Flue-Cured Tobacco Leaves Classification: A Generalized Approach Using Deep Convolutional Neural Networks

Siva Krishna Dasari, Koteswara Rao Chintada and Muralidhar Patruni

Abstract In this paper, a solution is defined based on convolutional neural networks (CNN) for the grading of flue-cured tobacco leaves. A performance analysis of CNN on 120 samples of cured tobacco leaves is reduced from 1450 × 1680 Red-Green-Blue (RGB) to 256 × 256, consisting 16, 32 and 64 feature kernels for hidden layers respectively. The neural network comprised of four hidden layers where the performance of convolution and pooling on first three hidden layers and fourth layer a fully connected as in regular neural networks. Max pooling technique (MPT) is used in the proposed model to reduce the size. Classification is done on three major classes' namely class-1, class-2 and class-3 for obtaining global efficiency of 85.10% on the test set consisting about fifteen images of each cluster. A comparative study is performed on the results from the proposed model with existing models, state of the art models on tobacco leaf classification.

Keywords Convolutional neural network · Flue-cured tobacco
Max pooling · Multi-layer perceptron · Tobacco leaves classification system

1 Introduction

Tobacco is one of the most successful commercial crops cultivated on this planet. China, India, Brazil and USA are the major producers of tobacco worldwide and these four nations alone contribute around 86% of the global production. India is the top two contributors to the global tobacco production and it's estimated that around 750 M kgs of tobacco is being produced in the area around 0.45 M hectares [1]. Harvested tobacco will undergo a process called curing for further stages of refinement.

There are few methods existing on grading of FCT leaves [2], proposed a method based on nearest neighbors which uses the mean of hue and chroma as color features and factors such as the ratio of leaf length with respect to the width and shape as features [3]. A fuzzy classification system which uses color, texture and

© The Author(s) 2018

S. Gurumoorthy et al., *Cognitive Science and Artificial Intelligence*,
SpringerBriefs in Forensic and Medical Bioinformatics,
https://doi.org/10.1007/978-981-10-6698-6_2

shape as features to design the model and grade the tobacco leaves using Support Vector Machine (SVM) to classify FCT leaves. There is a model proposed by [4] which uses symbolic data and uses Munsell coloring system to extract average and standard deviation of hue as color features.

In this work, a solution is proposed based on CNN for the grading of tobacco. As CNN have much fewer connections and parameters compared to standard feed forward neural networks with similar sized layers, it's easy to train them. They can handle high dimensional inputs cleverly. The feature selection is done unsurprisingly without explicitly stating the features with the help of filter kernels. They have many advantages as they adjust the filter weights to recognize features such as color, shape and size.

2 Convolutional Neural Networks

2.1 Convolutional Layer (CL)

In CNN each neuron doesn't sense the total image as in MLP which is fully connected and process all the pixels of the image. Every neuron gets connections from a portion of the image (local to particular neuron) which makes CNN faster to train (Fig. 1). The connectivity along the depth axis is same as the depth of the input volume (number of color channels).

2.2 Pooling Layer

Pooling Layer reduces the size of the input volume there by reducing the number of parameters and computation. It controls over-fitting of the data. While resizing, it operates on each depth slice independently. There are different methods of doing pooling, the proposed model uses MAX Pooling (MP) [5]. If a p × p filter is used for MP, then it take one maximum value out of all p2 values on the input volume. In pooling depth dimension remains same (Fig. 2).

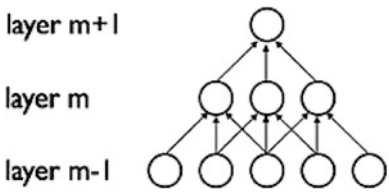


Fig. 1 Sparse connectivity of CNN

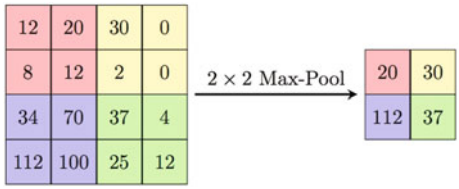


Fig. 2 MAX pooling on a depth slice

2.3 Back Propagation on Conv Layer

For an error function E and having the current error value of the layer m, calculate the error values of the previous layer and gradients for each weight. The error for previous layer is partial of E with respect to neuron output $\frac{\partial E}{\partial y_{ij}^m}$. Here is the formula for calculating the gradient for each weight by using chain rule [6].

