

# Automated Tobacco Grading Using Image Processing Techniques and a Convolutional Neural Network

Charlie S. Marzan and Conrado R. Ruiz Jr.

**Abstract**—Tobacco grading is very important for crop market price determination. It is beneficial for graders who need to manually classify tobacco leaves according to their grades. As such, a grading system using image processing techniques and a Convolutional Neural Network (CNN) is proposed in this study which aims to surpass existing results in tobacco grading. The system consists of image acquisition, pre-processing, leaf detection, segmentation, and classification. Tobacco leaf images were directly taken at the tobacco grading room and pre-processed for subsequent tasks. Through a Haar cascade classifier and applying image processing techniques, air-cured tobacco leaves are automatically detected and extracted in images. This method produced satisfactory results as it can successfully detect single and multiple tobacco leaves taken under different positions and scale. All detected tobacco leaves underwent various image processing to precisely segment leaves from the rest of the image. The experimental results also reveal that using segmented and nonsegmented images, CNN classifier can effectively grade tobacco leaves as high as 96.25% accuracy rate and on average, took 7.43 ms to classify a single tobacco leaf. This approach outperforms current methods in grading tobacco leaves.

**Index Terms**—Convolutional neural network, image processing techniques, leaf detection, tobacco grading.

## I. INTRODUCTION

Tobacco industry is one of the strongest pillars of the world economy because of its high commercial value. In fact, [1] reported that tobacco companies are among the highest profit making organizations in the world in 2017. The Imperial Tobacco Group in the United Kingdom was proclaimed as the largest global tobacco company with about 39.2 billion US dollars worth of sales. Moreover, Philip Morris International and British American Tobacco came in second and third in rank respectively. A high-quality tobacco leaf production must continue to take advantage of the growing export markets and must address quality requirements for domestic cigarette manufacturing. Therefore, quality evaluation or grading of tobacco leaves plays a crucial role in quality assurance of tobacco productions.

Grading tobacco leaves remains to be a difficult task because it is still manually performed even in this modern age. After curing, the leaves are carefully graded according to their appearance which includes the size, texture, shape, maturity, leaf structure, oil content, injury tolerance, waste

tolerance, and color intensity [2], [3]. This process requires highly skilled graders for quality assurance. However, there are several factors that influence the grader's judgment such as the grader's emotion, human eyesight, and the lighting condition in the work area [4]. This form of human observation is highly subjective and is mainly based on experiences.

In order to overcome the drawbacks of such traditional method, new techniques are needed to automate the grading process. This has been proven by the previous studies [4]-[9] where they applied image processing and machine learning techniques to develop an automatic grading system for flue-cured tobacco leaves. They proposed algorithms to extract the leaves' color, shape, texture, injury and waste tolerance, stalk position, and maturity. Relative to classification, machine learning algorithms like Fuzzy [4], Support Vector Machine [5], K-Nearest Neighbor [6], Generalized Regression Neural Network [7], and Backpropagation Neural Network [8] were used to classify tobacco leaves. Because the classification features were selected based on human experience, these approaches improved the accuracy of tobacco grading. However, the classification accuracy is still not high enough and is vulnerable to artificial feature selection.

Developed in recent years, Convolutional Neural Networks (CNNs) have become popular especially in computer vision because they achieved remarkable performances on different tasks such as image classifications in the field of agriculture [9]-[11]. CNN is an end-to-end pipeline that can automatically discover the useful features for classification, whose advantage lie in local spatial coherence in the input images, which allow them to have fewer weights as some parameters are shared. Inspired by the development of CNN in image-based recognition and not much studies on the application of CNN in tobacco leaf grading, this study attempts to use CNN in grading air-cured tobacco leaves and compare with the existing models.

## II. RELATED WORK

There have been previous studies about grading tobacco leaves using different image processing and neural network based classification techniques.

The study of [7] proposed a tobacco grading system based on image processing and Generalized Regression Neural Network (GRNN). Images used in the experiments are real-time acquisition from the Hena province. The 108 digital images underwent image processing to extract their shape, color, texture, and spectral features. After training the GRNN, the classification accuracy rate reached 88%. The accuracy was improved by putting the input characteristic values

Manuscript received August 25, 2019; revised October 17, 2019.

Charlie S. Marzan is with Don Mariano Marcos Memorial State University, Philippines (e-mail: charlie\_marzan@dlsu.edu.ph).

