Research Article

Deep Convolutional neural network (CNN) in tea leaf chlorophyll estimation: A new direction of modern tea farming in Assam, India

Utpal Barman*

Girijananda Chowdhury Institute of Management and Technology, Guwahati (Assam), India

E mail: utpalbelsor@gmail.com

How to Cite

Barman, U. (2021). Deep Convolutional neural network (CNN) in tea leaf chlorophyll estimation: A new direction of modern tea farming in Assam, India. *Journal of Applied and Natural Science*, 13(3), 1059 - 1064. https://doi.org/10.31018/jans.v13i3.2892

Abstract

This study presents the uprising of leaf chlorophyll estimation from traditional mechanical method to machine learning-based method. Earlier chlorophyll estimation techniques such as Spectrophotometer and Soil Plant Analysis Development (SPAD) meter demand cost, time, labour, skill, and expertise. A small-scale tea farmer may not afford these devices. The present study reports a low-cost digital method to predict the tea leaf chlorophyll using 1-D Convolutional Neural Network (1-D CNN). After capturing the tea leaf images using a digital camera in a natural light condition, a total of 12 different colour features were extracted from tea leaf images. A SPAD was used to estimate the original chlorophyll value of the tea leaves. The paper shows the correlation of original tea leaf chlorophyll with the extracted colour features of the tea leaf images. Apart from 1-D CNN, the Multiple Linear Regression (MLR) and K-Nearest Neighbor (KNN) were also applied to predict the tea leaf chlorophyll and compared their results with the 1-D CNN. The 1-D CNN model outperformed with an accuracy of 81.1%, Mean Absolute Error (MAE) of 3.01, and Root Mean Square Error (RMSE) of 4.18. The investigation system is very simple and cost-effective. It can be used in tea farming as a digital SPAD for faster and accurate leaf chlorophyll estimation in an easy way.

Keywords: Chlorophyll, Deep learning, Machine learning, SPAD

INTRODUCTION

Due to the use of the traditional mechanical farming system, the overall quality and productivity of farming are reducing(Barman and Choudhury, 2020). The same thing happens in the case of tea farming also. The traditional mechanical farming system takes more time and effort (Barman and Choudhury, 2019). This serious problem can be solved using proper farming and modern technology. Though modern technologies help farmers to enhance the quality and productivity of farming, a rural small-scale tea farmer in Assam may not afford the advanced technologies due to its high costs and need for expertness. The urban tea farmers often avail themselves of the facilities of modern technologies, laboratories, and expert training for better productivity, but the rural tea farmers often face difficulties due to the non-availability of modern farming technologies, laboratories, and experts. A good amount of plant photosynthesis is one of the important factors of tea farming and the causes of tea health issues. Expert's help

was reported and authenticated, but they may not be available in all areas. The Soil Plant Analysis Development (SAPD) meter and spectrophotometer are used for leaves chlorophyll estimation, but these are very costly and time-consuming (Agarwal and Gupta, 2018; Dey *et al*., 2016; Vesali *et al*., 2017). Small-scale tea farmers may not afford these devices. To overcome the mentioned issues, the previous investigators used different machine learning techniques such as Linear Regression (LR) and Artificial Neural Network (ANN) (Dey *et al*., 2016; Mohan and Gupta, 2019; Vesali *et al*., 2017), Stacked Sparse Autoencoder(Choi *et al*., 2019), Deep Neural Network(Choi *et al*., 2019) in their studies for the chlorophyll estimation of rice(Mohan and Gupta, 2019), pomegranate(Peng and Wang, 2019), spinach (Agarwal and Gupta, 2018), betel(Dey *et al*., 2016), etc. But the use of Convolutional Neural Network (CNN) for the leaf chlorophyll is not reported till now. Based on the aforesaid issues, the present research objective was formulated to estimate tea leaf chlorophyll by finding the correlation of tea leaf colour with the original

Fig. 1. *Tea leaf images and SPAD calibration*

chlorophyll value of the tea leaves using a 1-D CNN.

