Comprehensive Review on Tea Clone Classification Systems

Sachini Tennakoon[1] and Sulochana Rupasighe[2]

[1] University of Westminster London, UK
[2] Informatics Institute of Technology Colombo, Sri Lanka
w1742227@my.westminster.ac.uk

Abstract. A wide range of tea clones have been developed over the years**.** As each tea clone produces a distinct quality of tea, it is critical to identify them in the field. Tea clones may have extremely similar characteristics, making it difficult for tea farmers and tea estate owners to distinguish them manually. This problem can be resolved by using machine learning to create an application that recognizes tea clones automatically. This paper conducts a comparative review of existing tea clone classification systems, followed by a study that identifies research gaps and potential future works.

Keywords: Tea clone variants, Tea clone identification, Image classification, Machine learning.

1 Introduction

Tea has a long history of popularity all over the world. In 2020, the worldwide tea market was estimated to be worth approximately USD 200 billion [1]. A "clone" is a plant population derived from tea seeds through vegetative propagation [2]. The tea plant is nearly self-incompletable, therefore two separate plants are normally required to generate the seeds [2]. As a result, different clones with the same crossing parents exist. This leads to having high physical similarities between some clones [3].

Different clones produce tea of varying quality and quantities. However, due to the high physical similarities between clones, ordinary people such as tea farmers and tea estate owners, find it difficult to manually identify them. A qualified professional with years of experience is required to identify them through visual inspection by examining physical sizes, textures, bone structures, and leaf color [4]. There aren't many professionals who can identify these tea clones. Even if they did, it would be costly, time-consuming, and prone to errors [3]. This problem can be addressed using an automated tea clone classification system.

Machine learning can be used to detect tea clones automatically. Several studies over the last decade have used tea leaves to detect diseases [5], [6] and tea categories [7], [8] using traditional machine learning algorithms such as K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM), and they have performed well in basic scenarios.

Deep learning is a popular approach in computer vision for object recognition. In recent years, this technology was used to classify different varieties of tea clones. Yuwana [9] introduced a system to classify tea clones using a Support Vector Machine (SVM) as the classifier, and reduced image features using an encoder. Ramdan [10] used AlexNet and VGGNet which are Convolutional Neural Network (CNN) architectures, while Ramdan [11] used ResNet and DenseNet which are also CNN architectures to classify tea clones. Rizal [12] and Ibrahim [13] also implemented a system to identify tea clones using a CNN architecture. Suryawati [14] proposed a system by applying CNN to the Generative Adversarial Network (GAN) for feature extraction and Fully connected layers (FCL) as the classifiers. Zilvan [3] proposed an autoencoder variation for feature extraction and FCL as a classifier.

This research offers a comprehensive comparison of existing tea clone classification models. The second section of this research reviews different tea clone variants and compares existing tea clone classification methods. Section 3 of the article discusses the observed research gaps, whereas Section 4 discusses this survey's limitations. Section 5 discusses potential future works, while Section 6 concludes the entire survey.

2 Tea Clone Classification

2.1 Tea Clone Variants

Countries like India, Bangladesh, China, Taiwan, Sri Lanka, Korea, Japan, Kenya, Indonesia, and Vietnam have developed different variants of tea clones[15]. Several studies have been conducted in recent years to identify various morphological characteristics of some tea clones.

Syahbudin [16] has conducted a study to identify the morphological characteristics of tea clones and their utilization, which are established closed to mount Lawu forest, East Java. Asamica, a clone originated from Sri Lanka, yabokita, a clone from Japan, Chin, a clone from China, Gambung Assamica (GMB) 3 and GMB 7 clones developed by Tea and Quinine Research Center Gambung, West Java, and Tea Research Institute (TRI) 2025, and TRI 2024 developed by the Tea research institute of Sri Lanka are the seven tea clones that were selected for this research. TABLE Icompares the morphological features (Root and stem, length of leaf, thickness of leaf, fine hairs on tea buds) of these seven clones.

