simm(FeatVec(f, f+L)) + \delta(X,Y)}{size(FeatVec)} \begin{cases} if\ simm > Tr\ then & 1 \\ otherwise & 0 \end{cases} \qquad (16)$$

The proposed classification model error rate is measured as

$$ErrorRate[L] = \frac{1}{X_i * Y_i} * \frac{\sum_{L=1}^{M} Class(X+1, Y+1) - Class(X, Y))^2}{\sum_{L=1}^{M} size(FeatVec)} \qquad (17)$$

4. Results

For 58% of rural households in India agriculture is the main source of income, helping to propel the country's economy to the ranks of the world's fastest-growing markets. The thriving Indian processing industry bodes well for the future of Indian food exports. Because of this, India's grocery market is now the sixth largest in the world, and the retail food sector in the country has seen a 70% increase in sales. Recently, technological advancements in the food processing industry have propelled it to the fifth spot in terms of output, intake, exports, and anticipated increase. The country's food service industry accounts for 32% of the total market. The country's soil, climate, and farming techniques are held in high regard for their involvement in the successful cultivation of numerous food crops. Predicting and estimating leaf diseases early on is crucial for managing plant diseases and minimizing their severity. The purpose of this research is to enhance the leaf image quality and performing segmentation for accurate feature extraction for precise detection of leaf diseases.

The proposed model is implemented in python and executed in Google Colab. The dataset is considered from the link https://www.kaggle.com/datasets/alihussainkhan24/red-rot-sugarcane-disease-leaf-dataset. The dataset contains more than 900 images and using rotation and augmentation technique 2,940 pictures are generated, each of which is a sugarcane leaf that falls into one of six different types. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization based Image Quality Enhancement model with Definite Feature Set model using EKNN (RbS-EAHE-EKNN) is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for accurate disease or non disease classification. The proposed model is compared with the traditional Logistic Regression Model (LRM) model. The proposed model when contrasted with existing model, proposed model exhibits better performance.

A directional scaling technique with adaptive sharpening is utilized in Leaf Image Scaling to simultaneously detect image flaws, upscale the image, and sharpen the image so that it appears to be operating at a greater resolution. An image composed entirely of graphic primitives, such as in a vector graphic, can be scaled using geometric change without any noticeable degradation in image quality. Raster graphics require the creation of a new image with a different number of pixels when they are scaled. The Figure 9 represents the scaling of leaf image for processing and feature extraction.

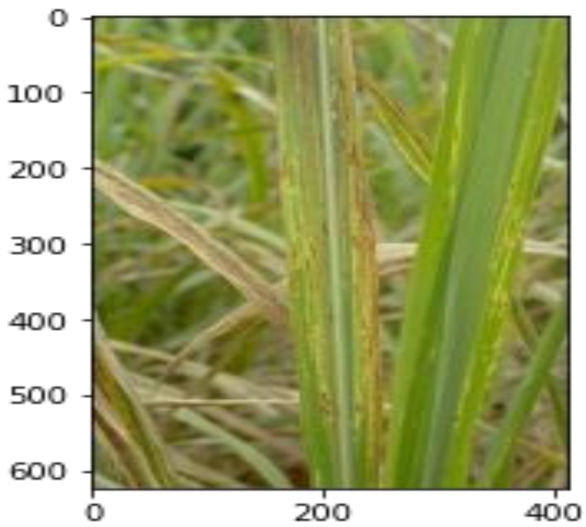

Fig 9: Image Scaling

Grey scale transformation is a method of converting images into a form which is used for extracting the image features accurately. As a result, only grayscales are left, with white being the lightest and black the darkest. This process is used for accurate extraction of features for disease classification. The grey scale image is shown in Figure 10.