Conrado R. Ruiz Jr. is with De La Salle University, Philippines (e-mail: conrado.ruiz@dlsu.edu.ph).

multiplying the corresponding Mean Impact Value (MIV) as the new input and retrained the network. This means that the neural network with the MIV method to screen the input was feasible. MIV can screen the input variables of neural network and establish artificial neural network model with less input and better effect. The advantages of GRNN include even sample data is not so much, it can still get better prediction, and the neural network can deal with unstable data.

Reference [8] employed Backpropagation Neural Network (BNN) in the classification of non-cured tobacco leaves as cigarette raw material. A total of 200 images from various grades of tobacco leaves served as the dataset of the study. Pre-processing was first performed in images before extracting the color, shape, and texture features. Using BNN classifier, the classification rate for tobacco leaf grading reached 77.50%. In addition, the latest study [9] utilized CNN in classifying flue-cured tobacco leaves in which 120 images were used in the training. The proposed CNN's architecture was composed of three convolutional layers, three pooling layers, and one fully-connected layer. The proposed model obtained 85.10% accuracy on the test set comprising 15 images.

The presented studies [7], [8] in tobacco grading did feature extraction to train the GRNN and BNN models. It is in contrast with the CNN used by [9] which does not take features as input. It eliminates the need for feature extraction or selection, making the whole process much more convenient and less prone to user bias. These studies show that CNNs have not only reduced the demand of image pre-processing, but also improved the classification accuracy.

### III. PROPOSED METHOD

The flowchart of the proposed method in tobacco grading is illustrated in Fig. 1. It comprises five stages: (i) image acquisition, (ii) image pre-processing, (iii) tobacco leaf detection, (iv) image segmentation, and (v) classification.

#### A. Image Acquisition

A total of 800 Burley (air-cured) tobacco leaves was photographed using Nikon D7000 DSLR camera with 2448 x 3696 pixel resolution and saved as JPEG file format. These leaves were taken at the tobacco grading room of the Universal Leaf Philippines, Inc. in the province of La Union, Philippines. This set of digital images has an equal representation of the grades in which each leaf was labeled by a grading expert. In this study, classifying air-cured tobacco leaf images is only limited to four grades such as BH, BRM, CHA, and BM. Fig. 2 shows sample tobacco leaf images with their corresponding grade.

Tobacco leaves are graded based on the position they grow on the stalk, quality, and color. Four positions were identified (e.g. lugs (X), cutters (C), leaf (B), and tip (T)), corresponding to the lower, middle, and upper portion of a stalk. The judgment for stalk position is mainly based on the color and texture.

Furthermore, high (H), medium (M), and low (L) are the three types of quality of tobacco. These types are dependent on the stalk position, color, injury and waste tolerance, and maturity. As to color, four categories were specified such as

tan (F), tannish-red (R), buff (L), variegated (K), and running green (V). The classification of tobacco leaves is symbolized as, for instance CHA, where C indicates that the leaf grew from cutters position, H denotes high quality, and A means that the leaf has spots.

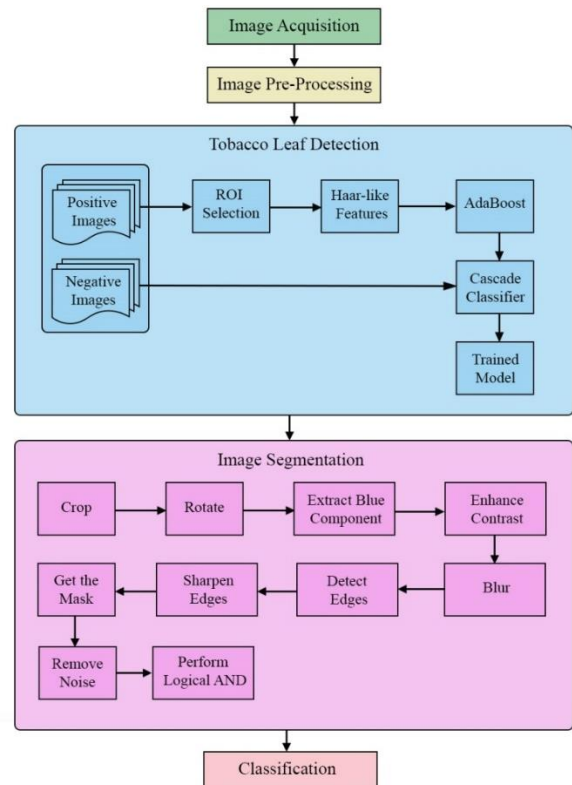


Fig. 1. Flowchart of the proposed method.

