Pixel Classification Methods for Automatic Symptom Measurement of Cassava Brown Streak Disease

Joviah Tuhaise
Department of Computer Science
Makerere University, Uganda
Email: tuhaise.joviah@cis.mak.ac.ug

John A. Quinn
Department of Computer Science
Makerere University, Uganda
Email: jquinn@cis.mak.ac.ug

Ernest Mwebaze
Department of Computer Science
Makerere University, Uganda
Email: emwebaze@cis.mak.ac.ug

***Abstract*—The rapid geographical expansion of the Cassava Brown Streak Disease (CBSD) pandemic has devastated cassava crops in East and Central Africa. To monitor CBSD, surveyors deal with single fields of plants at a time. A score of severity of disease is visually assigned to the plant based on the percentage of necrotised root of the cross-sections cuttings of the root examined. This method has problems with operator variability. This study investigates various computer vision techniques which leads to the development of a system whose feasibility is later assessed and it involves five stages. In stage one, using mobile deployment, images were acquired. Stage two involves use of different techniques to obtain an annotated data set where each pixel is later classified as healthy or necrotised in stage three. In stage four, the performance of these classifiers was evaluated based on the Area Under the Curve (AUC) analysis and accuracy results, Mean Absolute Error (MAE) and R^2 score. To assess operator variability, by comparing two sets of predictions of two different surveyors, a confusion matrix was used to determine how these predictions differed. With the R^2 score of 0.789, Nearest Neighbors classifier performed better than the other classifiers and thus it was recommended as the best method assign a score of necrosis to the diseased cross-section cuttings of the cassava root.**

I. Introduction

Cassava (Manihot esculenta) is an extremely important crop in Africa, 200 million people in the continent depend on it. In Sub-Sahara Africa, cassava can represent up to 60% of the daily calorie intake, and is largely consumed locally. Cassava grows in moderately poor soils with limited labor requirements and it tolerates drought. Thus cassava is a significant food security crop, mainly in drought-stricken areas [1][2].

The leading diseases affecting cassava in these areas are CBSD and Cassava Mosaic Disease (CMD). CBSD is a risk to food security, because the severity of root damage caused by the disease escalates the longer it stays in the field. CBSD, which is caused by a virus, was at first confined to coastal, low altitude areas in East Africa, but since the mid-2000s the disease has spread speedily, affecting Tanzania, Uganda, Kenya, Rwanda and Burundi. The disease has at present infected around 80% of crops in Uganda and around 20% of crops in Rwanda and Burundi [3]. CBSD is a more significant cause of crop loss in these regions than was earlier believed [4] since the disease causes both quantitative and qualitative decrease in total root yield by rotting of roots thus making them unmarketable and unpalatable. For cassava plants infected with CBSD, the major part affected is the tuber/root of the plant. To monitor CBSD, surveyors deal with single fields of plants at a time. When out in the field, normally they dig up a set of plants in selected gardens and examine five cross-section cuttings of the root. A score of severity of disease is allocated to the plant based on the average percentage of necrotised root of all five cross-sections examined. However, visual assessment of the symptoms to determine the score of severity of disease of a root by an expert may differ from the score of severity by another thus rendering this method inconsistent. To quantify the problem of operator variability, we asked two experts to assign scores of severity of disease to the same cassava cross-section cutting of the root, the results were obtained and the level of disagreement was obtained.

This paper presents an innovation to overcome this challenge. We present computer vision techniques for using camera-enabled mobile devices to automatically assign a score of severity of necrosis directly, allowing confidence on survey workers with lower levels of training, and thus reducing survey costs. Precisely, given expert-annotated single images of infected cassava tubers, we demonstrate classification based on pixel information.

II. Related work

According to Mahlein et al. [5], Spectral Vegetation Indices (SVIs) have been shown to be useful for an indirect detection of plant diseases. Nevertheless, these indices have not been evaluated to detect or to differentiate between plant diseases on crop plants. The study developed specific Spectral Disease Indices (SDIs) for the detection of diseases in crops. Sugar beet plants and the three leaf diseases Cercospora leaf spot, sugar beet rust and powdery mildew were used as sample diseases. Hyperspectral signatures of healthy and diseased sugar beet leaves were assessed with a non-imaging spectroradiometer at different developing stages and disease severities of pathogens. To develop hyperspectral indices for the detection of sugar beet diseases the best weighted combination of a single wavelength and a normalised wavelength difference was thoroughly searched testing all potential groupings. The optimised disease indices were tested for their ability to detect and to classify healthy and diseased sugar beet leaves. With a high accuracy and sensitivity healthy sugar beet leaves and leaves, infected with Cercospora leaf spot, sugar beet rust and powdery mildew were classified. Spectral disease indices were also successfully applied on hyperspectral imaging data and on non-imaging data from a sugar beet field.