Table 1. Morphological characteristics of tea clones in East Java, Indonesia

Tea Clones	**Root and Stem**	**Length of leaves**	**Thickness of leaves**	**Fine hairs on the tea buds**
Asamica	Big and sturdy stem	11.5 cm	Thin, toothed margin, and oblong	Thin white hair
Yabokita	Small stem	10.6 cm	Thick and oblong, dark green on mature leaves	Thin white hair
Chin	Small stem	7.4 cm	Thick, narrow, small	Thin white hair

GMB 3	Small stem	15.7 cm	Thin, oblong with long-pointed, long leaves	Thin white hair
GMB 7	Small stem	10.1 cm	Thin, oblong with long pointed, shorter than GMB 3 clone leaves	Thick white hair
TRI 2025	Small stem	16.5 cm	Thickest leaves	Thin white hair
TRI 2024	Small stem	13.9 cm	Slightly thinner than TRI 2025 clone	Thin white hair
Yabokita	Small stem	10.6 cm	Thick and oblong, dark green on mature leaves	Thin white hair

Abdul [17] conducted research to identify parametric characteristics of stems and leaves from six Bangladeshi clones: MZ/39, E/4, D/13, B2T1, BR2/97, SDL/1, and BT2. Stem circumference, the height of the first branching position, leaf length, leaf breadth, leaf length/breadth ratio, leaf size, leaf angle in degree, internodal length, shoot density, and yield of green leaf per hectare were the characteristics identified.

Zivan [3] conducted research to classify GMB 3 and GMB 9 clones using deep learning. There it has shown the morphological characteristics of these two clones. TABLE II shows the morphological characteristics of these two clones.

Table 2. Morphological Characteristics of GMB 3 and GMB 9

Clone	Leaf Characteristics
GMB 3	• Leaf shape is elliptic (1:3.02) • Leaf width is 3.06 cm • Leaf length is 9.25 cm • Phyllotaxis is 73.3o • Leaf area is 15.45 cm2 • The edge of the leaf is serrated with obtuse and irregular • The face of the leaf is flat • Number of bones is seven pairs • The shape of the base of the leaf is pointed • The tip of the leaf is taper • Leaf color is dark green
GMB 9	• Leaf shape is elliptic (2.08:1) • Leaf length is 3–7 cm • Leaf area is 34.71 cm2 • Phyllotaxis is 21–47o • The shape of the base of the leaf is obtuse to sharp • Number of leaf bone is 16–24 pairs • The edge of the leaf is regular sharp serrated • The shape of the tip of the leaf is tapered • The face of the leaf is bumpy • Leaf color is yellowish green

2.2 Recent Work on Tea Clone Classification

The following is a broad analysis of tea clone classification systems. Each model is explained individually. This followed by a comparison table comparing the techniques, improvements, and future works of all the deep learning models.

Bottleneck RGB Features for Tea Clone Identification. : Yuwana [9] introduced a system to classify two varieties of GMB series clones. This research has proposed an encoder-based feature reduction approach to reduce RGB image dimensions. This approach was used because the original dataset comprises a large number of dimensions that demand a large storage capacity and is affected by the time it is consumed to process data. The output features are used as input to the SVM classifier to classify the two clones, namely GMB 3 and GMB 9. This research shows that reducing the RGB image dimensions can achieve better performance than using a full dimension RGB image. This study has contributed 1297 images of GMB series tea clones, including 597 images of GMB 3 tea leaves and 699 images of GMB 9 tea leaves.