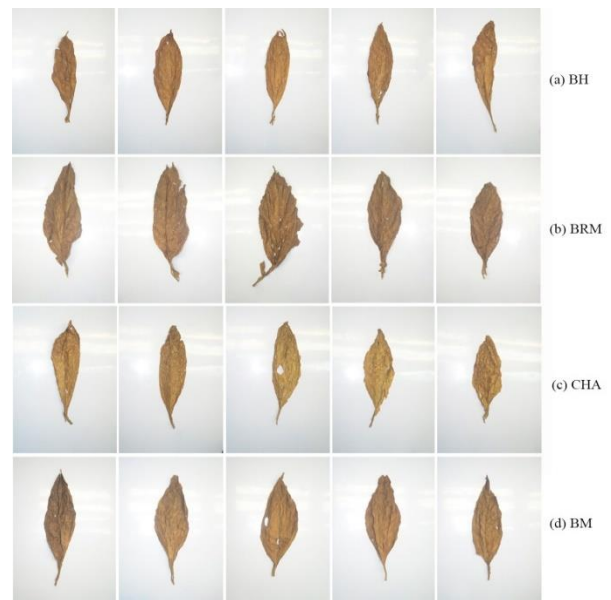


Fig. 2. Samples images in the dataset.

#### B. Image Acquisition

The dataset consists of 2448 × 3696 pixel RGB images. Since the resolution is quite high, this could lead to unexpected results. To unify the approach before attempting to detect leaves in images, all images are scaled down to 300 x 452 dimension. The ratio of the new image to the old image was first calculated before performing the actual resizing. Thus, the images' original aspect ratio was not altered in the

scaling process. By scaling images to a standard size, more accurate results can be obtained by the algorithm designed to detect and extract individual leaves.

### C. Tobacco Leaf Detection

The Haar cascade classifier method [12] was used to automatically detect tobacco leaves in images and was implemented using OpenCV Python. Before training, the training set containing both the positive and negative images was first prepared. The positive images contained the Region of Interest (ROI), in this case a tobacco leaf image, whereas negative images are those which do not contain the desired object. The training set comprised 1,000 images, among them are 600 positive and 400 negative images. The test set, on the other hand, contained 100 positive and 50 negative images carrying the best and worst cases. The positive images were taken in the dataset discussed in subsection A whereas negative images were individually obtained from Google image search. All images were then resized keeping the ratio of width to height.

Aside from the images, a text file containing the location of tobacco leaf in images (e.g. upper left pixel of ROI, width and height of tobacco leaves) was prepared. During the training, the classifier scans the features (shown in Fig. 3) in positive images and creates specific target values by using the sum values of black areas and the white areas in the features. The classifier produced most optimized values for detecting and tracking the object by changing the sizes of the features.

A positive image can have many features with different scales and positions. Hence, the calculations slow down the detection process. To solve this problem, the AdaBoost algorithm was used to reduce the number of irrelevant features in a detection. The principal idea of boosting was to combine many weak classifiers to produce a more powerful one. The training continues until classifiers with a number of specified stages were obtained. All these classifiers were then combined into a larger classifier. This results in cascading classifier, where the classification results of the initial classifier were forwarded to the classifier at the succeeding stage. Using a lot of positive and negative images facilitate the detection of tobacco leaves in images.

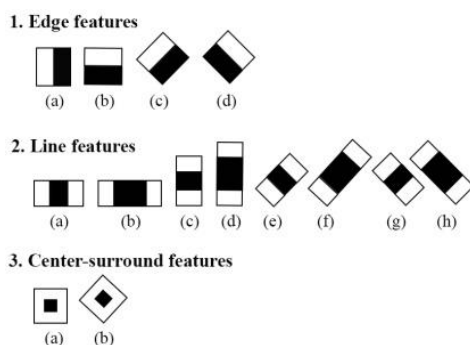


Fig. 3. Haar-like features [13].

### D. Image Segmentation

Image segmentation is the process of separating the object from the rest of the image [14]. The tobacco leaf images underwent segmentation processing stage because all detected leaves should be facing upward. The bounding box enclosing the ROI was first extracted or cropped. Since the Haar cascade classifier can detect leaves posed in different

angles or positions, rotating into the desired alignment was followed to ensure all images used for subsequent processing have a uniform position. In doing such, the extracted RGB image was then converted to a binary image and applied threshold before finding the boundaries of the leaf image. The two boundary points that are farthest from each other was obtained and drew a line from one end to the other. Then, the angle was computed and rotated the image to make a leaf facing upward. However, there were some instances that the leaf was facing downward. To deal with this problem, the tip and petiole of a leaf were identified so that the parameter value in the leaf rotation will be adjusted. The output of the rotation algorithm was an RGB image.