Fig. 1. Image sample as captured by camera

Smith and Camargo [6] developed an image-processing based algorithm to automatically identify plant disease visual symptoms. The study described an image-processing based method that identifies the visual symptoms of plant diseases by analysis of colored images. Results showed that the developed algorithm was able to identify a diseased region even when that region was represented by a wide range of intensities.

A hybrid intelligent system from color imagery for grape leaf disease detection was suggested by [7]. The system consisted of three core parts: grape leaf color segmentation, grape leaf disease segmentation and classification of diseases. The system was able to classify the image of grape leaf into three classes which were scab disease, rust disease and no disease.

Smith and Camargo [8] performed an image pattern classification for the identification of disease causing agents in plants. A machine vision system for the classification of the visual symptoms of plant diseases was implemented by analysis of colored images. A set of image features was extracted from each diseased region. Feature selection was then performed to identify which of these provided most information about the image domain. A Support Vector Machine (SVM) was used as a learning machine and cross-validation was the discrimination method used to identify the best classification model.

III. Assessment of Operator Variability

A. Methods

1) Image Acquisition: Image samples of cross sectional cut cassava tubers placed on a plain board were captured from Namulonge Crops resources Research Institute, Uganda. 15 root discs from the same genotype of cassava were cross sectionally cut, placed on a plain board and captured with a standard digital camera.

2) Expert Annotation: Expert annotation was a process where an expert visually assigned a score of severity of disease to the sample images used.

B. Results

1) Confusion Matrix: To assess operator variability, by comparing two sets of predictions of two different surveyors, a confusion matrix was used to determine how these predictions differed. The results are shown in Figure **??** and Table I:

TABLE I. Surveyor Score confusion matrix

Scores	1	2	3	4	5
1	1	0	0	0	0
2	0	6	1	0	0
3	0	0	2	3	0
4	0	0	1	3	2
5	0	0	0	0	2

Fig. 2. Cropped image

Fig. 3. Binary image

Referring to Table I:

- The *rows* correspond to the score results as assigned by Surveyor2.
- The *columns* correspond to the score results as assigned by Surveyor1.
- The *diagonal elements* in the matrix represent the number of same score results that both surveyors assigned to the same image.
- The *off-diagonal elements* represent the score results that were assigned differently by both surveyors to the same images.
    - Off-diagonal row elements represent Surveyor1's score results that differed from Surveyor2's score results. E.g. In the second row, one image was assigned a score *2* by Surveyor1 and Surveyor2 assigned the same image a score of *3*.
    - Off-diagonal column elements represent Surveyor2's score results that differed from Surveyor1's score results. E.g. In the fourth column, three images were assigned a score of *4* by Surveyor2 and Surveyor1 assigned the same image a score of *3*.

Based on the outcome of the confusion matrix, it is seen that different scores are assigned to the same image by two different surveyors. This shows how this method has problems with operator variability and therefore an automated system is a more feasible solution as compared to the surveyor visual assignment method.

IV. Automated Symptom Measurement

A. Methods

1) Image Segmentation: In this study, images were cropped manually using a cropping tool. However, cropped images had a white background and working with the whole image also brings inaccurate results too. To eliminate the white back ground, only non-pure white pixels in the image were extracted automatically, Samples of the resulting cropped image and binary image are shown in Figure 2 and Figure 3 respectively.

A threshold was applied separately to each cropped root image so as to obtain its respective binary image as shown above. Coordinates of the non-pure white pixels were obtained. Using these coordinates, from the respective binary image, each pixel was labeled healthy or necrotised.

2) Ground Truth Data: Image pixel data was extracted from the original image and its corresponding binary image. From the original image, the RGB pixel data with the corresponding location coordinates,(i, j) were extracted.

3) Classifier Training: In the classification, the method used for splitting data set into training and testing was the k-fold cross-validation sometimes called rotation estimation, the data set was randomly split into mutually exclusive subsets of approximately equal size[9]. The classifiers used were; Nearest Neighbors[10], Decision tree[11][12], Random Forest [13][14], Naive Bayes [15] and Support Vector Machine (Linear SVM)[16][17].

