# Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease

Godliver Owomugisha, John A. Quinn, Ernest Mwebaze
Department of Computer Science
Makerere University
Email: [owomugisha.godliver, jquinn, emwebaze]@cis.mak.ac.ug

James Lwasa
Department of Information Systems
Makerere University
Email: lwasaj@yahoo.com

***Abstract*—Machine learning has been applied in agriculture in various areas including crop disease detection and image processing systems have been developed for some crops. These crops include cotton, pomegranate plant, grapes, vegetables, tomatoes, potatoes and cassava among others. However, no machine learning techniques have been used in an attempt to detect diseases in the banana plant such as banana bacterial wilt (BBW) and banana black sigatoka (BBS) that have caused a huge loss to many banana growers. The study investigated various computer vision techniques which led to the development of an algorithm that consists of four main phases. In phase one, images of banana leaves were acquired using a standard digital camera. Phase two involves use of different feature extraction techniques to obtain relevant data to be used in phase three where images are classified as either healthy or diseased. Of the seven classifiers that were used in this study, Extremely Randomized Trees performed best in identifying the diseases achieving 0.96 AUC for BBW and 0.91 for BBS. Lastly, the performance of these classifiers was evaluated based on the area under the curve (AUC) analysis and best method to automatically diagnose these banana diseases was then recommended.**

## I. INTRODUCTION

Banana is the fourth most grown crop in the world after wheat, rice and maize and Uganda happens to be the second largest producer of bananas after India [1]. The crop is used as a staple food source in the country however, its growth is threatened by banana bacterial wilt (BBW) disease caused by *xanthomonas campestris pv musacearum (XCM)* [2]. The wilt originated in Ethiopia and in Uganda it was reported by Tushemereirwe et al. [3] in Kayunga district in 2001. The disease has also moved into Congo, into Rwanda and Tanzania. BBW affects all types of banana and spreads very fast causing a devastating effect hence, many farmers have lost their crops and this has led to reduction of food availability and income for banana farmers. The disease is also coupled with many costs including labor for cutting down and disposing off infected plants, de-budding the male flowers and disinfecting cutting tools.

Following the outbreak of BBW epidemics, the government of the Republic of Uganda through the Ministry of Agriculture, Animal industry and Fisheries (MAAIF) in conjunction with National Agricultural Research Organization and other key stake holders constituted a national task force in December 2001, which in November 2003 formulated long-term strategy and action plan to eradicate the disease. This strategy includes a national coordinated effort of continuous monitoring of the epidemics: awareness raising and training campaigns, empowering all stakeholders at all (district, sub county, parish and village) levels to control the disease [4].

The problem of identifying diseases in plants is a very well known one. Farmers wait for that time when the disease gets to a late stage and the symptoms are visible to realize that the crops are diseased. However, not much can be done to control the situation by that stage, hence this study aimed at early disease detection. The symptoms are visible in the leaves, male bud, fruit and stem. The disease begins with any leaf and causes them to turn yellow, brown and later they wilt.Young affected plants become stunted and may not produce any fruits. Apart from the BBW disease, Tushemereirwe et al. [3] mention other diseases that have led to the decline in banana production in many banana growing countries in the world including Uganda and these include: banana strake virus disease and banana black sigatoka (BBS). BBS blackens parts of the leaf and normally, drying starts from the edges and eventually the entire leaf is killed.

This paper is divided into five sections: starting with the introduction. Section two presents research that has been done on crop disease detection. Section three describes how different feature extraction and classification methods were applied to achieve the objectives of the study. Results of the techniques used are evaluated in section four and the last section recommends methods that worked best and future work.

## II. RELATED WORK IN COMPUTER VISION FOR AGRICULTURAL DISEASE DETECTION

Computer vision systems have been used increasingly in the food and agricultural areas for quality inspection [5], [6] and evaluation purposes as they provide suitably rapid, economic, consistent and objective assessment. They have proved to be successful for the objective measurement and assessment of several agricultural products [7]. With the advantages of superior speed and accuracy, a significant number of researchers have been attracted to apply machine vision techniques in crop disease detection.

A support vector machine technique has been used for classification and identification of foliar diseases in cotton [8]. The classification process starts by finding the best feature vector for each class and then creates the final classification system from the best results obtained. To accomplish this, the following were considered: decomposition of images into

 ISBN: 978-0-7972-1533-7

multiple channels (R, G, B, H, S, V, I3a, I3b, and GL), application of the discrete wavelet transform up to the third level, computation of the energy for each sub-band and feature vectors. This is followed by creation of the SVM classification environment, listing of the images used for training and testing and evaluation of the best feature vectors.

