Weed Pixel Level Classification Based on Evolving Feature Selection on Local Binary Pattern With Shallow Network Classifier

Basil Andy Lease, WK Wong, L Gopal, and W. R. Chiong
Curtin University Malaysia, CDT 250, Miri 98009, Sarawak, Malaysia
E-mail: basil.andy.lease@postgrad.curtin.edu.my

Abstract. This paper proposes an evolving feature selection on a rotation-invariant Local Binary Pattern (LBP) with genetic algorithm (GA) and a non-linear classifier to perform pixel-wise classification on biomass pixels. Early true leaf growth of carrots and weeds consists of 60 images provided by a public benchmark Crop/Weed Field Image Dataset (CWFID) [1] was used. The GA encodes LBP feature parameters generated from normalized difference vegetation index (NDVI) image slices as genome consisting of number of neighbours, radius, size of image slice, number of LBP combinations. LBP is a lightweight rotation invariant texture feature descriptor which encodes discriminative local texture information between crops and weeds. The study evaluated multiple ensemble models evolved by GA, where GA evolves the LBP feature parameter selection, and the number of LBP features as input variables used. The classifier can handle crop and weed overlaps, which are often present in commercial fields. Weeds often grow in close proximity and overlap crops and are similar in size, which adds complexity in discriminating them. Our experiments have shown that combinations with a back-propagation neural network with a symmetrical two hidden-layer network achieved the best classification accuracy when compared to other non-linear classifiers. By utilizing a single type of feature (LBP texture feature), our resulting Artificial Neural Network (ANN) ensemble classifier is able to achieve a classification accuracy of 83.5 %.

1. Introduction
Certain parts of the world are currently experiencing a food crisis. This is due to the high food production demands unmet with ever-growing world population. The agricultural field needs improvements in agriculture production technology to meet such demands. Thus, precision agriculture and smart agriculture is currently an important field of research. Freshly sown field with young crops are often plagued with naturally growing weeds which compete for living resources. The early growth stages are critical as it will affect crop yield. The conventional method of farmers to eliminate weeds is by blanket spraying chemical herbicides using a tractor as it is cost-effective and easy to use. However, they pose environmental, and health concerns as excessive use may lead to groundwater contamination [2], river contamination [3], and food contamination [4]. The current advancement of technology allows for computer vision-enabled autonomous robots to manage and control weeds in the field. This would enable far greater precision than blanket spraying as robots would be able to target and spray weeds with the correct dose effectively. Thus, reduce waste and excessive use of chemical herbicides. To achieve this level of autonomy, an accurate weed discrimination system is required. The main challenge
is often related to occlusion and complex natural environment of weeds among the crop. Thus, the system needs to understand the discriminative features between weeds and crops.

This paper focuses on the development of a weed classification system in the natural complicated environment setting of crops and weeds. This paper suggests a strategy that does not involve image segmentation of crop and weeds. Instead, we propose that the texture characteristics can separate weeds and plants through Local Binary Pattern (LBP) obtained from the picture windows centred around each pixel containing crop data. LBP is selected as it has desirable characteristics such as lightweight, invariant of rotation, capable of encoding pattern replication and providing data on local texture. The image window would encode the texture data of neighbouring crops as a function. Our model utilises genetic algorithm to select the optimal combinations of texture features and its feature parameters to best describe the differences of crops and weeds. A non-linear classifier will then evaluate the chosen features to predict the label for each pixel.

2. Literature Review

Features are often used to describe the property of an image and encode it in numerical vectors for image classification problems. Classification models then use the encoded data to understand relationships and patterns within the encoded data to distinguish between the classes marked. In many classification tasks, Local Binary Pattern (LBP) is often used as a visual descriptor for texture. Some popular uses in LBPs are face detection [5], facial recognition [6], visual inspection [7], motion analysis [8] and texture analysis [9]. LBP encodes the local textures of each surrounding pixel as a feature vector. The literature has demonstrated that LBP to be a robust texture descriptor in many applications, thus its potential should be studied in smart agriculture applications.