MATERIALS AND METHODS

Study area, image acquisition, and SPAD estimation

The study was conducted in Assam, a land of tea gardens, by collecting tea leaves from the various tea gardens of Assam. We plucked the immature tea leaves from the tree and simultaneously kept them ready for image acquisition using a digital camera. A Nikon DSLR camera with auto mode camera settings was used to captured 750 tea leaf images in sunlight, placing the leaves in a white paper after being plucked (Fig. 1). Tea images were in 4000x6000 dimensions with a 72-dpi level and jpg format. In parallel, the SPAD estimation was done for each tea leaf after the device calibration (Fig.1). The 750 SPAD estimation was saved in a .csv file and transferred to the computer system along with the tea leaf images for the pre-processing, feature extractions, and predictions (Fig. 2).

Image pre-processing and feature extraction

Due to the high dimension of the tea leaf images, the processing time of the images for feature extraction and estimation might take more time. To reduce the pro-cessing time, the images were resized. Since the paper aims to estimate tea chlorophyll in natural light, no other pre-processing techniques were applied to the images. Based on the earlier research output(Barman and Choudhury, 2020), the colour features of the tea leaf images were extracted to correlate with the chlorophyll of the tea leaves. Here, two colour models were used to extract the features of the tea leaves, including the means of RGB, the standard deviations of RGB, the standard deviations of HSV, kurtosis, skewness, the variance of each tea leaf using the equations mentioned in Table 1.

The average feature of the leaves is shown in Fig. 3. The variance of a pixel from its mean was more in tea leaf images which affect the overall training and estimation. To overcome this issue, the features were standardized by removing the mean of the feature and scaling the features to unit variance. After all this process, the tea leaf dataset was created with a size of 750x12. The 750 denoted the total numbers of samples and the 12 denoted the numbers of features. The features were saved in .csv file format and kept ready for regression using the 1-D CNN.

1-D CNN model for tea leaf chlorophyll estimation

The Convolutional Neural Network (CNN) was used for classification problems (Barman *et al*., 2020). Here, we used CNN model for prediction. Before, applying the 1-D CNN model, the entire 750 leaf feature data was divided into three categories such as training, validation, and testing set (Table 2). A total of 480, 120, and 150 images were used for training, validation, and testing
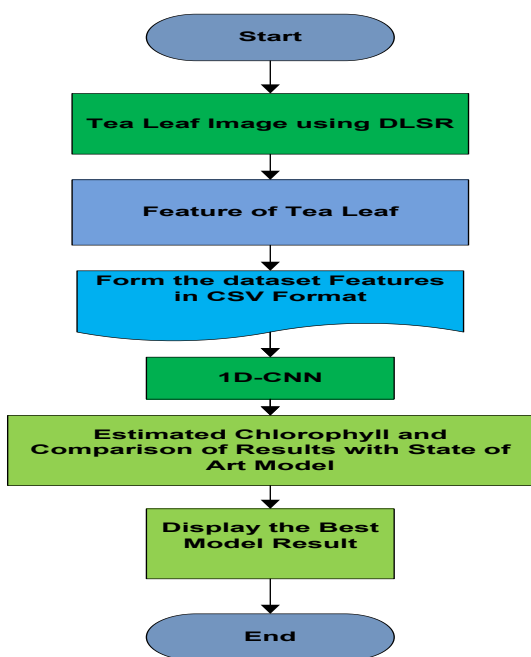