After rotation, the B component in the RGB image was extracted. The reason why only B component was used is that according to the statistical information of tobacco leaf images [15], most of the pixels which belong to a tobacco leaf have the B component value ranging 17-88. Furthermore, the pixels with the B component value from 60-80 has 100% probability to tobacco leaf. Other pixels having the value from 0-16 or 89-255 has a very high probability to background or noise.

Using the gamma correction, the contrast of the image was enhanced. The resulting image was slightly blurred with a Gaussian filter to minimize the effects of small holes or irregularities in images. To sharpen the edge of the leaf image, Canny edge detection was applied and added to the blurred image. The mask of the leaf image was then created using the Otsu's algorithm to determine the optimal threshold value. A binary image was the output of this image processing technique.

The binary image contains the ROI and some noise. Connected components except for the largest component were all removed as these considered as noise. By treating the white pixels in the binary image as ones and the black as zeros and then performing a logical AND of every pixel in the mask with the identically-positioned pixel in the rotated RGB image, a new image consisting only the ROI can be derived. This completed the isolation or segmentation of the leaf object from the rest of the image.

### E. Classification

In this study, Tensorflow was used to build a CNN-based tobacco leaves classifier. TensorFlow is an open sourced deep learning framework developed by Google, allowing a user to quickly and efficiently implement various algorithms fundamental to neural networks. Given the wide range of functions already made available, as well as the community support, TensorFlow was chosen over other well-known frameworks at this time.

Before training the classifier, the segmented leaves produced in the previous section were scaled down preserving the ratio and padded some amount of black pixels to the left and right sides of the image resulting in a 128 x 128 pixel image. This pre-processing aimed to generate standardized input data for the training process of the neural network. The dataset containing 800 images were split into a training, validation, and test set. Sixty percent images was used for training while the remaining percentage was divided into validation and test set. In a normal machine learning task, a model is trained on the training set and evaluated by testing

it on the test set. A validation set is used to ensure that the best possible model is obtained. This was done by using the validation set to fine-tune the parameters of the model during training, by periodically evaluating on the validation set.

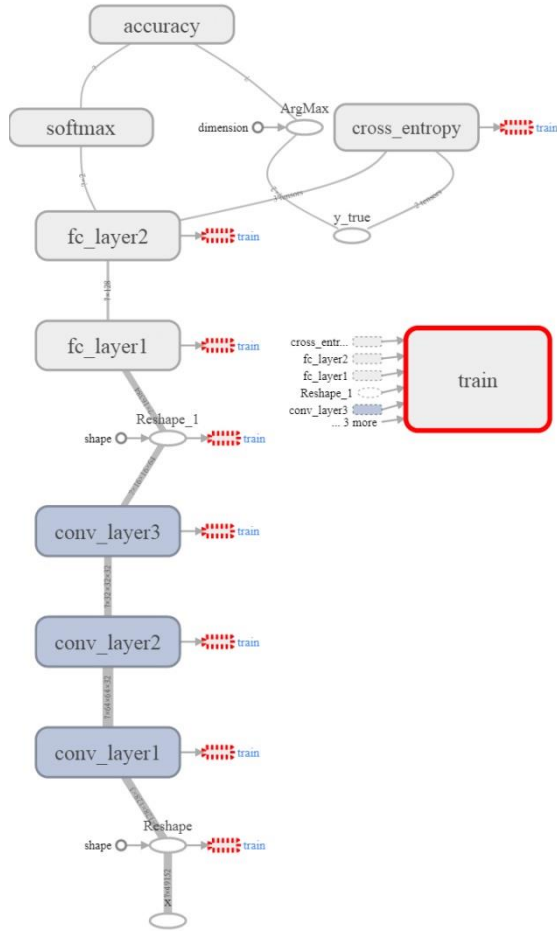


Fig. 4. CNN computational graph.

Fig. 4 shows the computational graph of the CNN after training. This graph was generated by Tensorboard which visualizes the layer output shapes at different levels, displayed in the greyed connections.