Current technology makes the solution to weed control tasks in the field of computer vision possible. Different kinds of features are frequently used in literature as the input to a classifier model. Ahmed et al. [10] proposed the approach to using the supporting vector machine (SVM) classifier to discriminate against weeds. They chose shape, colour and moment as the feature vector in their work. The SVM model was for six plant species. This technique averaged 97.3 % for the resultant classification model. The writers note that the technique could be more robust if more pictures were used. They also noticed that the model was affected by image noise and plant holes.

Hemming [11] used colour and biological morphology features for carrot, cabbage and weed. The classifier is based on weighted fuzzy classification. Colour features were shown to distinguish plants from the soil, and the classifier shows an improvement in classification accuracy when using colour features in addition to morphology features. Their technique has led to a classification rate of 51-95 %, which varies based on weed density. They have observed that this method is prone to occlusion of the leaves.

McCool et al. [12] suggested using lightweight DCNN models for pixel-level crop-weed classification tasks. The authors suggested that the classification model be trained in three phases. The first phase utilizes a pre-trained DCNN model that delivers cutting-edge performance at the cost of high computational requirements. This model is not suited for mobile automated field robotics as the frame rate is relatively low at 0.12 frames per second (fps). In the second stage, therefore, the pre-trained DCNN model is used as a teacher model to form a lightweight DCNN student model. This technique is known as model compression resulting in reduced accuracy but significantly reduced complexity. The third phase includes various lightweight designs coupled as a mix of models resulting in greater classification performance. The classification accuracy of the mixture of lightweight models is 90 %. The research discovered that while cutting-edge DCNN model could attain 93.9 % ranking accuracy, it could only operate low frame rates. Whereas, the lightweight DCNNs combination was able to compute with fewer
parameters at a faster framerate of about 1.07 to 1.83 fps.

Haug et al. [13] suggested a less complex design for plant classification without the use of segmentation techniques. Crops and weeds often grow in close proximity and overlap are prevalent. The authors, therefore, suggested the use of a Random Forest classifier to classify pixels instead of segments. Pixel classification eliminates difficulties combined with segmentation techniques like overlapping and challenging background. The classifier predicts the position of the spatial pixel based on the characteristics of overlapping neighbourhood pixels. Markov Random Field was used to merge similar predictions. The model achieved an average classification accuracy of 93.8%.

Current literature has outlined problems related to plant holes, occluded leaves, and high complexity models. Therefore, we suggest a model that seeks to reduce certain problems by combining an effective texture descriptor, LBP, in an ensemble classification model with the evolving feature selection in order to classify each pixel without segmentation as inspired by Haug et al. [13].

3. Classification Model

3.1. Image Dataset

The images used are from a crop/weed field image dataset by Haug et al. [1]. The dataset includes 60 images of top-down images of row crops on organic carrot farms. The images were captured with a multi-spectral camera, which contains red and near-infrared visual information, mapped to 3 colour channels. Channel 1 represents red, while channel 2 represents near-infrared, and channel 3 is a repetition of channel 1. A sample image is shown in Figure 1. The complicated spatial composition of overlapping and close growth of the weeds and crops are also apparent in the figure. The data set provides labelled images, in which a human expert annotated each pixel. The labels are "1" for weed and "0" for crops, which simplifies the task to a binary classification.

3.2. Background Removal

The original image includes soil and other visual information that is not useful in discriminating weeds. Thus, removing non-contributing image information is essential. The image includes red and near-infrared channels, so a Normalized Difference Vegetation Index (NDVI) was adopted to calculate the differences between plant and soil variations in spectral reflection. The NDVI
used to obtain soil vegetation mask is calculated, as shown in Equation 1. Besides separating
the plant from the soil, the method also simplifies the image to a single-channel greyscale image.

\[
NDVI = \frac{NIR - Red}{NIR + Red}
\]  

(1)

Most background information is filtered out by NDVI. Otsu’s thresholding method[14] further
filters out the remaining background pixels, which produces a binary vegetation mask which
represents the vegetation pixels in 1 (white) and the background as 0 (black). The binary mask
is used to filter the soil pixels on the NDVI image, resulting in a picture shown in Figure 2
where in NDVI space, only the vegetation pixels are present. The resulting image, therefore,
includes only contributing plant information in a single channel NDVI space. Reduction to a
single channel is essential because the feature extractor requires it.