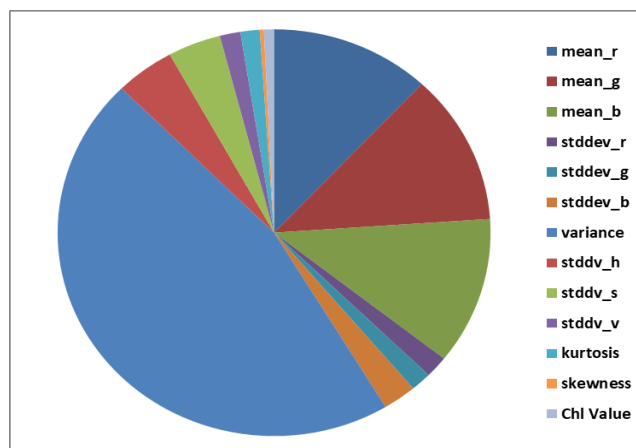**Fig. 2.** *Flow diagram of the proposed system*

Fig. 3. *Average features of 5 tea leaf*

set, respectively. The model consisted of two convolution layers without any pooling layer with 64-layer size, 2 kernel size, and relu activation function. After the convolutional, the model was flattened and simultaneously added two hidden layers with 64 and 32 hidden neurons and a relu activation function, respectively. Finally, the output layer with 1 hidden neuron was used to predict the tea chlorophyll. The model was trained for 10 epochs with a learning rate of 0.01and a batch size of 10. The architecture of the 1- D CNN model is presented in Fig. 4.

RESULTS AND DISCUSSION

Results of 1-D CNN
As mentioned in the above section, the 1-D CNN model was trained for 10 epochs with a batch size of 10. The

model's performance was evaluated using Mean Square Error (MSE) and Mean Absolute Error (MAE). The model is optimized with an ADAM optimizer to reduce the error in the model. The MAE and MSE of the epoch are presented below.

The above MSE and MAE values of the training, validation, and testing set claim no presence of overfitting and underfitting in the model due to the very little error difference among training, validation, and testing. Initially, the error of the 1-D CNN model was high, but with the increasing of epochs, the errors were also reduced. The validation performance was at epoch 10 with a validation MSE of 15.14, training MSE of 15.48, and testing MSE of 15.42. The MAE of the model in the three sets at epoch 10 were 2.91, 2.84, and 2.89. The overall results of the testing set are presented below (Table 3).

Table 1. Equations to extract different colour features for the tea leaf chlorophyll estimation

Parameter	Equations	Eq. No.
Mean	$\sum_{i,j=0}^{N-1} i.m_{i,j}$	(1)
Standard Deviation (SD)	$\sum_{i,j=0}^{N-1} (i - Mean)^2 m(i,j))^{1/2}$	(2)
Kurtosis	$\sum_{i,j=0}^{N-1} (i - Mean)^4 m(i,j)$	(3)
Skewness	$\sum_{i,j=0}^{N-1} (i - Mean)^3 m(i,j)$	(4)
Variance	$\sum_{i,j=0}^{N-1} m_{i,j}(i - \sigma)^2$	(5)