Convolutional Variational Autoencoder-Based Feature Learning for Automatic Tea Clone Recognition. : Zilvan [3] proposed an approach to classify the same types of clones used in Yuwana [9] which are GMB 3 and GMB 9 clones. The proposed system includes two main parts: 1. variant of the autoencoder (VAE) [18] for feature learning, which is trained in an unsupervised environment, 2. FCN classifier. The VAE was built using CNN architecture. CNN was used because it processes input images in three dimensions: width, height, and channel, which is more suitable than using a typical neural network. This proposed architecture was named Convolutional Variational Autoencoder (CNNVAE). The output of CNNVAE is used as the input for the FCN classifier. This classifier has used the last three fully connected layers of VGGNET16 architecture. This research has used different CNNVAE architectures with different levels of depth by changing the number of convolutional layers in both encoder and decoder. This research shows that using CNNVAE-3 (The number of convolutional layers is three in both encoder and decoder) is the most effective architecture. And the performance of CNNVAE-3 is more robust than VGGNet16 for blurred image tests without pre-training (Gaussian Blurring and Median Blurring). Moreover, this study has contributed 1297 images of GMB 3 and GMB 9 series tea clones.

Deep Convolutional Adversarial Network-Based Feature Learning for Tea Clones Identifications: Suryawati [14] also proposed a new approach for the same types of clones used in Yuwana [9] and Zilvan [3]. In this research, they have proposed an unsupervised feature learning algorithm derived from Deep Convolutional Generative Adversarial Network (DCGAN) [19]. DCGAN is built by applying CNN to the GAN (Generative Adversarial Network) architecture. For classification, FCN (Fully Convolutional Network) classifier was used. The result shows that DCGAN architecture performs better compared to Autoencoder. This study has contributed 1297 images of GMB 3 and GMB 9 series tea clones.

Deep CNN Based Detection for Tea Clone Identification: Ramdan [10] classified three GMB series tea clones, using CNN architecture. They have applied two architectures of CNN namely AlexNet (6 layers) and VGGNet (16 layers) to investigate the effect of depth of the convolutional layers on the performance of the classification system. The results show that the depth of the AlexNet and VGGNet model influences their ability to produce a good model performance and a deeper network can improve accuracy but tends to overfit. Therefore, VGGNet is shown to be slightly better. This research also showed that the learning rate setting is influential for deep neural network training to achieve good performance. This study has contributed 1966 images of GMB series tea clones, including 598 images of GMB 3 tea leaves, 699 images of GMB 9 tea leaves, and 699 images of GMB 11 tea leaves.

Tea Leaves GMB Series Classification Using Convolutional Neural Network: Rizal [12] is the first research to classify all the 11 types of GMB series clones (GMB 1 to GMB 11). The system has been built using CNN architecture. This has three hidden layers and a fully connected layer with the softmax activation function. This research has contributed 1689 images of all 11 types of tea clones.

Tea Clone Classification using Deep CNN with Residual and Densely Connections: Ramdan [11] has developed a classification system, to classify 6 different types of GMB series clones using a CNN architecture. They have applied 2 architectures of CNN, namely ResNet and DenseNet. This study demonstrates that increasing the depth of a network to improve accuracy is possible if CNN has direct connections at various layers and contributions from feature maps reintroduced into deeper layers via connections can improve model performance. This research has also shown the importance of choosing the correct learning rate. This research has contributed 3281 images of all the 6 types of clones.

Computer Aided System for Gambung Tea Identification using Convolutional Neural Network: Ibrahim [13] introduced a system to classify Gambung and Non-Gambung Series tea leaves. The system has been built using CNN with two hidden layers and one fully connected layer with the sigmoid activation function. This research has contributed 289 images of GMB series tea clones, including 63 of non-Gambung series leaves and 220 of Gambung series tea leaves.

The techniques, improvements, and future works of all the deep learning models explained above are summarized in TABLE III.