3.3. Feature Extraction

The image is divided into sub-images, which contains plant pixels. This paper will use the notion
of the image window to represent the sub-images. Every image window, as shown in Figure 3
comprises of the same pixel length and height where the centre pixel is the label class. The
image window size is determined in Equation 2.

\[
W_{\text{length}} = (n \times 2) + 1
\]  

(2)

Where n represents the half length of the image window.

From the image window, a rotation-invariant Local Binary Pattern (LBP) [15] is used as a
feature extractor encoding local pixel texture patterns in the to create a feature vector. LBP
compares its neighbouring pixels to a reference pixel situated in the middle of the image window
to form a binary vector. LBP can be defined as follows in terms of equations:

\[
LBP_{P,R}(x_{c}, y_{c}) = \sum_{p=0}^{p-1} s(i_{p} - i_{c}) \times 2^{p}
\]  

(3)
Figure 3. The image window representation where L is the label to labeled middle pixel and n is the number of pixels.

\[ s(v) = \begin{cases} 
1 & : v \geq 0 \\
0 & : v < 0
\end{cases} \]  \hspace{1cm} (4)

From Equation 3, \( P \) is the number of neighbours, and \( R \) is the radius from the centre. The subscript \( c \) denotes the centre and \( p \) denotes the current pixel neighbour. \( i_c \) represents the pixel value of the center reference pixel \( i_p \), which is located at \((x_c, y_c)\). The LBP process requires the image window to be divided into cells. Equation 5 calculates the number of cells.

\[ \text{numCells} = \text{floor} \left( \frac{W_{\text{length}}}{\text{CellSize}} \right)^2 \]  \hspace{1cm} (5)

Where \( W_{\text{length}} \) represents the length of the image window and \( \text{CellSize} \) represents the chosen cell size.

The LBP extraction method can be summarized as follows:

(i) Locate the plant pixel.
(ii) The plant pixel will correspond to a label based on the annotated image.
(iii) The plant pixel will be the centre of the image window.
(iv) Divide the image window into cells.
(v) The middle pixel is set as the reference, compare every other N neighbouring pixels through an R radius circular manner.
(vi) For pixels which are larger than the middle will be valued as 1, otherwise 0 in a continuous sequence given by Equation 4.
(vii) This will form a binary sequence of N number of neighbours which is converted to decimal given by Equation 3.
(viii) From the obtained number, a histogram is computed over each cell, and the frequency is added to each bin.
(ix) L2 normalization is applied to normalize the histogram.
(x) The resulting feature vector represents the normalized summation of the histogram of each cell for the particular image window.
(xi) Stride and repeat for all plant pixels.
Table 1. Table of parameters

| LBP Parameters         | Lower Boundary | Upper Boundary |
|------------------------|----------------|----------------|
| Image Window Size (S)  | 3x3            | 101x101        |
| Radius (R)             | 1              | 5              |
| Number of Neighbours (N)| 8              | 16             |

3.4. Feature Selection
Feature selection is a significant component in the model where most contributing features are chosen to allow the classifier to learn the differences between the two classes. As the search space is large, genetic algorithm is used to search for the ideal feature parameters in the ensemble. In parameter tuning, the LBP extraction method enables tuning of window size S, reference pixel R radius, and number of neighbours N. The cell size is set to 1. A list of LBP features with varying S, R, N parameters were generated for timely optimization. The image windows S size ranges from 3x3 to 101x101 pixels with 2 pixel increments for each dimension. Zero paddings are applied to the image window for out of bound regions. The radius of the neighbouring pixels R range from 1 to 5. The number of neighbours N ranges from 8-16 determines the number of pixels to be included for the binary pattern calculation. The LBP features produced are invariant to rotation. Therefore, all feasible combinations are produced for feature selection with the stated parameter boundaries.