$$\frac{\partial E}{\partial \omega_{ab}} = \sum_{i=0}^{s-t} \sum_{j=0}^{s-t} \frac{\partial y}{\partial x_{ij}^m} \frac{\partial x_{ij}^m}{\partial \omega_{ab}}$$

$$= \sum_{i=0}^{s-t} \sum_{j=0}^{s-t} \frac{\partial E}{\partial x_{ij}^m} y_{(i+a)(j+b)}^{m-1}$$

$$\frac{\partial E}{\partial x_{ij}^m} = \frac{\partial E}{\partial y_{ij}^m} \cdot \frac{\partial y_{ij}^m}{\partial x_{ij}^m}$$

$$= \frac{\partial E}{\partial y_{ij}^m} \cdot \frac{\partial}{\partial x_{ij}^m} \left(\sigma \left(x_{ij}^m \right) \right)$$

$$= \frac{\partial E}{\partial y_{ij}^m} \cdot \sigma^1 \left(x_{ij}^m \right)$$

The error at current layer $\frac{\partial E}{\partial y_{ij}^m}$ it's straight forward to compute $\frac{\partial E}{\partial x_{ij}^m}$ at present layer by knowing the first derivative of the activation function, i.e. $\sigma^1 \left(x_{ij}^m \right)$.

$$\frac{\partial E}{\partial \omega_{ab}} = \sum_{i=0}^{s-t} \sum_{j=0}^{s-t} \frac{\partial y}{\partial x_{ij}^m} \frac{\partial x_{ij}^m}{\partial \omega_{ab}}$$

$$= \sum_{i=0}^{s-t} \sum_{j=0}^{s-t} \frac{\partial E}{\partial x_{ij}^m} y_{(i+a)(j+b)}^{m-1}$$

Therefore, the error at the current layer computes the gradient with respect to weights on the layer, using chain rule to back propagate.

3 Tobacco Leaves Classification System

3.1 Convnet Architecture and Model

Tobacco Leaves classification system (TLCS) is based on deep CNN. Torch 7 Frame work in deep learning for building the system [7]. The network consists a total of nine layers. The full architecture of this model is depicted as shown in the Fig. 3 with the exclusion of Input and Output layers. There are total seven layers, out of which three convolutional, three pooling and one fully connected layer.

The design depth of the network is as follows:

Input Layer which is followed by a CL Conv1, Pooling Layer Pool1, CL Conv2, Pooling Layer Pool2, CL Conv3, Pooling Layer Pool3, a Rectified Linear Unit (ReLU) unit sandwiched between the Convolutional and Pooling Layers as a fully CL and the Output Layer.

In this model the following hyper parameters filter size or receptive field of 5, Pooling filter of size 2, with a stride of 1 and non Zero-padding applied. The CL Conv1 which filters the 256 × 256 × 3 input image with 16 Convolutional Kernels each of size 5 × 5. Total number of neurons in this layer is 252 × 252 × 16.

The Pooling Layer Pool1 which filters the input volume from Conv1 16 Pooling Kernels each of size 2 × 2.

Total number of neurons in this Layer is 126 × 126 × 16.

There is ReLU unit which works like an activation function applied on the output volume of the CL [8]. This outputs the same dimension as of Input volume with activation function applied on each pixel of the Input volume.

The output volume of Pool1 is fed into Conv2 Layer which has 32 filters, each filter operates on some portion of 126 × 126 × 16 input volume.

This Layer has a total of 122 × 122 × 32 neurons. Pool_2 Layer which filters the output of ReLU with 32 filters has a total of 61 × 61 × 32 neurons.

Conv3 Layer which filters the output of Pool2 functions with 64 filters has a total of 57 × 57 × 67 neurons. Pool3 Layer which filters the output of ReLU functions with 64 filters has a total of 28 × 28 × 64 neurons. The output volume of Pool3 is

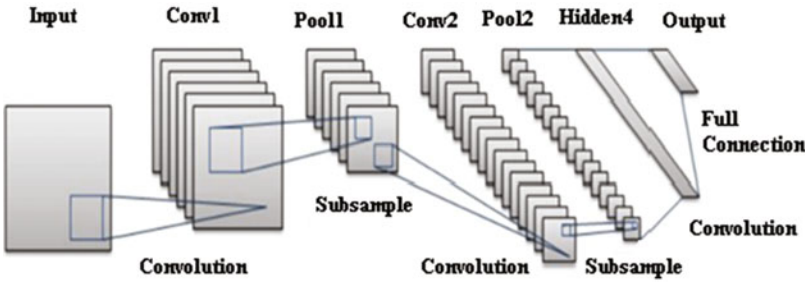


Fig. 3 Convolutional neural network architecture of tobacco leaves classifier

being fed into a fully connected layer having 128 neurons. All the outputs of fully connected layers are piped into the output layer which is a three way classifier with three neurons.