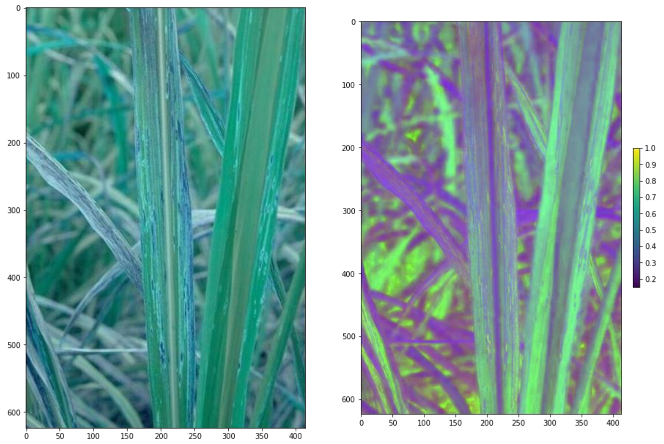

Fig 10: Grey Scale Leaf Image

Image segmentation divides a whole leaf image into smaller, more manageable chunks, generally according to the properties of individual pixels. Segmenting an image involves grouping together similar pixels into bigger groups, which removes the need to analyze individual pixels. Creating an image by cutting it up into smaller, more manageable pieces is called tiling. Segmentation of an image begins with the identification of non-splittable regions within the image. The Image Segmentation Accuracy Levels of the proposed and traditional models are shown in Figure 11.

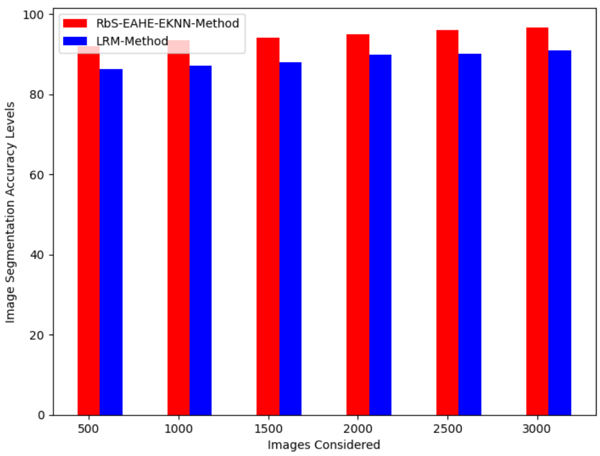

Fig 11: Image Segmentation Accuracy Levels

Image contrast can be enhanced with the help of EAHE, a pre-processing technique. The image's brightness levels are rebalanced by computing multiple histograms, each of which corresponds to a different region of the image. Contrast is increased in specific areas, setting the adaptive technique apart from traditional histogram equalization. The Enhanced Adaptive Histogram Equalization based Image Quality Enhancement Accuracy Levels of the existing and proposed models are shown in Figure 12.

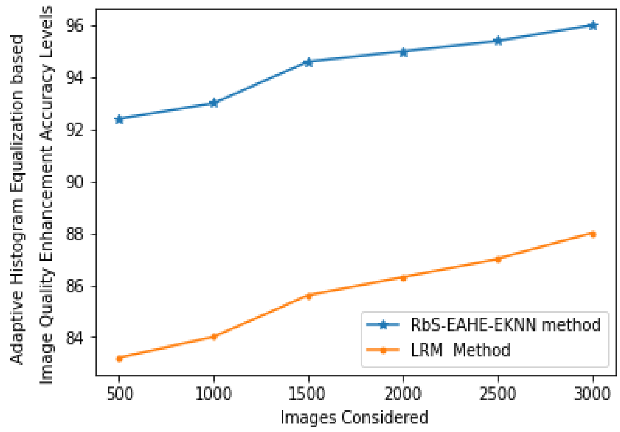

Fig 12: Enhanced Adaptive Histogram Equalization based Image Quality Enhancement Accuracy Levels

The segmented image with highlighting the useful portion that is used to extract the features with elimination of noise is performed. The segmented image is shown in Figure 13.