While training, images from the four classes were fed to a convolutional layer which is followed by two more convolutional layers. After creating a convolutional layer, biases were added to the neuron, which are also learnable or trainable. Then, the output of max-pooling is fed to RELU as the activation function. After which, the last convolutional layer was flattened to convert the multi-dimensional tensor into a one-dimensional tensor. Two fully-connected layers were added at the end. The second fully-connected layer has only four outputs which represent each class.

To get the probability of each class, softmax was applied to the output of the fully-connected layer. The class having higher probability is the prediction of the network. The softmax function provides outputs ranging from 0 to 1. Thereafter, the cost that will be minimized to reach the optimum value of weights was defined. This study used a simple cost that was calculated using a Tensorflow function `softmax_cross_entropy_with_logits()` which takes the output of last fully-connected layer and actual labels to calculate `cross_entropy`. In terms of optimization function, AdamOptimizer was employed for gradient calculation and weight optimization. Training images were passed in a batch

of 16 in each iteration.

#### F. Performance Evaluation

The correct classification and misclassification in tobacco leaf detection and grading were summarized in a confusion matrix. The confusion matrix is a  $n \times n$  matrix in which each column represents the number of instances in an actual class and each row represents the instances in a predicted class. The evaluation metric derived from the confusion matrix was accuracy. The accuracy rate was computed by comparing the number of True Positives ( $TP$ ) and True Negatives ( $TN$ ) with the total number of  $TP$ ,  $TN$ , False Positives ( $FP$ ), and False Negatives ( $FN$ ). Accuracy is defined by the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

Computations and implementations of the proposed approaches in tobacco leaf detection, segmentation, and classification were performed in Windows 10 64-bit operating system, 2.80 GHz Intel(R) Core(TM) i7-7700HQ CPU, 16 GB RAM, and NVIDIA GeForce GTX 1050 4GB video RAM.

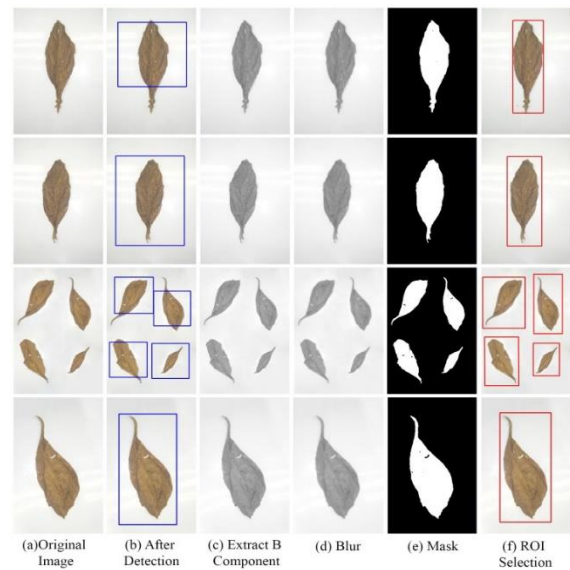


Fig. 5. Results of tobacco leaf detection in images.

#### A. Tobacco Leaf Detection

Fig. 5 shows the results of tobacco leaf detection using a Haar cascade classifier and image processing techniques. As can be seen in column (a), the input images contain a single leaf and multiple leaves taken under different positions and scale. These images were fed into the classifier and the results are seen in column (b). Accuracy rate of 91.33% was achieved by the classifier and was derived from 87 TP, 13 FP, 50 TN, and 0 FN. The result implies a high rate of correctly detecting tobacco leaves and nontobacco leaves in images. Even if the classifier can distinguish tobacco leaves, the ROI selection is not stable in some images. Some portions of the tip and petiole of the tobacco leaves were not captured which made the detection unstable. To resolve this problem, image

processing techniques were applied to improve the results. This includes extracting B component (column c), blurring (column d), masking or thresholding (column e), and selecting ROI after finding the leaf's boundaries (column f). Viewed in totality, the performance of the proposed leaf detection algorithm reached 100.00% accuracy rate as it managed to detect single and multiple tobacco leaves under different positions and scale.