3.2 Details of Learning

The model is trained using gradient descent with a batch size of 1, weight decay of 0.0001 and with a momentum of 0.5.

Initialized learning rate is 0.01.

The generated formula for weights becomes

$$\vartheta_{i+1} = 0.5V_i - 0.0001 \cdot \in \cdot \omega_i - \in \cdot < \frac{\partial \mathrm{L}}{\partial \omega} | \omega_{\mathrm{i}} >_{\mathrm{D_i}}$$

$$\omega_{i+1} = \omega_i + \vartheta_{i+1}$$

where

i Iteration index
$\in$ Learning rate
ϑ Momentum variable

$< \frac{\partial L}{\partial \omega} | \omega_i >_{D_i}$: is the average over ith batch $\mathrm{D_i}$ of the derivative of the objective with respective ω valuated at ω_i [9].

The network is trained on 120 images and It took approximately 4 h for 300 epochs on a machine with 2.3 GHz Intel Quad Core Processor and 3 GB of Main Memory.

Figure 4 illustrates the training accuracy over the number of epochs.

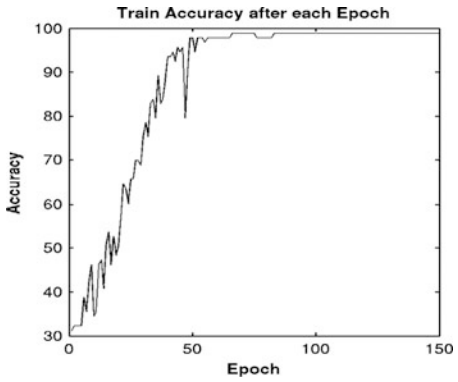


Fig. 4 Training accuracy over the number of epochs

4 Resultant Outputs

The proposed model on the test data set consisting of 47 image samples belonging to three different classes i.e. (class-1: 15 samples, class-2: 17 samples class-3: 15 samples). It took on average about 68.46 ms to test or classify each test sample.

A global correctness of 85.10% on 47 samples. Here is the Confusion matrix for the three classes.

$$\begin{matrix} & \text{class1} & \text{class2} & \text{class3} \\ \begin{matrix} \\ \\ \end{matrix} & \begin{pmatrix} 13 & 0 & 2 \\ 0 & 14 & 3 \\ 0 & 2 & 13 \end{pmatrix} \end{matrix}$$

Out of 15 samples of class-1, two samples are incorrectly classified into class-3. Out of 17 samples of class-2, 3 samples are incorrectly classified into class-3 from the 15 samples of class-3 two are incorrectly classified into class-2. Here is the bar chart depicting in Fig. 5.

For the test sample plotted a graph for the percentage accuracy with respect to number of epochs. At around 100 epochs the accuracy get saturated and have minimal fluctuations after that. This is represented in the Fig. 6.

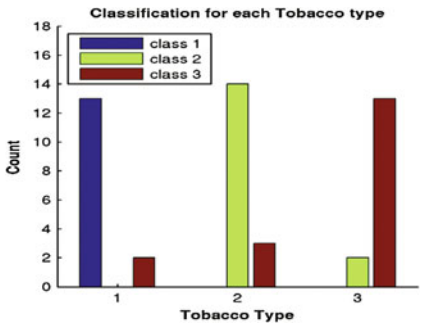


Fig. 5 Test sample correctness

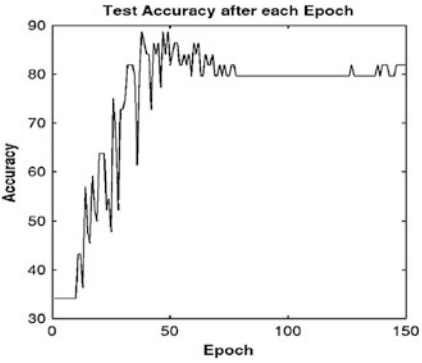


Fig. 6 Accuracy plot for a number of epochs

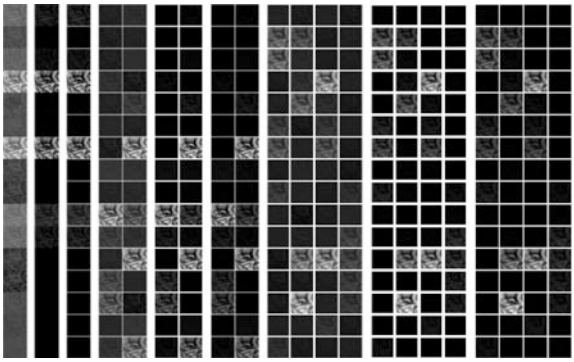


Fig. 7 A sample input volume as it passes through the layers

Figure 7 shows the outputs from different layers as the test sample passes through each layer (Conv Layer, ReLU Unit, Pooling Layer).