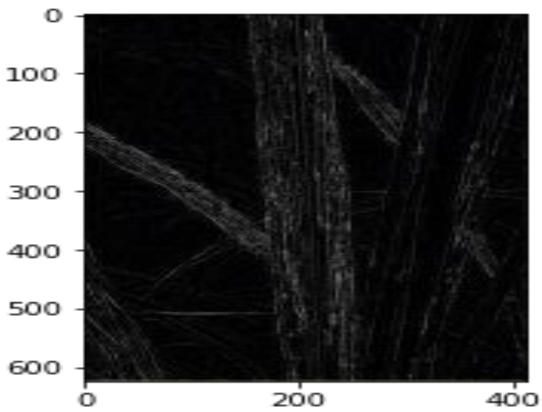

Fig 13: Segmented Image

The process of reducing unorganized information to a set of quantifiable qualities that can be handled without losing any aspect of the original data's significance is referred to as feature extraction. When applied to cleaned data, the EKNN automated learning model excels. The first test feature keeps track of the distances between each test instance and its nearest neighbor in the first class. The second test feature is a summation of the distances between each test instance and its two nearest neighbors in the first class. The Figure 14 represents the Feature Extraction Time Levels of the existing and proposed models.

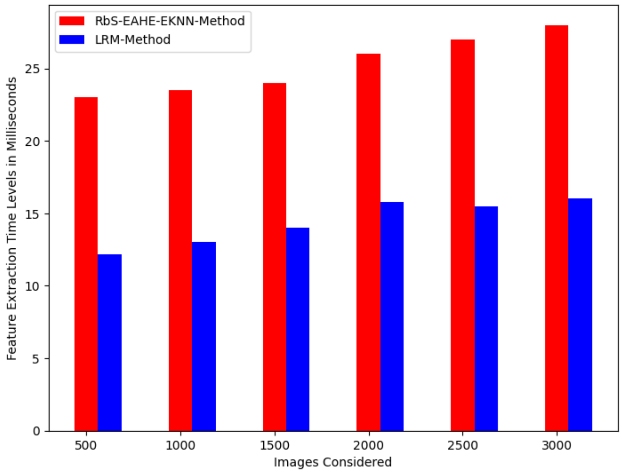

Fig 14: Feature Extraction Time Levels in Milliseconds

The image horizontal and vertical segmentation edges are performed and the leaf disease portion can be identified using the relevant portion. The highlighted portion is shown in Figure 15. The figure 16 represents the extracted feature values.

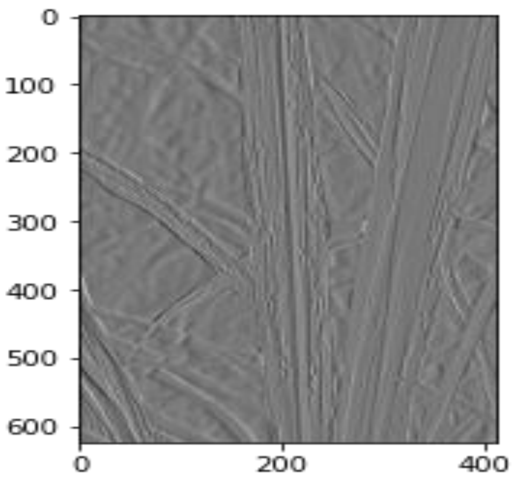

Fig 15: Leaf Highlighted Portion

```
array([[False,  True, False, False, False, False,  True,  True],
       [False, False, False, False, False, False, False, False],
       [False, False, False, False, False, False, False, False],
       [ True,  True,  True,  True,  True,  True,  True,  True],
       [ True,  True,  True,  True,  True,  True,  True,  True],
       [False,  True, False, False, False, False,  True, False],
       [False,  True, False, False, False, False,  True, False],
       [False, False, False, False, False, False, False, False]])
```

Fig 16: Extracted Feature Values

Correlation based feature selection is a filtering technique used to perform classification of leaf diseases based on the relation between the features. A subset of features is assessed only on the basis of their inherent qualities in the data. There are three distinguishing features of a strong correlation. They can reveal information on the nature, strength, and even direction of a link between two variables. The Correlation based Feature Selection Accuracy Levels of the proposed and existing models are shown in Figure 17.