To evaluate classifier performance, four performance measures were implemented, i.e., Receiver Operating Characteristic (ROC), AUC and predictive accuracy score, which evaluated how good the classifiers performed when distinguishing necrotised pixels from healthy pixels. MAE and R^2 score, the coefficient of determination evaluated how good the classifiers performed in the overall prediction of the percentage of necrotisation in the root. The data and source code are available at https://github.com/tjovia/CBSD.git

B. Results

1) Actual Necrosis Percentage: To calculate the actual percentage of necrosis of a given test image, the binary pixel labels were used. From Equation 1,

$$\text{Actual\% of Necrosis} = \left(\frac{number\ of\ pixels\ labeled\ necrotised}{Total\ number\ of\ pixels}\right) \times 100 \quad (1)$$

2) Accuracy per pixel: To choose the best performing classifier, basing on accuracy per pixel, results of AUC and predictive accuracy score with cross-validation were obtained. In this section the results for both methods will be presented. A ROC graph is a technique for visualising, organising and selecting classifiers based on their performance. Given a classifier and an instance, there are four possible outcomes, i.e., true positive (TP), false negative (FN), true negative (TN) and false positive (FP) [18].

Table II shows the results for AUC obtained for the different classifiers for the RGB color space. If a classifier yields a 1.0, then it is a perfect test. 0.9 to 0.99 is an excellent test, 0.8 to 0.89 is a good test, 0.7 to 0.79 is a fair test, 0.6 to 0.69 is a poor test, 0.5 to 0.59 is a failed test and below 0.5 the classifier is negatively correlated with the target.

With Predictive Accuracy Score, the accuracy of the test approximates how effective the algorithm is by showing the probability of the true value of the class label; summing it all up, it evaluates the overall effectiveness of the algorithm [19]. The higher the probability, the higher the predictive accuracy score. Four sample images were used to determine the probabilities as shown in Table III.

TABLE II. AUC OF CLASSIFIERS FOR RGB COLOR SPACE

Sample Image	Image1	Image2	Image3	Image4
Naive Bayes	0.96	0.96	0.96	0.96
Linear SVM	0.97	0.96	0.97	0.97
Decision Tree	0.96	0.96	0.97	0.96
Nearest Neighbors	0.96	0.96	0.96	0.96
Random Forests	0.97	0.97	0.97	0.97

TABLE III. ACCURACY SCORE OF CLASSIFIERS FOR RGB COLOR SPACE

Sample Image	Image1	Image2	Image3	Image4
Naive Bayes	0.90	0.89	0.90	0.89
Linear SVM	0.92	0.91	0.91	0.92
Decision Tree	0.92	0.92	0.92	0.92
Nearest Neighbors	0.92	0.92	0.92	0.92
Random Forests	0.92	0.92	0.92	0.92

3) Accuracy per root sample: Performance measures used to determine best performing classifier, basing on accuracy per root sample, were Mean Absolute Error (MAE) and R^2 Score, the Coefficient of Determination. In this section the results for both methods will be presented. The MAE was used to measure how close predictions of the overall percentage of necrotisation of a root were to the actual percentage of necrotisation of the root. The MAE estimated over N is defined as;

$$MAE(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| . \quad (2)$$

where $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value.

The results for (MAE) for the different classifiers are shown in Table IV.

The R^2 score, the coefficient of determination was calculated and this provided results on how well the overall percentage of necrotisation is predicted by the model. If $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value, then the score R^2 estimated over N is defined as,

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (3)$$

where $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$.

The results for R^2 score for the different classifiers are shown in Table V.

Naive Bayes classifier and Nearest Neighbors classifier had the lowest (MAE) of *0.049* while the other classifiers had a slightly higher MAE. This meant that Naive Bayes classifier and Nearest Neighbors classifier were more reliable models compared to the other classifiers. By comparing the classifiers used, the Nearest Neighbors classifier had the highest result of the R^2 score of *0.789*. This means that *79%* of the total

TABLE IV. MEAN ABSOLUTE ERROR FOR THE DIFFERENT CLASSIFIERS

Classifier	Mean Absolute Error
Naive Bayes	0.049
Linear SVM	0.059
Decision Tree	0.052
Nearest Neighbors	0.049
Random Forest	0.053

TABLE V. R^2 SCORE FOR THE DIFFERENT CLASSIFIERS

Classifier	R^2 score
Naive Bayes	0.691
Linear SVM	0.605
Decision Tree	0.688
Nearest Neighbors	0.789
Random Forest	0.662