Table 2. MSE and MAE of the 1-D CNN

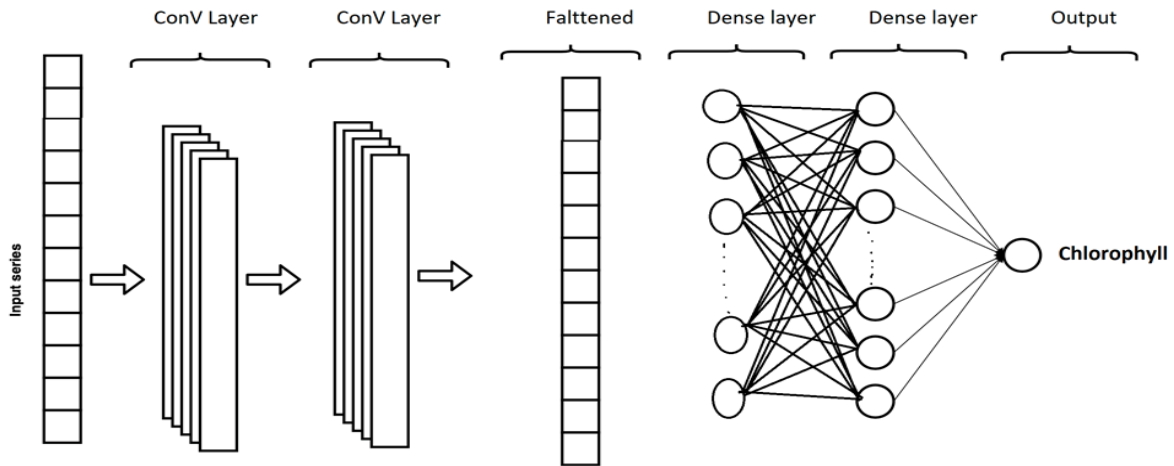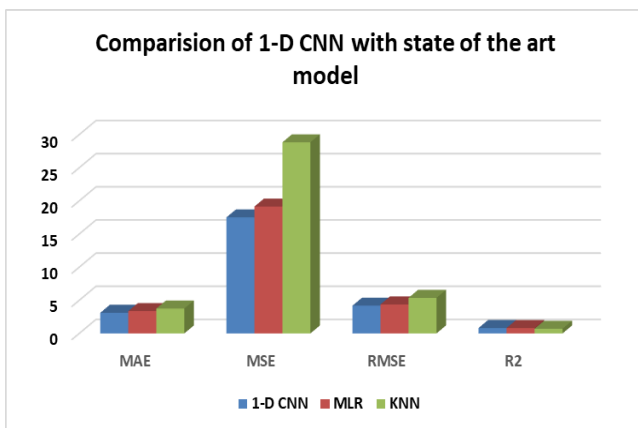| Epoch	Training MSE	Validation MSE	Testing MSE	Training MAE	Validation MAE	Testing MAE
1	19.14	20.14	21.03	3.42	3.52	3.59
2	17.21	19.21	18.75	3.38	3.35	3.56
3	17.13	19.72	20.45	3.14	3.24	3.45
4	13.92	18.42	19.87	3.08	3.12	3.57
5	16.42	16.87	15.48	3.08	3.14	3.43
6	16.40	15.48	16.48	3.06	3.01	3.25
7	16.42	15.24	15.48	3.05	2.97	3.28
8	15.89	15.14	15.87	3.05	2.98	3.29
9	15.86	15.85	16.23	2.95	3.00	2.97
10	15.48	15.14	15.42	2.91	2.84	2.89

Fig. 4. *Architecture of 1-D CNN*

Fig. 5. *Comparison of 1-D CNN with state of art model*

Comparison with state of art models

Along with the MSE and MAE, the 1-D CNN was also evaluated using the Root Mean Square Error (RMSE), and Coefficient Regression (R^2). The final MAE, MSE, RMSE, and R^2 of the 1-D CNN model were 3.01, 3.97, 15.97, and 0.811. Barman and Choudhury, (2020) used the LR and ANN model for citrus leaf chlorophyll prediction with an average accuracy of 80% and 99%, respectively. Vesali *et al*., (2015) also reported the LR and ANN for the prediction of corn leaf with a maximum accuracy of 97%. In both studies(Barman and Choudhury, 2020; Vesali *et al*., 2015), the contact imaging technique was used to capture leaf images, but we captured the images using a digital camera in natural environmental light conditions. Barman *et al*., (2021) also reported the use of multilayer ANN, MLR, Support Vector Regression (SVR), and K-Nearest Neighbor (KNN) models for tea leaf chlorophyll prediction. They reported MLR as the best model for tea leaf chlorophyll estimation with a maximum accuracy of 80%, MAE of 3.3, and an RMSE of 4.37. In their study, in the case of ANN and SVR, the colour correlations with the chlorophyll values were not acceptable. They mentioned only 50% accuracy in ANN and 53% accuracy in SVR with