### B. Segmentation

Segmentation of images was done to uniformly produce leaf images facing upward since the proposed algorithm can able to detect multiple leaves under different angles. The results of the processing steps of image segmentation are displayed in Fig. 6. The aim of image segmentation in this

context was first to extract the ROI after being bounded by the red colored rectangle, such as the images in Fig. 6.a and Fig. 6.b. Then, the image was rotated so that its major axis would be parallel to the vertical axis of the image frame (see Fig. 6.c and Fig. 6.d). The resulting image was first separated into its RGB components, and the blue (B) component was extracted to form a grayscale image (see Fig. 6.e). Thereafter, images underwent contrast enhancement, blur, and edge detection for sharpening the edges of the tobacco leaves (see Fig. 6f to 6i). Experience showed that thresholding every blue component image would produce a binary representation of the leaf object whose outline was not degraded. Some salt and paper noise are visible in the thresholded image (see Fig. 6j) and these were eliminated in Fig. 6k.

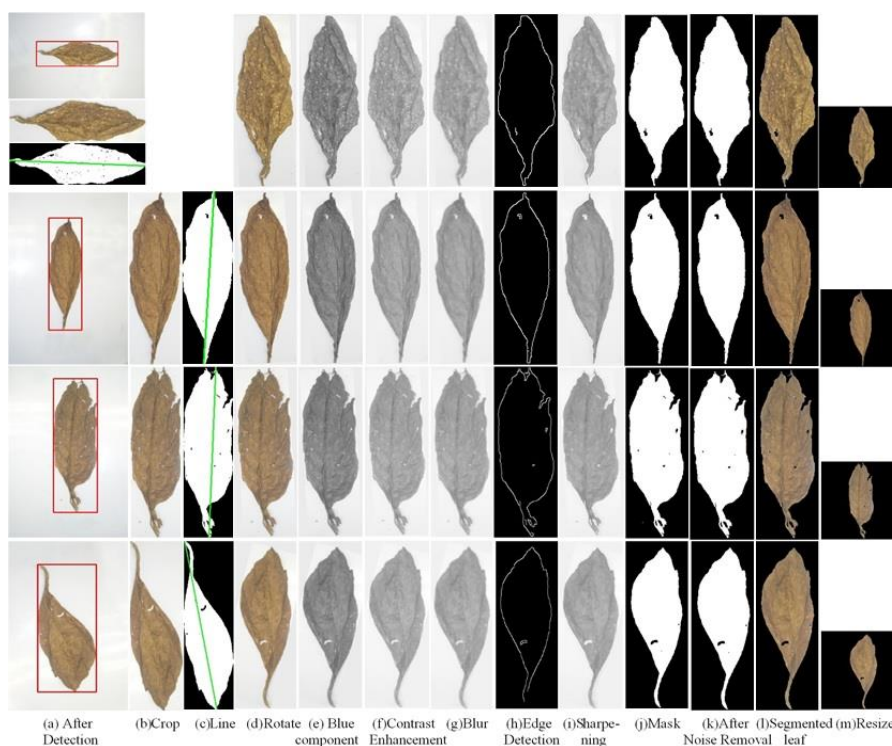


Fig. 6. Results of image processing techniques used for segmentation.

It was assumed that, by this stage, the leaf would be the largest object in the image. Images in Fig. 6.k can be regarded as a mask, which is a representation of the leaf shape (with all its holes) in the rotated image (d). Since image (d) and image (k) were of the same size in terms of pixel dimensions, multiplying each pixel value in image (d) by the corresponding value in image (k) now yielded image (l). All images were resized into  $128 \times 128$  (see Fig. 6.m) for consistency of the image sizes in building the CNN-based classifier. This completed the segmentation of the leaf object from the detected tobacco leaf image.

### C. Classification

Experiments on classification were done using three datasets, namely segmented (800) images, nonsegmented (800) images, and combination of segmented and nonsegmented (1,600) images. They were trained separately to determine which dataset can generate the highest classification rate in grading tobacco leaves. The performance of the classifier was obtained after 5,000 iterations. As a result, confusion matrices were derived from

the 20% test set of each dataset.

TABLE I: CONFUSION MATRIX OF THE CNN-BASED CLASSIFIER USING SEGMENTED IMAGES

Grades	BH	BM	BRM	CHA
BH	37	3	0	2
BM	3	37	0	0
BRM	0	0	40	1
CHA	0	0	0	37
Accuracy	94.37%			

TABLE II: CONFUSION MATRIX OF THE CNN-BASED CLASSIFIER USING NONSEGMENTED IMAGES

Grades	BH	BM	BRM	CHA
BH	35	3	1	2
BM	3	36	1	2
BRM	1	1	38	1
CHA	1	0	0	35
Accuracy	90.00%			

In Table I, the performance of the classification of the CNN-based classifier on the test set containing segmented

images arrived at 94.37% accuracy. It took an average of 7.40 ms to grade each tobacco leaf. On the other hand, the classifier's performance went down to 90.00% when nonsegmented images were used (see Table II) and on the average, the classifier spent 7.42 ms to grade each leaf. Surprisingly, in Table III, the classifier reached 96.25% accuracy when both segmented and nonsegmented datasets were employed. As to the time spent to classify single tobacco leaf, the classifier took 7.43 ms.

