Non-destructive detection of carotenoid content in tea leaves using a compact spectrometer

Yuta Tsuchiya, Rei Sonobe, Haruyuki Seki, Yuhei Hirono

**KEY WORDS:** 1D-CNN, DBN, reflectance, tea cultivars

**ABSTRACT**

Carotenoid content and composition are determined by the developmental stage, tissue type, and environmental stimuli and then play an important role in photosynthetic light-harvesting and stress reduction. Furthermore, carotenoid rich foods decrease the risk of developing certain types of cancer. Therefore, quantifying carotenoid contents has many applications in agriculture, ecology, and health science. However, traditional methods for quantifying carotenoids require destruction of samples, and they are time consuming and expensive. On the other hand, hyperspectral remote sensing offers some non-destructive methods that could be alternatives.

 In this study, we used regression models based on one-dimensional convolutional neural network (1D-CNN) or Deep Belief Nets (DBNs) to estimate carotenoid content from reflectance data from an inexpensive system based on a compact spectrometer (Colorcompass-LF, a total price for the proposed solution was approximately 1600 USD). The 1D-CNN -based model was better for this purpose: it achieved a ratio of performance to deviation of 1.43, a root mean square error of 1.08 μg/cm², and a coefficient of determination (R²) of 0.50.

**1.Introduction**

 Plant pigments are important elements of the plant photosynthetic apparatus and its nutritional functions. Among them, carotenoids not only contribute to the absorption of light energy for photosynthesis, but also play an important role in photoprotection of photosynthesis and dissipate excess light energy. Carotenoid content is known to be closely related to plant stress and photosynthetic capacity and is an indicator for assessing the physiological status of plants. In tea plants, carotenoid content is a heritable trait, and by monitoring carotenoid levels in various tea cultivars and varieties, breeders can identify plants with desirable carotenoid profiles and use this information for selective breeding(Y. Chen et al., 2023; Yang et al., 2021). This process can lead to the development of tea cultivars with improved nutritional quality, stress, and overall performance. Furthermore, carotenoid-rich teas are often preferred by consumers due to their enhanced appearance, flavour, and health benefits(Baker & Günther, 2004). Methods such as high-performance liquid chromatography have been used to measure carotenoid content. However, such methods are time-consuming and costly, require destructive samples of leaf tissue, and cannot capture changes in pigments over time. Remote sensing, on the other hand, is not only non-destructive and low-cost, but also allows measurement of large areas at a time. Hyperspectral remote sensing utilizing spectral reflectance characteristics has been proposed based on machine learning, spectral reflectance index, and radiative transfer models, but the commercial spectroradiometers used in previous studies are made overseas and have not been implemented in society due to cost and maintenance aspects. Recently, highly sensitive, cheap, and fingertip-sized spectrometers, such as the C12880MA-10 (Hamamatsu Photonics), have been released, and their potential should be evaluated. In this study, reflectance measurements were obtained from the Colorcompass-LF, which is based on the C12880MA-10. Algorithm choice is one of the important processes to obtain the estimation results with high accuracies from reflectance data. Deep learning-based algorithms have been successful in effectively expressing complex relationships, and their strong performance in the evaluation of vegetation properties has been reported. Furthermore, deep learning has become increasingly prevalent following the rapid development of big data and computing power in the past few years (Chu et al., 2022; Jiang et al., 2022).One-dimensional convolutional neural network (1D-CNN) is one of the most effective architectures based on deep learning and has been used to evaluate soil properties using Vis–NIR reflectance(Ng et al., 2019; Pullanagari et al., 2021). Deep belief nets (DBNs) also have a probabilistic generative architecture composed of multiple layers of stochastic latent variables (J. H. Chen et al., 2017) and have performed well in hyperspectral remote sensing (Y. Chen et al., 2014; Sonobe et al., 2020). In general, high-specification computers are required to generate regression models based on deep learning algorithms. Google Colaboratory is a free online cloud based Jupyter notebook environment that allows the generation of regression models based on graphics processing units. This server was used to generate regression models based on 1D-CNN for our proposed method of low-cost field-scale monitoring.

**2.Materials and Methods**

2.1. Measurements and Datasets

The experiments were conducted at the Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Shimada, Japan (Figure 1). The tea field comprised 39 ridges and a different cultivar was cultivated on each ridge, however, Yabukita was cultivated in the two ridges. We therefore collected samples from 38 tea cultivars. While most of the cultivars, including Sencha and Matcha, are grown for green tea, Sunrouge (the product of *Camellia taliensis* × *C. sinensis*) yields a pink tea, and Benifuuki, Benihikari, and Benihomare produce black tea.