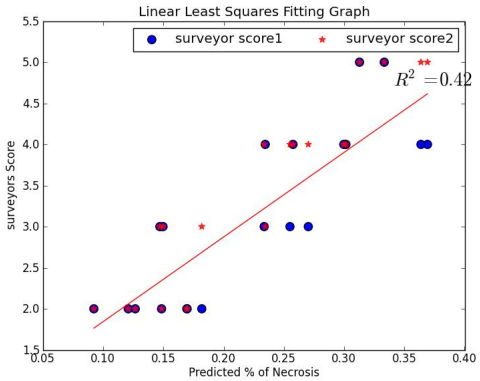

Fig. 4. Surveyor score and predicted necrosis percentage

variation in Actual Percentage Necrosis is determined by the linear relationship between Nearest Neighbors Percentage Necrosis and the Actual Percentage Necrosis. Because of the highest result of the R^2 score compared to other classifiers, Nearest Neighbors classifier proved to be the more reliable the model.

The score of necrosis that was visually assigned by an expert from Namulonge Crops resources Research Institute, the predicted score of necrosis and the actual score of necrosis for the sample images was compared. The performance measure used was Linear least squares fitting and the goal was to ascertain the relationship between different pairs of these variables. And these were:

- Surveyor1 and 2 Score and Predicted necrosis Percentage.
- Actual necrosis Percentage and Predicted necrosis Percentage.
- Actual necrosis Percentage and Surveyor1 & Surveyor2 Score.

Linear least squares fitting graphs were plotted as shown in Figures 4, 5 and 6.

R^2 is a measure of how close the data are to the line and the higher the R^2, the better the model fits the data. Based on the

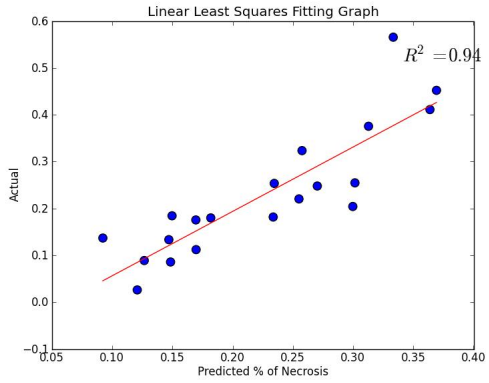

Fig. 5. Actual necrosis percentage and predicted necrosis percentage

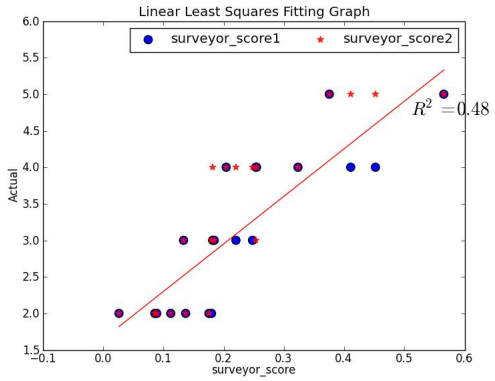

Fig. 6. Actual necrosis percentage and surveyor score

TABLE VI. R^2 SCORE FOR THE DIFFERENT RELATIONSHIPS

Relationship	R^2 score
Surveyor Score and Predicted necrosis Percentage	0.42
Actual necrosis Percentage and Predicted necrosis Percentage	0.94
Actual necrosis Percentage and Surveyor Score	0.48

results in Table VI, the R^2 score of actual necrosis percentage and predicted necrosis percentage is very high compared to actual necrosis percentage and surveyor score. This means from the results in Table VI, the model designed will predict the overall percentage of necrosis of a root well. Based on the results of the R^2 score after the linear squares fitting, it is seen how an automated system is a more feasible solution as compared to the method of surveyor visual score assignment. From the results of the R^2 score of actual necrosis percentage and predicted necrosis percentage being *0.94* and the highest as compared to actual necrosis percentage and surveyor score with an R^2 score of *0.48*, an automated score assignment to the necrotised root is a more feasible solution as compared to the method of surveyor visual score assignment.