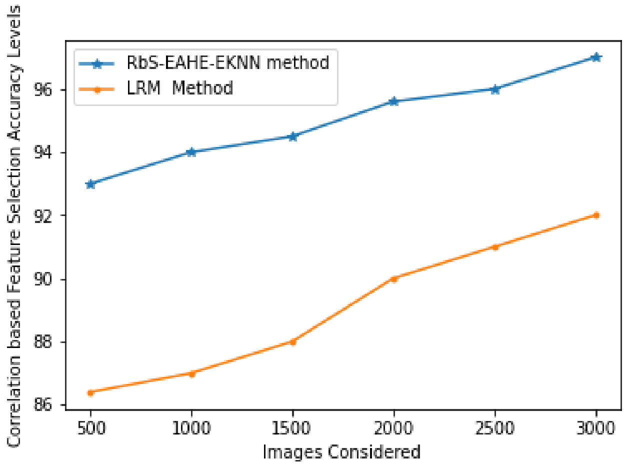

Fig 17: Correlation based Feature Selection Accuracy Levels

The effectiveness of a classification method can be summarized in a table called a confusion matrix. The confusion matrix for the leaf disease classification is represented in Figure 18.

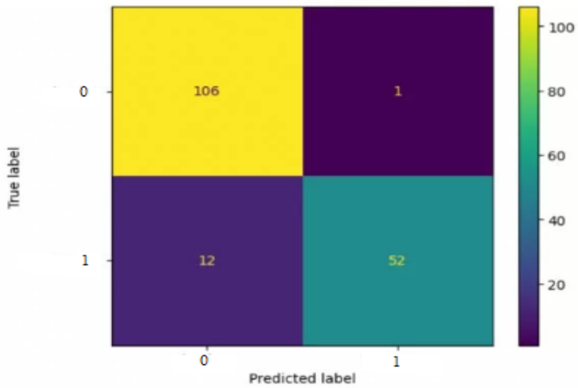

Fig 18: Confusion Matrix

The class contribution algorithm-based enhanced KNN is a more advanced variant of the original KNN. By assigning importance to each feature in the dataset under consideration, the EKNN enhances classification precision. Each feature relative importance is learned via the EKNN built-in learning component. The EKNN approach considers both the training data and the unknown instance when weighing the relative value of several perspectives in determining how to categorize the latter. The EKNN considers not only the k nearest neighbor features, but also the features that has highest weight. The EKNN Classification Time Levels of the proposed and existing models are shown in Figure 19.

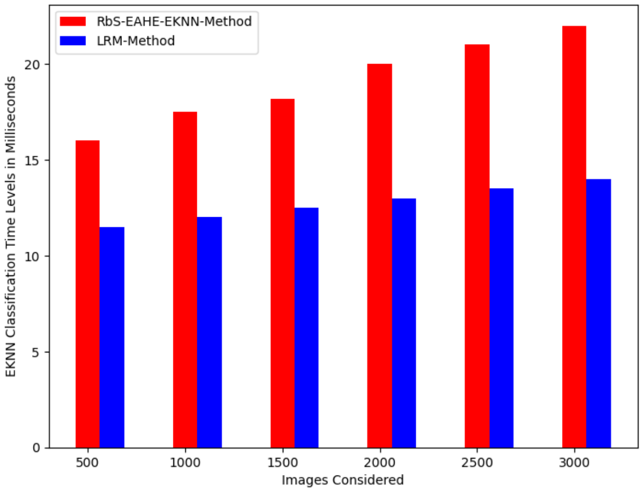

Fig 19: EKNN Classification Time Levels in Milliseconds

The proposed classification parameters are performed as diseased or non diseased. The diseased leaf is represented as class_1 and non diseases is represented as class_0. The evaluation parameters are shown in Table 1 and Figure 20.

Table 1: Evaluation Metrics

Parameters Considered	Models Considered	
	RbS-EAHE-EKNN	**LRM Model**
Accuracy	97.4	91.4
Precision	99	94.2
Recall	94.6	89.5
Sensitivity	93.7	83.4
Specificity	95.3	86.6

```
print(classification_report(y_test, y_predict, target_names=class_name
print(confusion_matrix(y_test, y_predict))

              precision    recall  f1-score   support

     class_0       0.85      0.94      0.89        18
     class 1       0.95      0.75      0.84        24

    accuracy                           0.87        54
   macro avg       0.87      0.90      0.87        54
weighted avg       0.88      0.87      0.87        54
```

Fig 20: Evaluation Parameters

Accuracy is measured by how frequently one's forecasts come true. Differences from a perfect condition are quantified by loss values. This is the sum of the error rates for each occurrence in both the training and validation sets. How well or poorly a model performs after each iteration of optimization can be inferred from the size of the loss. The accuracy statistic provides a clear measure of the algorithm's performance. The Figure 21 shows the Training and Validation Accuracy and Loss levels.