Table 3. Summary of Tea Clone Classification Systems

Research	Technology	Improvements	Future works
Yuwana [9] .	Encoder for dimension reduction and SVM as the classifier	Increased the performance by reducing the RGB image dimensions using an encoder	1. Classify all clone classes of the GMB series. 2. Use other features such as texture, shape, and leaf angle to do the classification.
Ramdan [10] .	Used two CNN architectures: AlexNet and VGGNet16.	Increased the performance by increasing the	1. Create a method to automate the process of setting CNN parameters.

		depth of the models.	2. Classify all classes of GMB clone tea leaf series.
Suryawati [14]	Deep Convolutional Generative Adversarial Network (DCGAN) for dimension reduction and the fully connected layers of VGG architecture as the classifier.	Proposed DCGAN by applying CNN to the GAN (Generative Adversarial Network) architecture	1. Investigate various deep learning architectures to improve performance. 2. Improve performance by adding more training data, increasing the depth of the encoder, and experimenting with other encoder variants.
Rizal[12].	CNN architecture with three hidden convolutional layers and one fully connected layer.	Classified all 11 classes of GMB series clones.	Not specified
Ibrahim [13] .	CNN architecture with two hidden layers	Classified GMB series clones and non-GMB series clones.	Not specified
Zilvan[3] .	CNN based Variational Autoencoder (CNNVAE) for dimension reduction and FCN for classification	1. Proposed CNNVAE by applying CNN to the VAE. 2. Classified blurred images and achieved good accuracy.	1. Evaluate the robustness of CNNVAE based feature learning using a different type of noise type. 2. Improve performance by investigating the various number of latent attributes
Ramdan [11]	Used two CNN architectures: ResNet and DenseNet.	Increased the accuracy by using ResNet and DenseNet.	Not specified

3 Exploring Research Gaps

Yuwana [9], Zilvan [3], Suryawati [14], Ramdan [10], Rizal [12] and Ramdan [11] have built tea clone classification systems using various techniques. These systems used RGB images of the tea leaves to train the model. However, RGB images are light-sensitive. As a result, using only RGB images as input data might impair the classification model's performance.

Ramdan [10] proposed an AlexNet and VGGNet architecture-based tea clone classification system. VGGNet outperformed AlexNet in terms of accuracy, although the model is prone to overfit. This indicates that as the depth of the architecture increases, the model begins to overfit. Suryawati [14] proposed a tea clone classification model that extracts features using an Encoder-Deep Convolutional Generative Adversarial Network. This model's performance can be enhanced further by adding more training data, increasing the depth of the encoder, and experimenting with other encoder variants. The performance of the tea classification model proposed by Zilvan [3] can be enhanced by testing with different latent attributes on the Convolutional Variational Autoencoder used for feature extraction.

There are various types of tea clones developed by different countries. Yuwana [9], Zilvan [3], Suryawati [14] identified GMB 3 and GMB 9 clones. Ramdan[10] classified GMB 3, GMB 9, and GMB 11 clones, whereas Ramdan[11] classified GMB 1, GMB 3, GMB 7, GMB 9, and GMB 11. Taking all existing tea clone identification systems into consideration, they have only categorized GMB series clones developed by the Tea and Quinine Research Center Gambung, West Java.

Throughout all the works, improving the accuracy of the classification model, reducing the tendency of overfitting of the model, improving the model to classify varying tea clones are the gaps that are evident. There is enough space to work on these gaps and improve the models in the tea clone classification domain.

4 Limitations of the Survey

Three studies on the differences in morphological features of clones were discussed in this survey. These works were chosen based on the clones used in the studies. However, there are other studies conducted based on the morphological characteristics of tea clones. Since the scope of this survey is primarily focused on automated tea clone classification systems, all studies based on morphological characteristics of tea clones were not taken into strong consideration throughout the survey.

5 Future Work

There hasn't been a lot of research done in the domain of tea clone classification. As a result, there is much more to be researched and improved. The focus is mostly on enhancing the classification model's accuracy. Because there is no available dataset for tea leaves of the same clone, all research requires novel datasets to perform classification. As a result, it is difficult to acquire a large amount of data, which is why most systems fail to increase accuracy. The goal of this paper is to give future researchers a quick grasp of innovative and credible tea clone classification works, as well as potential future work pathways along which the domain's research can be extended.