V. CONCLUSION

The research has proved that an automatic symptom measurement system specifically one that assigns a score to a necrotised cassava tuber infected with CBSD is a feasible solution, and has consistency advantages as compared to

the surveyor visual score assignment method. If used, it would avert the challenge of inconsistent data collection by surveyors and would speed up the process of developing new cassava varieties that are resistant to CBSD. It has been shown how different classification techniques have been applied to automatically assign the score to the cassava tuber. Five classifiers were tested to get results; all classifiers performed well though some performed better than others. Nearest Neighbors classifier and Naive Bayes classifier had the lowest MAE, however Nearest Neighbors classifier had the highest result of the R^2 score and based on that, it was proved to be the more reliable the model as compared to the other four classifiers. The model was assessed and the results proved that it was a more feasible solution as compared to the method of visual assignment of the score of necrosis. In the assessment of the feasibility of the model, based on the results of the confusion matrix and the R^2 score after the linear squares fitting, it was shown how an automated system is a more feasible solution as compared to the method of surveyor visual score assignment. From the confusion matrix, the difference in scores by the surveyors to the same image showed how the surveyor visual score assignment has problems with operator variability. From the results of the R^2 score of actual necrosis percentage and predicted necrosis percentage being the highest as compared to actual necrosis percentage and surveyor score with an R^2 score, it was demonstrated that an automated score assignment to the necrotised root is a more feasible solution as compared to the method of surveyor visual score assignment.

The work can be incorporated in to a mobile version. This is planned to be done by using Open Data Kit (ODK) which is a is an open source suite of tools that enables data collection on mobile phones and data submissions to a central server. This could then improve on the monitoring of cassava brown streak disease by providing real-time information because not only experts but volunteers or farmers can take the images and then they can be uploaded on to a server at the research institute where they can be processed. Furthermore, it could improve on the prediction and optimisation of plant protection measures since there is consistency of results.

ACKNOWLEDGMENT

The authors would like to thank the AI-DEV group in the School of Computing and Informatics Technology, Makerere University for giving thoughts, advice and ideas for improvement of the study and Namulonge Crops resources Research Institute, Uganda for the support.

REFERENCES

[1] A. L. Chávez, T. Sánchez, G. Jaramillo, J. Bedoya, J. Echeverry, E. Bolaños, H. Ceballos, and C. A. Iglesias, "Variation of quality traits in cassava roots evaluated in landraces and improved clones," *Euphytica*, vol. 143, no. 1-2, pp. 125–133, 2005.

[2] N. Nassar and R. Ortiz, "Cassava improvement: challenges and impacts," *The Journal of Agricultural Science*, vol. 145, no. 2, pp. 163–171, 2007.

[3] B. for Farming in Africa, "Cassava," June 2013. [Online]. Available: http://www.b4fa.org/biosciences-and-agriculture/cassava/

[4] P. Ntawuruhunga and J. Legg, "New spread of cassava brown streak virus disease and its implications for the movement of cassava germplasm in the east and central african region," *USAID, Crop Crisis Control Project C3P*, 2007.

[5] A.-K. Mahlein, T. Rumpf, P. Welke, H.-W. Dehne, L. Plmer, U. Steiner, and E.-C. Oerke, "Development of spectral indices for detecting and identifying plant diseases," *Remote Sensing of Environment*, vol. 128, no. 0, pp. 21 – 30, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425712003793

[6] J. Smith and A. Camargo, "An image-processing based algorithm to automatically identify plant disease visual symptoms," *Biosystems Engineering*, vol. 102, no. 1, pp. 9–21, 2009.

[7] A. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, and A. Srikaew, "Grape leaf disease detection from color imagery using hybrid intelligent system," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008. ECTI-CON 2008. 5th International Conference on*, vol. 1. IEEE, 2008, pp. 513–516.

[8] A. Camargo and J. Smith, "Image pattern classification for the identification of disease causing agents in plants," *Computers and Electronics in Agriculture*, vol. 66, no. 2, pp. 121–125, 2009.

[9] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, vol. 14, no. 2, 1995, pp. 1137–1145.

[10] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan kaufmann, 2006.

[11] M. N. Anyanwu and S. G. Shiva, "Comparative analysis of serial decision tree classification algorithms," *International Journal of Computer Science and Security*, vol. 3, no. 3, pp. 230–240, 2009.

[12] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[13] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[14] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[15] D. M. Farid, L. Zhang, C. M. Rahman, M. Hossain, and R. Strachan, "Hybrid decision tree and naïve bayes classifiers for multi-class classification tasks," *Expert Systems with Applications*, 2013.

[16] V. N. Vapnik, "Statistical learning theory," 1998.

[17] P. Bartlett and J. Shawe-Taylor, "Generalization performance of support vector machines and other pattern classifiers," *Advances in kernel methods: support vector learning*, pp. 43–54, 1999.

[18] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[19] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," in *AI 2006: Advances in Artificial Intelligence*. Springer, 2006, pp. 1015–1021.