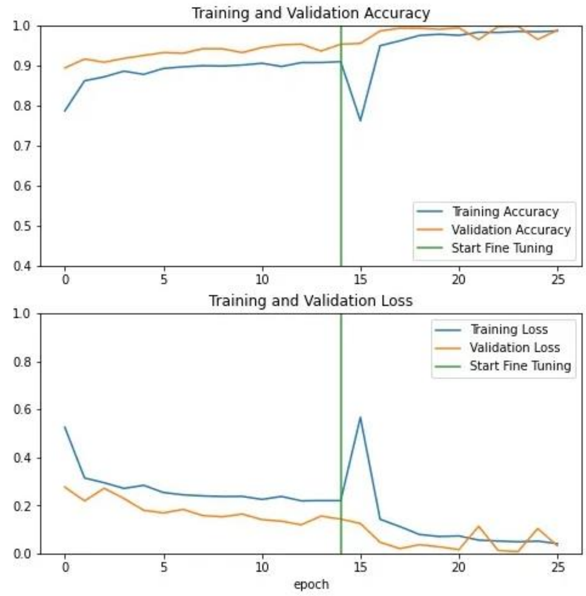

Fig 21: Training and Validation Accuracy and Loss

Predictions based on classifications are the most common and helpful model in leaf disease detection as disease or non disease. Predictions based on a classification sort data into distinct categories. The key is creating training data that includes both the attributes and the expected result. The EKNN Classification Accuracy Levels of the existing and proposed models are shown in Figure 22.

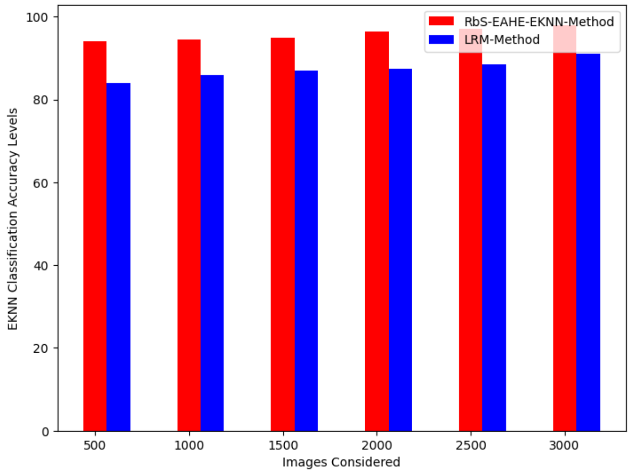

Fig 22: EKNN Classification Accuracy Levels

5. Conclusion

Many countries, particularly India, rely heavily on sugarcane as a food source. Because it enhances flavor and retains moisture in food, sugar plays a major role in the human diet. Despite setbacks, India's sugarcane industry remains vital to the nation's economy.

Numerous diseases are currently a problem for sugarcane crops. Recognizing plant diseases is a vital skill for those working in agriculture. Disease detection is a time-consuming, labor-intensive process that requires constant attention from farmers. Unfortunately, this method can be quite pricey for commercial farms and has a potential of producing inaccurate data. It can be time consuming and expensive for farmers to travel long distances to reach specialists when they need assistance. This research integrates image processing techniques with a practical agricultural application to detect leaf diseases in sugar cane crops. In this research, Region based Segmentation with Enhanced Adaptive Histogram Equalization based Image Quality Enhancement model with Definite Feature Set model using EKNN is proposed for considering the sugarcane images and enhancing the image quality to perform accurate feature extraction for accurate disease or non disease classification. The proposed model enhances the image quality by 97% accuracy in classification. In future, integrated histogram equalization and morphological operations will results in better segmentation and accuracy levels and feature dimensionality reduction can be applied with reduced feature set and integrated classifiers can be used for better accuracy. The time complexity can be reduced further by applying lightweight classification operations with better performance levels.