We used a spectrometer with a complementary metal-oxide semiconductor (CMOS) sensor (C12880MA-10, Hamamatsu Photonics) and a shape-memory alloy (SMA)-SMA fibre patch cable (M25L05, Thorlabs) with a 0.22 numerical aperture to measure reflectance with a leaf clip (Figure 2, colorcompass-LF). On 10 May, 20 June, and 28 June, we collected 234 samples from the third leaf of the tea trees (six samples from each cultivar except Yabukita, for which 12 samples were collected).



**Figure 1. Aerial view of the tea field sampled in this study**.

****

**Figure 2.** Measurements of the reflectance data using the Colorcompass-LF.

Leaf discs were collected after the reflectance measurements were completed, and the absorbance of N, N-dimethylformamide extract was measured using a dual-beam scanning ultraviolet–visible spectrophotometer (UV-1280, Shimadzu, Kyoto, Japan). Wellburn’s method(Wellburn, 1994) was applied to quantify the carotenoid content based on absorption [1-3].

Car= (1000A480-1.12Chla-34.07Chlb)/245 (1)

Chla=12A663.8-3.11A646.8  　　 (2)

Chlb=20.78A646.8-4.88A663.8　　　　　 (3)

where A is the absorbance and the subscripts are the wavelengths (in nm). A stratified random sampling approach, which is a method of sampling that involves the division of all measurements into smaller sub-groups (strata), was applied. The strata were based on treatments.

The measurements were divided into two groups: a training dataset (75%) and test dataset (25%) following a previous study, and this approach was repeated one hundred times to ensure robust results.

2.2. Model Development

2.2.1 One-Dimensional Convolutional Neural Network (1D-CNN)

 CNN has been applied to automatically detect features of interest from the given data, and 1D-CNN can provide accurate results for 1D data; 1D-CNN has an input layer, hidden layers (convolutional, pooling, fully connected and normalization) and an output layer. Convolution was applied to the reflectance data to extract a feature map using a convolution filter, and then each unit in the convolutional layer was connected to local features in the feature map. After the convolution operation, a pooling layer was used for the dimensional reduction of the feature map, which effectively reduces the computational cost and minimises the overfitting of the network while preserving important information. In this study, the max-pooling technique and ReLU were applied. It was reported that 1D-CNN was effective to estimate the concentrations of the major and minor pigments from the reflectance and absorption coefficient spectral inputs The architecture was composed of 10 hidden layers that included four convolutional layers, four max-pooling layers, and two fully connected layers; two dropout rates, 0.4 and 0.2, were used following previous studies. The regression models based on 1D-CNN were generated using Google Colaboratory.

2.2.2 Deep Belief Nets (DBNs)

 DBNs consist of multi-layer unsupervised restricted Boltzmann machines and produce an optimum model in comparison to a model based on random weights for the weight initialisation of a deep neural network. DBNs can be effectively used to perform layer by-layer pre-training intended to initialise the training of a backpropagation algorithm. DBNs have been applied to extract vegetation properties, such as quality (chlorophyll-a content) and stress (chlorophyll-a: b), from hyperspectral data for improved tea tree management, and some pre-processing could be reduced. The initial configurations were the learning rate (0.1), the maximum iteration number of the pre-training dataset (100), the learning rate of the pre-training dataset (0.01), the maximum iteration number of the training dataset (100), and the batch data size (10) following previous studies. DBN regression was implemented using the “darch” package in R version 4.0.6

2.3. Statistical Criteria

 The model performance was evaluated using the ratio of performance to deviation (RPD, Equation (4) ), as RPD is a widely applied indicator with a clear definition (e.g., category A [RPD > 2.0], category B [1.4 ≤ RPD ≤ 2.0], and category C [RPD < 1.4]). RPD directly compares the index performance of different datasets and is used especially to examine robustness across different datasets. In addition to RPD, the root mean square error (RMSE, Equation (5)) and coefficient of determination (R², Equation (6)) were calculated using:

RPD=$ \frac{SD}{SEP}$ , (4)

 RMSE=$\sqrt{\frac{1}{n}\sum\_{i=0}^{n}\left(ŷi – yi\right)^{2} }$ , (5)