Neural networks have been chosen as the classifier of choice since no prior knowledge of data processes is necessary, and they can establish relationships from the features given. The neural network classification structure is comprised of a model of three layers, namely, an input layer, 2 hidden layers and an output layer. The input layer provides one type of LBP feature from the GA selection. Each of the two hidden layers has 10 interconnected neurons. Finally, the output layer consists of a 2-class output, although it is a binary classification task, a 2-class output is needed to generate certainty scores of each class. At the output layer, tangent sigmoid was used as the activation function, and the model loss uses the mean squared error of the classification accuracy. This structure forms a simple NN unit that is used in the ensemble.

The ensemble consists of several simple stacked NN units. Contrary to the traditional stacked neural networks, this ensemble model does not supply all simple NN units with the same input. The output of all simple NN units is combined with a sum average of each output class. The highest value would determine the selection of the predicted class.

3.5. Training of Model
3.6. Neural Network Classifier
The divisions for train and test are 40 images and 20 images respectively. The training consists of 2 stages, which are feature selection by GA and NN training through backpropagation. The full ensemble classification model is shown in Figure 4. For the first stage, genetic algorithm (GA) optimization was used to find the optimal feature parameter combinations. The GA starts with a randomly generated population with varying feature parameters. Each generation will consist of a GA population size of 20 individuals where each individual will its NN units trained, and its fitness evaluated. Fitness is defined as the classification accuracy given by the ensemble NN units. Stopping criteria will be met when there is no improvement of over 50 generations. The features selected by GA will be the inputs of each NN unit and the outputs are the 2 class certainty scores (i.e. crop or weed). The 2 class certainty scores from each NN units are combined with all the outputs in the ensemble model through sum average. The resultant score represents the classifier accuracy. This classification accuracy is then used by the GA to determine the best combinations of parameters and features to be used.
4. Results and Discussion

Evaluation metrics provide a standard for discussing and comparing each model’s strengths and weaknesses. The standard metrics are accuracy, accuracy, recall and F1-score from Equations (6)-(9) respectively. Accuracy indicates how well weeds and plants can be properly classified in the classification model. Precision refers to how well classification of the weed class can be determined without a crop being misclassified as a weed. The precision metric is particularly essential in order to avoid accidental crop spraying. The recall metric, or sometimes called sensitivity, will determine how much weeds are sprayed considering the actual number of weeds in the field. The F1 score is the weighted average of recall and precision. The F1 metric takes into account the imbalanced dataset and measures recall and precision as one measure.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{7}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{8}
\]

\[
\text{F1 Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \tag{9}
\]

Where \(T\) is "True", \(F\) is "False", \(P\) is "Positive", and \(N\) is "Negative".

Our proposed method used a combination of simple NN units in an ensemble. The number of simple NN units in the ensemble are 5, 10, and 15. From the results shown in Table 2, a simple NN unit with a single LBP feature at best could only achieve an accuracy of 80.8 %. The LBP as a feature extractor shows potential in weed classification tasks as even the single
Table 2. Comparison of classification scores from literature studies on CWFID [1] dataset

| Methods                          | Accuracy | Precision | Recall | F1-score |
|----------------------------------|----------|-----------|--------|----------|
| Random Forest Tree [1]           | 85.9 %   | 79.6 %    | 80.8 % | 80.2 %   |
| PWDS [16]                        | 85.8 %   | 91.1 %    | 86.8 % | 88.8 %   |
| ANN with 15 units in ensemble    | 83.5 %   | 83.5 %    | 97.8 % | 90.1 %   |
| ANN with 10 units in ensemble    | 83.0 %   | 82.8 %    | 98.2 % | 89.9 %   |
| ANN with 5 units in ensemble     | 82.3 %   | 84.0 %    | 95.8 % | 89.5 %   |
| ANN without ensemble             | 80.8 %   | 84.5 %    | 91.8 % | 88.0 %   |

type of LBP feature combination paired with NN could achieve relatively high precision, recall and F1-score when compared to Random Forest Tree [1]. In terms of classification accuracy, our best ensemble model with 15 NN units at 85.9 % did not surpass in terms of accuracy when compared to Random Forest Tree [1] of 85.9 % and PWDS [16] of 85.8 %. The ensemble model was able to increase the overall accuracy by 2.7 % from 80.8 % to 83.5 % with 15 units in the ensemble. Even though a higher number of ensemble increase the overall accuracy, it can be seen that the precision maintained relatively the same, averaging around 83 %. The increase in accuracy is due to the increasing recall metric. Recall, as mentioned previously, is the sensitivity or hit rate. This means that the classifier is very good at determining the targeted weed class, but as reflected in accuracy, it does not perform well in classifying the outlier crop class.