an RMSE value of 5.54 and 6.67 in ANN and SVR, respectively. The LR and ANN model was also used for rice chlorophyll estimation with an accuracy of 77% and 79% for the LR and ANN(Mohan and Gupta, 2019). Choi *et al*. (2019) investigated the Deep Neural Network (DNN) with 4 convolutions layers and 2 maxpooling layers for chlorophyll estimation with a maximum accuracy of 95% and an MSE of 23. The use of Stacked Sparse Autoencoder and SVR were reported by Peng and Wang, (2019) for pomegranate leaf chlorophyll estimation with an accuracy of 80%. In the above studies, the researchers extracted RGB or HSV colour features from the leaves images and combined these features to make another colour feature such as G-R/G+R(Vesali *et al*., 2015), R+B(Barman and Choudhury, 2020), etc. This study extracted the original colour feature without combining RGB or HSV colour features from the tea leave images. A maximum accuracy of 81% with an MAE of 3.11 in the case of 1-D CNN was reported in the study. The present study also reported an MSE of 17.50 with an RMSE of 4.18 for the 1-D CNN regression. The results of 1-D CNN were compared with the MLR and KNN regression results in the case of tea leaf chlorophyll prediction (Fig. 5 and Table 4). Table 4 indicates that the MAE of the KNN (3.75) was more than the MLR (3.37) and 1-D CNN (3.11). The accuracy of the KNN was the minimum (70%) as compared to the MLR (80%) and 1-D CNN (81%). The MAE, MSE, RMSE, and R^2 values of the 1-D CNN reported itself as a best model for tea leaf pre-

Table 3. MAE, MSE, RMSE, and R^2 of the 1-D CNN

Parameters	Value
MAE	3.11
MSE	17.50
RMSE	4.18
R^2	0.81

diction (Table 4). The comparisons of the 1-D CNN models with results of the previous studies are presented in Table 5 which indicate that the proposed achieved more accuracy in the case of 1-D CNN and MLR than the results reported by Yadav *et al*. (2010).

Conclusion

In the present study, the 1-D CNN, along with MLR and KNN was used to estimate the chlorophyll value of the tea leaves. The results of the applied models proved that leaf colours could be used to predict the leaf chlorophyll. The images of this study were captured in natural environmental light conditions using DSLR that itself a novel of this study. The 1-D CNN model performed well as compared to the state-of-the-art model. This investigated model is simple and low-cost. It can replace the other traditional method for faster leaf chlorophyll estimation accurately and in easy manner.

ACKNOWLEDGEMENTS

Conflict of interest
The authors declare that they have no conflict of interest.

REFERENCES

1. Agarwal, A. & Gupta, S. D. (2018). Assessment of spinach seedling health status and chlorophyll content by multivariate data analysis and multiple linear regression of leaf

Table 4. Comparison of the results of the 1-D CNN, MLR, and KNN

Model / Parameter	MAE	MSE	RMSE	R^2
1-D CNN	3.11	17.5	4.18	0.81
MLR	3.37	19.133	4.37	0.80
KNN	3.75	28.83	5.36	0.70

Table 5. Comparative analysis leaf chlorophyll prediction

Author(s)	Plant	Features	Algorithm	Accuracy
(Agarwal and Gupta, 2018)	Spinach	L*a*b	LR	85%
(Mohan and Gupta, 2019)	Rice	RGB	LR	77%
(Mohan and Gupta, 2019)	Rice	RGB	ANN	79%
(Peng and Wang, 2019)	Pomegranate	Deep Features	Autoencoder	94.2%
(Vesali *et al*., 2017)	Corn	Red	LR	94%
(Vesali *et al*., 2017)	Corn	Hue	LR	96%
(Vesali *et al*., 2017)	Corn	G-R	LR	96%
(Vesali *et al*., 2017)	Corn	G-R/G+R	LR	97%
(Rigon *et al*., 2016)	Soybean	HSB and Lab colour	LR	96%
(Vesali *et al*., 2015)	corn	RGB	ANN	82%
(Riccardi *et al*., 2014)	Quinoa	RGB	LR	97%
(Ali *et al*., 2012)	Tomato	Logsig={G-R/3-B/3}/255	LR	96%
(Ali *et al*., 2012)	Lettuce	Logsig={G-R/3-B/3}/255	LR	89.6%
(Ali *et al*., 2012)	Broccoli	Logsig={G-R/3-B/3}/255	LR	91.4%
(Yadav *et al*., 2010)	Potato	RGB	LR	77%
Proposed Model (1-D CNN)	Tea Leaf	12 Original Colours	1-D CNN	81%
Proposed Model (MLR)	Tea Leaf	12 Original Colours	MLR	80%
Proposed Model (KNN)	Tea Leaf	12 Original Colours	KNN	70%