TABLE III: CONFUSION MATRIX OF THE CNN-BASED CLASSIFIER USING BOTH SEGMENTED AND NONSEGMENTED IMAGES

Grades	BH	BM	BRM	CHA
BH	75	5	0	1
BM	4	75	0	0
BRM	1	0	80	1
CHA	0	0	0	78
Accuracy	96.25%			

The result entails high probability of classifying correctly the tobacco leaf grades. It is worthy to mention that CNN produced satisfactory outcome in grading tobacco leaves in terms of accuracy and time using the combined segmented and nonsegmented images.

Further, Table III gives the number of tobacco images that were misclassified using both segmented and nonsegmented images. It is interesting to note that the misclassification rate was only 3.75% which is relatively low. The matrix also shows that the CNN classifier has a little problem distinguishing between BH and BM classes, but can make the distinction between other tobacco grades pretty well. Leaves graded as BRM got the highest classification performance. This pattern can also be observed from Table I and Table II.

Meanwhile, Fig. 7 shows sample misclassified images in the test set comprised of segmented and nonsegmented images where it can be perceived that BH and BM leaves have slight differences on their color. This corresponds to the confusion matrix in Table III.

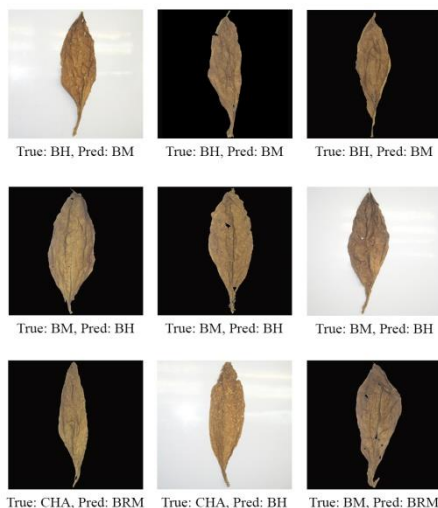


Fig. 7. Sample misclassified images in the test set carrying segmented and nonsegmented images.

#### D. Comparison

Table IV shows the comparison of different existing models with the proposed model in grading tobacco leaves. Only the existing studies which used neural network in the

classification phase were considered. It can be gleaned from the table that the proposed model in classifying tobacco leaves according to their grades surpassed the performance of the different models. It only shows that having a high training sample size would contribute to a better performance of the convolutional neural network. The performance of the CNN-based classifier really outperformed the classification techniques used by the previous studies in tobacco grading.

#### V. CONCLUSION AND FUTURE WORK

An automated tobacco grading using image processing techniques and Convolutional Neural Network was proposed in this study. Python, OpenCV, and TensorFlow were used as the simulation environment for this work. Using a Haar cascade classifier and image processing techniques, the accuracy rate of 100.00% proves the capability of the classifier to distinguish tobacco leaves from other objects. Results also show that the classifier could detect multiple leaves taken under various positions and scale. This automatic leaf detection facilitates the segmentation process as the ROI was accurately extracted. Applying different image processing techniques made the leaf segmentation successful.

TABLE IV: COMPARISON OF DIFFERENT NEURAL NETWORK MODELS WITH THE PROPOSED MODEL

Study	Training Sample	Classification Technique	Accuracy
Grading tobacco leaves based on image processing and GRNN (2012) [7]	108	Generalized Regression Neural Network	93.50%
Classification quality of tobacco leaves as cigarette raw material based on ANN (2017) [5]	200	Backpropagation on Neural Network	77.50%
Flue-cured tobacco leaves classification: A generalized approach using DCNN (2018) [12]	120	CNN	85.10%
Proposed model	960	CNN	96.20%

Among the three datasets used for classification, the CNN-based classifier generated a higher accuracy when it was trained in segmented and nonsegmented images. In fact, it reached an overall accuracy of 96.25% correct classification and on average, took 7.43 ms to classify a single leaf. The proposed model outperformed the different existing neural network models in classifying tobacco leaves. This result only shows that CNN achieved an accurate and computationally effective way of grading tobacco leaves.