The limitation of our current proposed work is at the method by which the ensemble outputs are combined. As the predicted classes are a sum average of all the NN units, the ensemble model is constrained to equal weights between each class and each NN unit. This would limit the flexibility of the ensemble model as it would not be able to allocate more weights to the appropriate contributing class and unit. In other words, the current ensemble model weighs all predictions from each output class of the NN units the same and does not boost the weak class well.

Our method achieved the highest F1-score of 90.1 % from 15 units in the ensemble when compared with Random Forest Tree [1] and PWDS [16]. This indicates that the classifier performs quite well in regards to a well-balanced precision and recall. Thus, improvements could be made towards the classification of the crop class to improve the classification accuracy further.

5. Conclusion
This paper demonstrated that LBP could potentially be used to identify weeds and crops in the agriculture field. LBP, as a feature used in this weed classification task, appears to perform quite well in classifying weeds as reflected by the recall score. Improvements could be further made to the structure of the ensemble to improve the classification of the weaker crop class further. For future work, we would propose the use of a weighted sum average to the ensemble paired with an evolving optimizer such as particle swarm optimization (PSO) to find the optimal weights. We hypothesize that certain combinations of LBP feature parameters may be able to discriminate a particular class better. Thus, with a weighted ensemble, the classification of the weaker class as the weaker units could be boosted by the weights.

References
[1] Haug S and Ostermann J 2014 A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks European Conference on Computer Vision (Springer) pp 105–116
[2] Guzzella L, Pozzoni F and Giuliano G 2006 Environmental Pollution 142 344–353
[3] Pereira W E and Hostettler F D 1993 Environmental science & technology 27 1542–1552
[4] Landrigan P J and Benbrook C 2015 New England Journal of Medicine 373 693–695
[5] Jin H, Liu Q, Lu H and Tong X 2004 Face detection using improved lbp under bayesian framework Third International Conference on Image and Graphics (ICIG’04) (IEEE) pp 306–309
[6] Liu L, Fieguth P, Zhao G, Pietikäinen M and Hu D 2016 Information Sciences 358 56–72
[7] Tajeripour F, Kabir E and Sheikh A 2008 EURASIP Journal on Advances in Signal Processing 2008 60
[8] Zhao G and Pietikainen M 2006 Local binary pattern descriptors for dynamic texture recognition 18th International Conference on Pattern Recognition (ICPR’06) vol 2 (IEEE) pp 211–214
[9] Mäenpää T and Pietikäinen M 2005 Texture analysis with local binary patterns Handbook of pattern recognition and computer vision (World Scientific) pp 197–216
[10] Ahmed F, Al-Mamun H A, Bari A H, Hossain E and Kwan P 2012 Crop Protection 40 98–104
[11] Hemming J and Rath T 2001 Journal of agricultural engineering research 78 233–243
[12] McCool C, Perez T and Upcroft B 2017 IEEE Robotics and Automation Letters 2 1344–1351
[13] Haug S, Michaels A, Biber P and Ostermann J 2014 Plant classification system for crop/weed discrimination without segmentation Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on (IEEE) pp 1142–1149
[14] Otsu N 1979 IEEE transactions on systems, man, and cybernetics 9 62–66
[15] Ojala T, Pietikäinen M and Mäenpää T 2002 IEEE Transactions on Pattern Analysis & Machine Intelligence 24 971–987
[16] Umamaheswari S, Arjun R and Meganathan D 2018 Weed detection in farm crops using parallel image processing 2018 Conference on Information and Communication Technology (CICT) (IEEE) pp 1–4