image features. *Computers and Electronics in Agriculture*, 152, 281–289.

2. Ali, M. M., Al-Ani, A., Eamus, D. & Tan, D. K. (2012). A new image processing based technique to determine chlorophyll in plants. *American-Eurasian Journal of Agricultural and Environmental Sciences*, 12(10), 1323–1328.

3. Barman, U. & Choudhury, R. D. (2019). Soil texture classification using multi class support vector machine. *Information Processing in Agriculture*, 7 (2), 318-322

4. Barman, U. & Choudhury, R. D. (2020). Smartphone image based digital chlorophyll meter to estimate the value of citrus leaves chlorophyll using Linear Regression, LMBP-ANN and SCGBP-ANN. *Journal of King Saud University-Computer and Information Sciences*, https://doi.org/10.1016/j.jksuci.2020.01.005

5. Barman, U., Choudhury, R. D., Sahu, D. & Barman, G. G. (2020). Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Computers and Electronics in Agriculture*, 177, 105661.

6. Barman, U., Sarmah, A., Sahu, D. & Barman, G. G. (2021). Estimation of Tea Leaf Chlorophyll Using MLR, ANN, SVR, and KNN in Natural Light Condition. *Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India*, 170, 287.

7. Choi, J.-H., Kim, J., Won, J. & Min, O. (2019). Modelling Chlorophyll-a Concentration using Deep Neural Networks considering Extreme Data Imbalance and Skewness. *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 631–634.

8. Dey, A. K., Sharma, M. & Meshram, M. R. (2016). An analysis of leaf chlorophyll measurement method using chlorophyll meter and image processing technique. *Procedia Comput. Sci*, 85, 286–292.

9. Mohan, P. J., & Gupta, S. D. (2019). Intelligent image analysis for retrieval of leaf chlorophyll content of rice from digital images of smartphone under natural light. *Photosynthetica*, *57*, 388–398.

10. Peng, Y. & Wang, Y. (2019). Prediction of the chlorophyll content in pomegranate leaves based on digital image processing technology and stacked sparse autoencoder. *International Journal of Food Properties*, 22(1), 1720–1732.

11. Riccardi, M., Mele, G., Pulvento, C., Lavini, A., d'Andria, R. & Jacobsen, S.-E. (2014). Non-destructive evaluation of chlorophyll content in quinoa and amaranth leaves by simple and multiple regression analysis of RGB image components. *Photosynthesis Research*, 120(3), 263–272.

12. Rigon, J. P. G., Capuani, S., Fernandes, D. M. & Guimarães, T. M. (2016). A novel method for the estimation of soybean chlorophyll content using a smartphone and image analysis. *Photosynthetica*, 54(4), 559–566.

13. Vesali, F., Omid, M., Kaleita, A. & Mobli, H. (2015). Development of an android app to estimate chlorophyll content of corn leaves based on contact imaging. *Computers and Electronics in Agriculture*, 116, 211–220.

14. Vesali, F., Omid, M., Mobli, H. & Kaleita, A. (2017). Feasibility of using smart phones to estimate chlorophyll content in corn plants. *Photosynthetica*, 55(4), 603–610.

15. Yadav, S. P., Ibaraki, Y., & Gupta, S. D. (2010). Estimation of the chlorophyll content of micropropagated potato plants using RGB based image analysis. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 100(2), 183–188.