In future work, additional images should be acquired, and classes should be increased to make the model more practical. Images taken under different illumination and complex background should also be considered as these conditions can be used in the real scenario. The authors also recommend performing further experiments like applying other deep learning algorithms.

#### ACKNOWLEDGMENT

The authors thank De La Salle University, Don Mariano Marcos Memorial State University, Commission on Higher

Education, National Tobacco Administration, and Universal Leaf of the Philippines Inc. for their support towards the fulfillment of this scholarly endeavor.

#### REFERENCES

- [1] H. Bhasin. Top 10 tobacco companies in the world in 2017. [Online]. Available: <https://www.marketing91.com/top-tobacco-companies/>
- [2] A. Yawootti and P. Kaewtrakulpong, "A machine vision system for thai flue-cured tobacco classification," in *Proc. Electrical Engineering/Electronics, Computer, Telecommunications, and Information Technology International Conference*, 2015, pp. 1-5.
- [3] S. Pan, M. Kudo, and J. Toyama, "Edge detection of tobacco leaf images based on fuzzy mathematical morphology," in *Proc. International Conference on Information Science and Engineering*, 2009, pp. 1219-1222.
- [4] F. Zhang and X. Zhang, "Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation," *Sensors*, vol. 11, no. 3, pp. 2369-2384, February 2011.
- [5] X. Ma, J. Shen, R. Liu, and H. Zhai, "Choice of tobacco leaf features based on selected probability of particle swarm algorithm," in *Proc. Control and Decision Conference*, pp. 3041-3044, 2016.
- [6] D. Guru, P. Mallkarjuna, S. Manjunath, and M. Shenoi, "Min-max representation of features for grading cured tobacco leaves," *Statistics and Application*, vol. 9, pp. 15-29, 2011.
- [7] J. Liu, J. Shen, Z. Shen, and R. Liu, "Grading tobacco leaves based on image processing and generalized regression neural network," in *Proc. International Conference on Intelligent Control, Automatic Detection and High-End Equipment*, 2012, pp. 89-93.
- [8] Y. Sari and R. A. Pramunendar, "Classification quality of tobacco leaves as cigarette raw material based on artificial neural networks," *Int. J. Comput. Trens Technol.*, vol. 50, no. 3, pp. 147-150, August 2017.
- [9] S. K. Dasari, K. R. Chintada, and M. Patruni, "Flue-cured tobacco leaves classification: A generalized approach using deep convolutional neural networks," *Cognitive Science and Artificial Intelligence Advances and Application*, pp. 13-21, 2018.
- [10] W. S. Jeon and S. Y. Rhee, "Plant leaf recognition using a convolutional neural network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 1, pp. 26-34, March 2017.
- [11] J. Champ, T. Lorieul, M. Servajean, and A. Joly, "A comparative study of fine-grained classification methods in the context of the lifecycle plant identification challenge 2015," presented at the Conference and

Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015.

- [12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2001, p. 1.
- [13] R. Lienhard and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. International Conference on Image Processing*, pp. 900-903, 2002.
- [14] V. Premalatha, M. Sumithra, S. Deepak, and P. Rajeswari, "Implementation of spatial fem for leaf image segmentation in pest detection," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 4, no. 10, pp. 471-477, October 2014.
- [15] S. Pan, M. Kudo, and J. Toyama, "Edge detection of tobacco leaf images based on fuzzy mathematical morphology," in *Proc. International Conference on Information Science and Engineering*, 2009, pp. 1219-1222.



**Charlie S. Marzan** was born on May 17, 1988 in La Union, Philippines. He obtained his bachelor degree of science in computer science in 2009 and master degree in Information technology in 2013, both at Don Mariano Marcos Memorial State University. He is current a PhD student in computer science at De La Salle University. He is passionate in research areas involving image processing and machine learning. He is currently a recipient of CHED K-12 Transition

Scholarship Program and an instructor at the College of Computer Science of DMMMSU-SLUC.



**Conrado R. Ruiz Jr.** obtained his bachelor's degree in computer science with a specialization in software technology (cum laude) from DLSU in 1999. He received his master's and doctorate degrees from the School of Computing of the National University of Singapore in 2001 and 2015, respectively. His several works in the areas of computer graphics, computational art, and multimedia information retrieval have been published. He was a BPI science

awardee in 1999 and received a special citation from the National Academy of Science and Technology during the Talent Search for Young Scientists in 2007. He is currently an associate professor at the College of Computer Studies of DLSU.