Al-Hiary et al. [9] proposed an automatic detection and classification of leaf diseases and the work is divided into three parts. This begins with the identification of the infected object(s) based upon K-means clustering procedure, extraction of the feature set of the infected objects using color co-occurrence methodology for texture analysis and finally detection and classification of disease type using artificial neural network (ANNs).

Aduwo et al. [10] present an automated vision-based diagnosis of cassava mosaic disease. The proposed algorithm is based on camera-phone input to provide a more efficient solution. The methodology begins with capturing leaf images with a standard digital camera. The captured image is then processed by applying various image processing techniques such as SIFT, SURF and HSV for shape feature extraction. The image is either classified as diseased or not based on other methods like a k-nearest neighbor classifier (KNN), support vector classifier (SVC) and Naive Bayes among others. A comparison on the different classifiers was done and results for the three main datasets were produced.

Others [7], [11] have demonstrated the value of image processing in inspecting and grading the quality of agricultural and food products. An automated system for the disease detection and grading in pomegranate plant was proposed in [11]. The techniques used here include color segmentation based on linear discriminant analysis, contour curvature analysis and a thinning process, which involves iterating until the stem becomes a skeleton.

The approach in [9] uses color co-occurrence methodology for texture analysis which makes it not applicable for banana leaves. However, the developed algorithm combined the features extracted in [8], [10] and this added strength to the results. In addition to the classifiers that were used, the study investigated on the behavior of other classification techniques on the dataset and recommended the best methods. This has not been done in the past for banana diseases thus making this research new.

## III. Methods and Results

The methodology aimed at detecting the BBW and BBS diseases using automated vision-based diagnosis techniques and work was divided into 4 parts: image acquisition, feature extraction, disease classification and evaluation of the classification performance.

### A. Image acquisition

A Canon digital camera of 12 megapixels was used to capture both healthy and diseased images from different banana plantations in Bushenyi district (Western part of Uganda) where these diseases are common. Samples were taken from 5 sub-counties at an average of 5 diseased plants per plantation. A total of 623 image samples was used for this study and data was organized in three sets. Set one holds 360 leaves from healthy plants, set two has 220 leaves diseased with BBW and set three has 43 leaves diseased with BBS. In order to capture clear images with descriptive details, the camera was kept in both horizontal and vertical resolution of 72dpi (dots per inch). The flash mode was off since images were taken during day time with enough natural light and the process did not involve any cutting/removal of leaves off the plant. One sample image from each set is given in the figures below.

Fig. 1. Healthy leaf

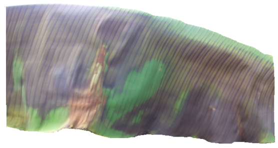

Fig. 2. Leaf affected by BBW disease

Fig. 3. Leaf affected by BBS disease

### B. Feature extraction and creation of feature vectors

Most of the time, the captured images may contain many objects especially in the background and working with such images leads to inaccurate/incorrect results. These images were cropped in order to obtain the leaf part only. However, cropped images then had a white background with pixel values of 255 and working with the whole image also brings inappropriate results too. To avoid this challenge a mask was applied onto the image in order to obtain the useful segment. The region with most green pixels was identified and basing on threshold value of gray $< 200$; green components of the pixel intensities are set to one and the background is set to zero. This converts an image into binary, thus indicating the segmentation of the leaf from the background. This mask was then applied onto the original image during histogram calculation as follows: the pixels with zero components were deleted (by multiplying the mask pixel values with the pixels of the original image) and only the region where the pixels are ones was considered during histogram calculation. Color histograms were extracted and transformation was from RGB to HSV, RGB to L*a*b*. Fig.4 is a mask of Fig.1

Shape was also considered for this study and the process of calculating shape features was based on three routines namely, thresholding at different levels, extracting of connected components, calculating morphological features for each connected component. First, each image is thresholded at

 ISBN: 978-0-7972-1533-7

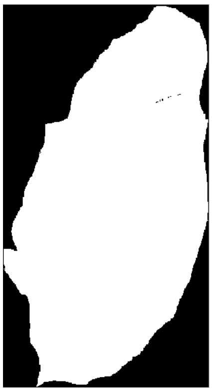

Fig. 4. Mask

gray level. Connectivity openings [12] were used to calculate all the components in each thresholded image. These are called the peak components and were used to construct a max tree which is a data structure designed for morphological image processing in order to efficiently compute features or attributes of the connected components (following the same methodology as [13]). This process was done for every image and various morphological features were calculated for the connected components. Five shape attributes were therefore chosen to be more important and these include: Area of minimum enclosing rectangle, elongation, small compactness, small perimeter and Moment of Inertia.