References

[1] S. Barburiceanu, S. Meza, B. Orza, R. Malutan and R. Terebes, "Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture," in IEEE Access, vol. 9, pp. 160085-160103, 2021, doi: 10.1109/ACCESS.2021.3131002.

[2] H. Phan, A. Ahmad and D. Saraswat, "Identification of Foliar Disease Regions on Corn Leaves Using SLIC Segmentation and Deep Learning Under Uniform Background and Field Conditions," in IEEE Access, vol. 10, pp. 111985-111995, 2022, doi: 10.1109/ACCESS.2022.3215497.

[3] Dodla. Likhith Reddy, & Dr. D Prathyusha Reddi. (2017). Texture Image Segmentation Based on threshold Techniques. *International Journal of Computer Engineering in Research Trends*, *4*(3), 69–75.

[4] S. Mousavi and G. Farahani, "A Novel Enhanced VGG16 Model to Tackle Grapevine Leaves Diseases With Automatic Method," in IEEE Access, vol. 10, pp. 111564-111578, 2022, doi: 10.1109/ACCESS.2022.3215639.

[5] C. Zhou, S. Zhou, J. Xing and J. Song, "Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network," in IEEE Access, vol. 9, pp. 28822-28831, 2021, doi: 10.1109/ACCESS.2021.3058947.

[6] Y. Wu, X. Feng and G. Chen, "Plant Leaf Diseases Fine-Grained Categorization Using Convolutional Neural Networks," in IEEE Access, vol. 10, pp. 41087-41096, 2022, doi: 10.1109/ACCESS.2022.3167513.

[7] N. H. Saad, N. A. M. Isa and H. M. Saleh, "Nonlinear Exposure Intensity Based Modification Histogram Equalization for Non-Uniform Illumination Image Enhancement," in IEEE Access, vol. 9, pp. 93033-93061, 2021, doi: 10.1109/ACCESS.2021.3092643.

[8] Anamika Sharma, & Parul Malhotra. (2017). LDA Based Tea Leaf Classification on the Basis of Shape, Color and Texture. *International Journal of Computer Engineering in Research Trends*, *4*(12), 543–546.

[9] S. H. Majeed and N. A. M. Isa, "Adaptive Entropy Index Histogram Equalization for Poor Contrast Images," in IEEE Access, vol. 9, pp. 6402-6437, 2021, doi: 10.1109/ACCESS.2020.3048148.

[10] Venkata Srinivasu Veesam, & Bandaru Satish Babu. (2017). A Relative Study on the Segmentation Techniques of Image Processing. *International Journal of Computer Engineering in Research Trends*, *4*(5), 155–160.

[11] R. -Q. He, W. -S. Lan and F. Liu, "MRWM: A Multiple Residual Wasserstein Driven Model for Image Denoising," in IEEE Access, vol. 10, pp. 127397-127411, 2022, doi: 10.1109/ACCESS.2022.3226331.

[12] Q. Wu and S. Zhu, "Multispectral Image Matching Method Based on Histogram of Maximum Gradient and Edge Orientation," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 5001105, doi: 10.1109/LGRS.2021.3077688.

[13] Q. Chang, X. Li and Y. Zhao, "Reversible Data Hiding for Color Images Based on Adaptive Three-Dimensional Histogram Modification," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 5725-5735, Sept. 2022, doi: 10.1109/TCSVT.2022.3153796.

[14] Namita M. Butale, & Dattatraya.V.Kodavade. (2018). Survey Paper on Detection of Unhealthy Region of Plant Leaves Using Image Processing and Soft Computing Techniques. *International*

Journal of Computer Engineering in Research Trends, *5*(12), 232–235.

[15] P. Taherei Ghazvinei, H. Hassanpour Darvishi, A. Mosavi et al., “Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network,” Engineering Applications of Computational Fluid Mechanics, vol. 12, no. 1, pp. 738–749, 2018.

[16] M. T. N. Ratchaseema, L Kladsuwan, L Soulard et al., “The role of salicylic acid and benzothiadiazole in decreasing phytoplasma titer of sugarcane white leaf disease,” Scientific Reports, vol. 11, no. 1, pp. 15211–15219, 2021.