$R^{2}=1-\left(\frac{\sum\_{i=1}^{n}\left(yi - yˆi\right)^{2}}{\sum\_{i=1}^{n}\left(yi - ӯ\right)^{2}}\right)$, (6)

where SD is the standard deviation of the measurements, SEP is the standard error prediction, n is the number of samples, yi is the real value, ŷi is the estimated value, and y is the mean of the measurements. Chang et al. claimed that Category B can be improved by using different calibration strategies, but properties in Category C may not be reliably predicted. The sensitivity of spectral wavelengths was evaluated using the variance principle. For wavelength i (nm), the sensitivity Si was calculated as follows:

Si =$\frac{Var(f(X\_{400},…,X\_{i},…,X\_{850}) - f(\overbar{X}) }{Var(γ) }$, (7)

where Var is the variation, f() is the prediction of spectra due to the variation in wavelength i (nm) with other wavelengths held constant at their mean values, f(X) is the estimated value based on the mean reflectance, and $γ $represents the measured carotenoid content. Following the calculation of Si, the scores were converted to percentage

**3.Results and discussion**

3.1. Carotenoid Content for Each Cultivar

Carotenoid content of the tea cultivars is shown in Figure 3. The carotenoid content per cm² of leaf area was 6.18 - 15.63 μg and the averages for the three observation dates were 11.23, 10.28, and 9.45 μg, respectively. These differed significantly from each other, based on the Tukey-Kramer test (*p* < 0.001). Many of the lowest values were observed in black-tea-oriented cultivars (e.g., the mean carotenoid content of Benifuuki, Benihikari, and Benihomare was 8.89, 9.10, and 9.39 μg/cm², respectively), but the lowest value was observed in Hokumei (on 28 June). The cultivars with high values included matcha- and sencha-oriented cultivars (e.g., the mean carotenoid content of Kanaemaru, Minamisayaka, and Minekaori was 12.38, 11.99, and 11.70 μg/cm², respectively), and the highest value was observed in Minekaori (on 10 May).



**Figure 3.** Box plots of leaf carotenoid content

3.2. Spectral Reflectance

The mean reflectance measured by the spectrometers is shown in Figure 4. There are differences between variety in the green peak around 550 nm and the near-infrared region around 750 nm. Focusing on the green peak, benihikari has the highest reflectance and sunrouge has the lowest. On the other hand, when focusing on the near-infrared region, makinoharawase had the highest reflectance, while benihikari had the lowest reflectance.

**Figure 4.** Mean reflectance spectra measured by the Colorcompass-LF.

3.3. Correlation between Carotenoid Content and Reflectance

Two valleys are apparent in the graph of correlation coefficients for all dates (500 - 650 and 700 - 710 nm; *r*: -0.60 to -0.51; *p* < 0.001). The correlations were weaker on the first sampling day than on the other dates.

3.4. Sensitivity Analysis

 The results of the sensitivity analysis are shown in Figure 5. The 1D-CNN results contained data only from 415 to 810 nm because abnormal values occurred in the wavelength range at both ends. Wavelengths around 550 nm were important when using 1D-CNN (Figure 5.). On the contrary, this was not observed for DBN (Figure 5.b)

 (a)1D-CNN

(b)DBN

**Figure 5.** Sensitivity analysis results for the combinations of algorithms and spectrometers for the combinations of (a) 1D-CNN and reflectance from Colorcompass-LF, (b) DBN and reflectance from Colorcompass-LF.

3.5. Accuracy Assessment

 Table 1 shows the estimation accuracy of each analysis method. DBN's RPD was 1.16, category C, which was not very good estimation accuracy. On the other hand, 1D-CNN had an RPD of 1.43 and was in category B, achieving a practical estimation accuracy. Furthermore, R² was 0.2 for DBN, whereas it was 0.5 for 1D-CNN, indicating that 1D-CNN had better estimation accuracy.