The minimum bounding rectangle also called minimum bounding box is the smallest rectangle that contains every point in the shape. For an arbitrary shape, eccentricity is the ratio of the length $L$ and width $W$ of minimal bounding rectangle of the shape at some set of orientations. Elongation, *Elo*, is based on eccentricity [14].

$$Elo = 1 - \frac{W}{L}$$

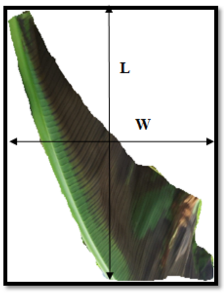


Fig. 5. Minimum bounding rectangle and corresponding parameters for elongation

The compactness measure of a shape is a numerical quantity representing the degree to which a shape is compact and one of the compact measures of shape is surface area / volume. Perimeter of an object is the distance around the outside of the object. Unlike regular shapes where at least two sides or angles are the same, irregular objects do not have these instances of symmetry and perimeter can be determined if one takes into consideration each edge of the shape. This can either be from the left or right, bottom or top. Moment of Inertia is area (mass) times the square of perpendicular distance to the rotation axis, $I = Ad^2$.

To create feature vectors, histogram data for color components H for HSV, R for RGB and L* for L*a*b* was extracted. These components were also combined, for example HS, HV or SV and some classifiers yield better results. Another comparison was done where classification was based on the extracted shape features combined with the color histogram features. To avoid dealing with huge data and overfitting, only 50 bins were used for each case and the histograms were normalized as well.

### C. *Disease classification*

Classifiers map an unlabeled instance of color histogram feature vectors (or a combination of color histogram feature vectors with shape features vectors) to a label. The seven classifiers used in this study were: Nearest Neighbors [15], Decision tree [16], [17], Random forest [18], [19], Extremely Randomized Trees [20], Naive Bayes [21] and support vector classifier (Linear SVM and RBF SVM) [22], [23], [24], [25]. The method used for splitting data set into training and testing was the k-fold cross-validation sometimes called rotation estimation method. The dataset was randomly split into mutually exclusive subsets (folds) of equal size of 10 [26]. The implementation platform was python with Opencv and Scikit-learn libraries. Data and source code used in achieving this are available at https://github.com/godliver/source-code-BBW-BBS.git.

## IV. Results

The choice made on which algorithm (classifier) performed best was based on the results of the AUC analysis. A comparison of the true positive rate and false positive rate for the different classifiers was done. If a classifier yields an AUC score of 1.0, then it has predicted perfectly. 0.5 is a random performance and below 0.5 means the classifier is anti-correlated with the target. Different tests were made for various color components with shape features but excellent performance was generated when the color components H and S for HSV were combined with the five shape attributes that were selected. The AUC results for the different classes (BBW, BBS and healthy) are shown in Fig 6, 7 and 8 respectively.

Of the seven classifiers, Extremely Randomized Trees yield a very high score. Both Random Forest and Extremely Randomized Trees algorithms are ensemble methods. Both algorithms are perturb-and-combine techniques specifically designed for trees. This means a diverse set of classifiers is created by introducing randomness in the classifier construction and the prediction of the ensemble is given as the averaged prediction of the individual classifiers [27]. Scikit-learn implementation combines classifiers by averaging their probabilistic prediction, instead of letting each classifier vote for a single class [18]. However, with Extremely Randomized Trees, randomness goes one step further in the way splits are computed. As in random forests, a random subset of candidate features is used, but instead of looking for the most discriminative thresholds, thresholds are drawn at random

 ISBN: 978-0-7972-1533-7

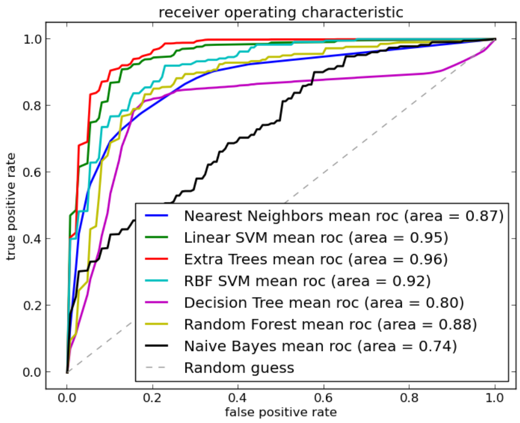


Fig. 6. HS -color components with shape attributes (AUC for BBW)