[17] A.A. Elsharif and S.S. Abu-Naser, “An expert system for diagnosing sugarcane diseases,” International Journal of Applied Engineering Research, vol. 3, no. 3, pp. 19–27, 2019.

[18] K. Bagyalakshmi, R. Viswanathan, and V. Ravichandran, “Impact of the viruses associated with mosaic and yellow leaf disease on varietal degeneration in sugarcane,” Phytoparasitica, vol. 47, no. 4, pp. 591–604, 2019.

[19] L. Li, S. Zhang, and B. Wang, “Plant disease detection and classification by deep learning-A review,” IEEE Access, vol. 9, pp. 56683–56698, 2021.

[20] K. Garg, S. Bhugra, and B. Lall, “Automatic quantification of plant disease from field image data using deep learning,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1965–1972, Waikoloa, HI, USA, January 2021.

[21] A.Sagar and D. Jacob, “On using transfer learning for plant disease detection,” BioRxiv, pp. 1–8, 2021.

[22] V. Tiwari, R. C. Joshi, and M. K. Dutta, “Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images,” Ecological Informatics, vol. 63, Article ID 101289, 2021.

[23] S.V. Militante and B.D. Gerardo, “Detecting sugarcane diseases through adaptive deep learning models of convolutional neural network,” in Procee2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), pp. 1–5, IEEE, Kuala Lumpur, Malaysia, December 2019.

[24] D.A.G. V. Padilla Magwili, A. L. A. Marohom, and G. Clyde Mozes, “Portable yellow spot disease identifier on sugarcane leaf via image processing using support vector machine,” in Procee2019 5th International Conference on Control, Automation and Robotics (ICCAR), pp. 901–905, IEEE, Beijing, China, August 2019.

[25] N. K. Hemalatha, R. N. Brunda, G. S. Prakruthi, B. V. B. Prabhu, A. Shukla, and O. S. J. Narasipura, “Sugarcane leaf disease detection through deep learning,” Deep Learning for Sustainable Agriculture, Academic Press, Cambridge, MA, USA, 2022.

[26] M. M. Ozguven and K. Adem, “Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms,” Physica A: Statistical Mechanics and its Applications, vol. 535, Article ID 122537, 2019.

[27] K. Thilagavathi, K. Kavitha, R.D. Praba, S.A.J. Arina, and R.C. Sahana, “Detection of diseases in sugarcane using image processing techniques,” Bioscience Biotechnology Research Communications, vol. 15, no. 10, pp. 2157–2168, 2020.

[28] N.B. Quoc, “Development of loop mediated isothermal amplification assays for the detection of sugarcane white leaf disease,” Physiological and Molecular Plant Pathology, vol. 113, Article ID 101595, 2021.

[29] Y. Shendryk, J. Sofonia, R. Garrard, Y. Rist, D. Skocaj, and P. Thorburn, “Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging,” International Journal of Applied Earth Observation and Geoinformation, vol. 92, Article ID 102177, 2020.

[30] K.-l. Wang, Q.-q. Deng, J.-w. Chen, and W.-k. Shen, “Development of a reverse transcription loop-mediated isothermal amplification assay for rapid and visual detection of Sugarcane streak mosaic virus in sugarcane,” Crop Protection, vol. 119, pp. 38–45, 2019.

[31] Verma, D. ., Reddy, A. ., & Thota, D. S. . (2021). Fungal and Bacteria Disease Detection Using Feature Extraction with Classification Based on Deep Learning Architectures. Research Journal of Computer Systems and Engineering, 2(2), 27:32. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/29

[32] Mirdha, V. ., Bhatnagar, D. ., Saleem, S. ., Sharma, B. ., & Jangid, K. G. . (2023). Circularly Polarized Antenna with Metallic Reflector for High-Gain Satellite Communication. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 91–97. https://doi.org/10.17762/ijritcc.v11i4.6391

[33] Pandey, J. K., Veeraiah, V., Talukdar, S. B., Talukdar, V., Rathod, V. M., & Dhabliya, D.(2023). Smart city approaches using machine learning and the IoT. Handbook of research ondata-driven mathematical modeling in smart cities (pp. 345-362) doi:10.4018/978-1-6684-6408-3.ch018 Retrieved from www.scopus.com