**Table1.** RPD, RMSE (µg/cm²), and R² values for the estimation results for each deep learning algorithm after 100 repetitions.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | RMSE | RPD | R² |
| 1D-CNN | 1.08 | 1.43 | 0.50 |
| DBN | 1.35 | 1.16 | 0.22 |

 The estimation accuracy of 1D-CNN was basically better than DBN. The important wavelength bands of the 1D-CNN model were also evaluated, and the green peak around 550 nm was found to be important for estimating carotenoid content in leaves. In fact, the correlation coefficient between reflectance at 550 nm and carotenoids was -0.6, indicating a negative correlation. Although no preprocessing was performed on the collected data, the RPD exceeded 1.4 in 1D-CNN. Some previous studies reported that 1D-CNN does not need complex spectra pretreatment and variable selection (Li & Li, 2022; Nofrizal et al., 2022; Sonobe & Hirono, 2023)and then the strength of this algorithm was also confirmed in this study.

**4. Conclusions**

In this study, we acquired hyperspectral data using Colorcompass-LF, a low-cost complementary metal-oxide semiconductor (CMOS) sensor and evaluated its performance. As a result of estimating the carotenoid content of tea leaves using DBN and 1D-CNN, high estimation accuracy was obtained with 1D-CNN. Since it was found that there is a negative correlation between the reflectance near 550 nm and the carotenoid content of tea leaves, additional research with different cultivation conditions and measurement periods is expected to further improve the estimation accuracy. The information provided by the Colorcompass-LF can be used for better nutrient management, enabling less experienced farmers to monitor field scale at a lower cost, making quality control and plant maintenance easier.

**References**

Baker, R., & Günther, C. (2004). The role of carotenoids in consumer choice and the likely bene Wts from their inclusion into products for human consumption. *Trends in Food Science & Technology*, *15*(10), 484–488.

Chen, J. H., Zhao, Z. Q., Shi, J. Y., & Zhao, C. (2017). A New Approach for Mobile Advertising Click-Through Rate Estimation Based on Deep Belief Nets. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/7259762

Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6). https://doi.org/10.1109/JSTARS.2014.2329330

Chen, Y., Niu, S., Deng, X., Song, Q., He, L., Bai, D., & He, Y. (2023). Genome-wide association study of leaf-related traits in tea plant in Guizhou based on genotyping-by-sequencing. *BMC Plant Biology*, *23*, 196.

Chu, H., Zhang, C., Wang, M., Gouda, M., Wei, X., He, Y., & Liu, Y. (2022). Hyperspectral imaging with shallow convolutional neural networks (SCNN) predicts the early herbicide stress in wheat cultivars. *Journal of Hazardous Materials*, *421*. https://doi.org/10.1016/j.jhazmat.2021.126706

Jiang, X., Duan, H., Liao, J., Guo, P., Huang, C., & Xue, X. (2022). Estimation of Soil Salinization by Machine Learning Algorithms in Different Arid Regions of Northwest China. *Remote Sensing*, *14*(2). https://doi.org/10.3390/rs14020347

Li, D., & Li, L. (2022). Detection of Water pH Using Visible Near-Infrared Spectroscopy and One-Dimensional Convolutional Neural Network. *Sensors*, *22*(15). https://doi.org/10.3390/s22155809

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, *352*. https://doi.org/10.1016/j.geoderma.2019.06.016

Nofrizal, A. Y., Sonobe, R., Yamashita, H., Seki, H., Mihara, H., Morita, A., & Ikka, T. (2022). Evaluation of a One-Dimensional Convolution Neural Network for Chlorophyll Content Estimation Using a Compact Spectrometer. *Remote Sensing*, *14*(9). https://doi.org/10.3390/rs14091997

Pullanagari, R. R., Dehghan-Shoar, M., Yule, I. J., & Bhatia, N. (2021). Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. In *Remote Sensing of Environment* (Vol. 257). https://doi.org/10.1016/j.rse.2021.112353

Sonobe, R., & Hirono, Y. (2023). Carotenoid Content Estimation in Tea Leaves Using Noisy Reflectance Data. *Remote Sensing*, *15*(17), 4303.

Sonobe, R., Hirono, Y., & Oi, A. (2020). Quantifying chlorophyll-a and b content in tea leaves using hyperspectral reflectance and deep learning. *Remote Sensing Letters*, *11*(10). https://doi.org/10.1080/2150704X.2020.1795294

Wellburn, A. R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, *144*(3). https://doi.org/10.1016/S0176-1617(11)81192-2

Yang, Y. Z., Li, T., Teng, R. M., Han, M. H., & Zhuang, J. (2021). Low temperature effects on carotenoids biosynthesis in the leaves of green and albino tea plant (Camellia sinensis (L.) O. Kuntze). *Scientia Horticulturae*, *285*. https://doi.org/10.1016/j.scienta.2021.110164