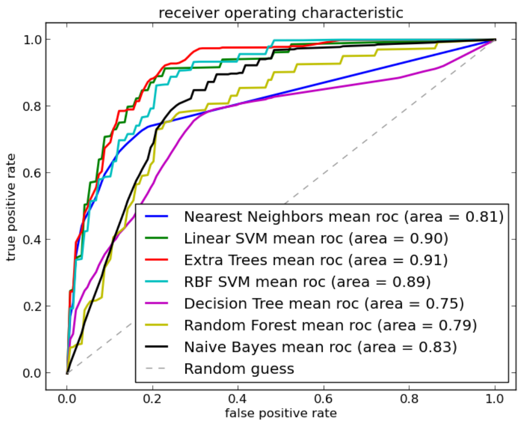


Fig. 7. HS-color components with shape attributes (AUC for BBS)

for each candidate feature and the best of these randomly-generated thresholds is picked as the splitting rule. This usually allows to reduce the variance of the model a bit more, at the expense of a slightly greater increase in bias [20]. Table 1 shows the results of Extremely Randomized Trees classifier dependent on the two leaf features. Whereas color has a greater impact than shape features, AUC performance is better when both features are combined.

## V. Conclusion

With a very high performance of 0.96, 0.91 and 0.99 AUC for BBW, BBS and healthy classes respectively, this research has proved that there is a consistent and more accurate way to auto-detect these banana diseases rather than relying on the previous strategies that have been used in [4]. It has been shown how different feature extraction methods and classification techniques are applied systematically in the attempt to solve this problem. It is evident that the algorithm is feasible and can well identify the two diseases. Features that have been selected that work best for this application are when H and S color components are combined with the five shape features that were chosen as most important. Among the seven

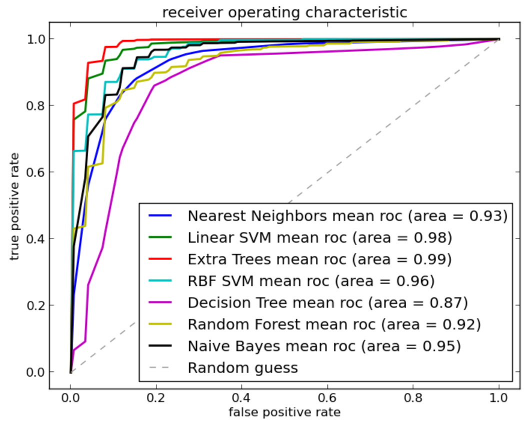


Fig. 8. HS-color components with shape attributes (AUC for healthy)

| | BBW | BBS | Healthy |
|---|---|---|---|
| Color | 0.94 | 0.90 | 0.97 |
| Shape | 0.90 | 0.84 | 0.96 |
| Color + Shape | 0.96 | 0.91 | 0.99 |

TABLE I. AUC for Extremely Randomized Trees (Extra Trees) classifier

classifiers that were used, Extremely Randomized Trees is recommended because of its high performance on this data set.

The platform for automation of vision-based diagnosis of BBW and BBS diseases provides a useful direction and this work can be extended so that this works on a mobile phone device. This adds flexibility to the application since farmers are able to move with their phones to the fields and minimizes the cost of training personnel to monitor banana plants in different regions. The tool could then provide real-time information as farmers don't need to wait for experts as they can always send images to the server and then get advice. There will always be consistency of results since everyone uses the same tool. Two experts might give two different judgements on the same image, but software will always give the same answer. Other improvements that can be brought to the current work include:

- Investigating on the possibility of bananas ever getting infected by both BBW and BBS diseases.
- Adding another class to cater for healthy but mature leaves that are beginning to age or leaves affected by drought stress
- Considering features of the other parts of the plant such as the stem.

## Acknowledgment

The authors gratefully acknowledge the AI-DEV group, department of computer science, Makerere university for the helpful suggestions on improving this work. The authors would also like to say thanks to the team at National Agricultural Research Laboratories for being kind and helpful during the data collection stage.