Columns-1, 2, 3 are convolutional, ReLU and Pooling Layers respectively with 16 filters. Columns-4, 5, 6 having 32 filters each and Columns-7, 8, 9 having 64 filters each.

5 Compartive Study

In the proposed system, the model takes 120 images for training sample. The existing models trained on more number of samples than considered in the proposed model and did more quality samples taken with better equipment.

Description: Table 1 shows that the proposed system produced 85.10% accuracy measure, though it is less in measure the concept of image enhancing in this work is scheduled with 120 as a size which surprisingly gave good result by considering back propagation.

Table 1 Comparison of different models with the proposed model

Title	Sample size	Classification technique	Accuracy (%)
A trainable grading system for tobacco leaves [2]	110	NNC	64
Proposed model (tobacco leave classification system using max pool technique)	120	DCNN	85.10
Recognition of the part of growth of flue cured tobacco leaves on support vector machine [10]	1712	SVM	86.62
Min-max representation of features for grading cured Tobacco leaves [4]	887	SC	87.18

6 Conclusion

Proposed TLCS performance even on limited training samples (image data set) compared to existing models. The proposed model achieved a global correctness of 85.10%. On application, the state of art image classification technique CCN for the model. This can be applied on samples in which it's hard to describe the features. As disclosed in Table 1, though the performance measure of the proposed model is next to [10]. It is clearly understood that the sampling in the above references handled huge clusters of image data whereas in the proposed model, it is considered strongly for 120 images. Here, the max pooling technique had shown its effect in abundant reduction of image size by reducing the unwanted kernel features of the respective hidden layers. The CL itself produced more betterment results compared with the three individual hidden layers. The major difference between the other models and the proposed model is that, the secure data of hidden layer could be accessed at any levels of all the algorithms whereas in this study it is clearly informed that there is no passage of data transfer between CL and hidden layers. This study gives us the parametric results as 85.10% which is accurate and appropriate both technically and theoretically as per the samples of tobacco data collected.

7 Future Scope

This model is further extended for application among larger training sample set to achieve better results. The result obtained through deep learning algorithms could be enhanced by using some optimization algorithms to the images which are considered. Consistent images i.e., with huge information and with less noise are considered to enrich the outputs at all levels.

References

1. CTRI-Rajahmundry. Tobacco in Indian Economy. http://www.ctri.org.in/fortobacco Economy.php 2015.
2. Zhang, J., et al. 1997. A Trainable Grading System for Tobacco Leaves. *Computers and Electronics in Agriculture* 16 (3): 231–244.
3. LeCun, Yann, et al. 1998. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* 86 (11): 2278–2324.
4. Guru, D.S., et al. 2011. Min-max Representation of Features for Grading Cured Tobacco Leaves. *Statistics and Applications* 9 (1&2): 15–29.
5. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, 1097–1105.

6. Boureau, Y-Lan, Jean Ponce, and Yann LeCun. 2010. A Theoretical Analysis of Feature Pooling in Visual Recognition. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), 111–118.
7. Prasad, V., T.S. Rao, and M. Babu. 2016. Thyroid Disease Diagnosis Via Hybrid Architecture Composing Rough Data Sets Theory and Machine Learning Algorithms. *Soft Computing* 20 (3): 1179–1189.
8. Prasad, V., Rao, T. S., and Reddy, P. P..2016. Improvised Prophecy Using Regularization Method of Machine Learning Algorithms on Medical Data. *Personalized Medicine Universe* 5, 32–40.
9. Lawrence, Steve, et al. 1997. Face Recognition: A Convolutional Neural-Network Approach. *Neural Networks, IEEE Transactions on* 8 (1): 98–113.
10. LeCun, Yann, et al. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1 (4): 541–551.



http://www.springer.com/978-981-10-6697-9

Cognitive Science and Artificial Intelligence
Advances and Applications
Gurumoorthy, S.; Rao, B.N.; Gao, X.-Z.
2018, VIII, 112 p. 50 illus., 28 illus. in color., Softcover
ISBN: 978-981-10-6697-9