## REFERENCES


[1] "The biology of bananas and plantains," *Uganda National Council for Science and Technology(UNCST) and Program for Biosafety Systems(PBS)*, 2007.

[2] L. Turyagyenda, G. Blomme, F. Ssekiwoko, E. Karamura, S. Mpiira, and S. Eden-Green, "Rehabilitation of banana farms destroyed by xanthomonas wilt in uganda," *Journal of Applied Biosciences*, vol. 8, no. 1, p. 230 235, 2008.

[3] W. Tushemereirwe, A. Kangire, J. Kubiriba, M. Nakyanzi, and C. Gold, "Diseases threatening banana biodiversity in uganda," *African Crop Science Jounal*, vol. 12, no. 1, pp. 19–26, 2004.

[4] O. O. W.K Tushemereirwe, D Ngambeki, "Awareness of banana bacterial wilt control in uganda: 2. community leaders perspectives," *African Crop Science Journal*, vol. 14, no. 2, p. 166, 2006.

[5] R. H. Asankhani and H. Navid, "Qualitative sorting of potatoes by color analysis in machine vision system," *Journal of Agricultural Science*, vol. 4, no. 4, 2012.

[6] T. Brosnan and D.-W. Sun, "Improving quality inspection of food products by computer visiona review," *Journal of Food Engineering*, 2003.

[7] N. V. G and H. K. S, "Quality inspection and grading of agricultural and food products by computer vision," *International Journal of Computer Applications*, vol. 2, no. 1, May 2010.

[8] A. A. Bernardes, J. G. Rogeri, N. Marranghello, and A. S. Pereira, "Identification of foliar diseases in cotton crop," *Topics in Medical Image Processing and Computational Vision*, vol. 8, pp. pp 67–85, 2013.

[9] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. Al-Rahamneh, "Fast and accurate detection and classification of plant diseases," *International Journal of Computer Applications*, vol. 17, no. 1, pp. 31–38, March 2011.

[10] J. R. Aduwo, E. Mwebaze, and J. A. Quinn, "Automated vision-based diagnosis of cassava mosaic disease," *Proceedings of ICDM Workshop on Data Mining in Agriculture*, 2010.

[11] S. Sanvakki, V. Rajpurohit, V. Nargund, R. Arunkunar, and P. Yallur, "A hybrid intelligent system for automated pomegranate disease detection and grading," *International Journal of Machine Intelligence*, vol. 3, no. 2, pp. 36–44, 2011.

[12] C. Ronse, "Set-theoretical algebraic approaches to connectivity in continuous or digital spaces," *Journal of Mathematical Imaging and Vision*, 1998.

[13] J. A. Quinn, A. Andama, I. Munabi, and F. N. Kiwanuka, "Automated blood smear analysis for mobile malaria diagnosis," *Mobile Point-of-Care Monitors and Diagnostic Device Design, CRC Press*, 2014.

[14] Y. Mingqiang, K. Kidiyo, and R. Joseph, "A survey of shape feature extraction techniques," *Pattern Recognition*, pp. 43–90, 2008.

[15] Z. Ma and K. Ata, "K-nearest-neighbours with a novel similarity measure for intrusion detection," *UKCI'13*, pp. 266–271, 2013.

[16] O. Maimon and L. Rokach, "Data mining and knowledge discovery handbook, second edition," April 2010.

[17] L. Ruey-Hsia and B. Geneva G, "Instability of decision tree classification algorithms," *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 570–575, 2002.

[18] B. S. Leo and L. Breiman, "Random forests," *Machine Learning*, pp. 5–32, 2001.

[19] L. Andy and W. Matthew, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[20] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[21] H. Zhang, "The optimality of naive bayes," *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, 2004.

[22] S. R. Gunn, "Support vector machines for classification and regression," *University of Southampton, Technical Report*, 1998.

[23] S. Keerthi, O. Chapelle, and D. DeCoste, "Building support vector machines with reduced classifier complexity," *Journal of Machine Learning Research*, vol. 7, pp. 1493–1515, 2006.

[24] C. wei Hsu, C. chung Chang, and C. jen Lin, "A practical guide to support vector classification," *National Taiwan University, Taipei 106, Taiwan*, 2010.

[25] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines: And other kernel-based learning methods," *Cambridge University Press*, 2000.

[26] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, pp. 1137–1143, 1995. [Online]. Available: http://dl.acm.org/citation.cfm?id=1643031.1643047

[27] Breiman, "Arcing classifiers," *Annals of Statistics*, 1998.