A summary of future works that can be done in the domain would be 1) to enhance existing algorithms to improve accuracy 2) to attempt new algorithms to improve accuracy. 3) In addition to RGB photos, use other strong features such as texture, shape, and leaf angle. 4) Improve the classification model so that it may be used to classify additional tea clones around the world.

6 Conclusion

There are various types of clones developed by different countries**. Some of these** tea clones may have extremely similar characteristics, making it difficult to distinguish between them manually. To address this problem, research has been conducted in the past to automate tea clone classification.

This paper thoroughly surveys 7 tea clone classification systems. That is followed by a comprehensive tabular analysis of techniques, limitations, and future works mentioned for each work, which will be quite useful to many future researchers of the domain.

Acknowledgment

. We would like to express our heartfelt appreciation to everyone who supported us throughout this research. Without their unwavering support, this research would have remained merely an idea.

References

[1] "Global tea market revenue 2012-2025," *Statista*. https://www.statista.com/statistics/326384/global-tea-beverage-market-size/ (accessed Mar. 14, 2022).

[2] Tea Research Institue of Sri Lanka, *Handboook on tea*. 2008.

[3] V. Zilvan *et al.*, "Convolutional variational autoencoder-based feature learning for automatic tea clone recognition," *Journal of King Saud University - Computer and Information Sciences*, p. S1319157821000227, Feb. 2021, doi: 10.1016/j.jksuci.2021.01.020.

1. [4] "My Bonfire." https://www.gamboeng.com/pages/detail/2015/59/146 (accessed Mar. 20, 2022).

[5] S. Hossain, R. M. Mou, M. M. Hasan, S. Chakraborty, and M. A. Razzak, “Recognition and detection of tea leaf’s diseases using support vector machine,” in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Batu Feringghi, Mar. 2018, pp. 150–154. doi: 10.1109/CSPA.2018.8368703.

[6] B. Karunamoorthy and D. Somasundereswari, “A defect tea leaf identification using image processing,” *Przeglad Elektrotechniczny*, vol. 89, pp. 318–320, Sep. 2013.

[7] X. Wu, J. Yang, and S. Wang, “Tea category identification based on optimal wavelet entropy and weighted k-Nearest Neighbors algorithm,” *Multimed Tools Appl*, vol. 77, no. 3, pp. 3745–3759, Feb. 2018, doi: 10.1007/s11042-016-3931-z.

[8] S. Wang, P. Phillips, A. Liu, and S. Du, “Tea Category Identification using Computer Vision and Generalized Eigenvalue Proximal SVM,” *FI*, vol. 151, no. 1–4, pp. 325–339, Mar. 2017, doi: 10.3233/FI-2017-1495.

[9] R. Sandra Yuwana, E. Suryawati, Ana Heryana, Vicky Zilvan, Dadan Rohdiana, and Heri Syahrian K., “Bottleneck RGB Features for Tea Clones Identification,” in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, Dec. 2019, pp. 259–262. doi: 10.1109/ISRITI48646.2019.9034667.

[10] Ade Ramdan *et al.*, “Deep CNNBased Detection for Tea Clone Identification,” *indones.j.electron.telecommun.*, vol. 19, no. 2, p. 45, Dec. 2019, doi: 10.14203/jet.v19.45-50.

[11] A. Ramdan, V. Zilvan, E. Suryawati, H. F. Pardede, and V. P. Rahadi, “Tea clone classification using deep CNN with residual and densely connections,” *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 289–296, Oct. 2020, doi: 10.14710/jtsiskom.2020.13768.

[12] S. Rizal, N. K. C. Pratiwi, N. Ibrahim, H. Vidya, S. Saidah, and R. Y. N. Fu’adah, “TEA LEAVES GMB SERIES CLASIFFICATION USING CONVOLUTIONAL NEURAL NETWORK,” *JOURNAL OF ELECTRICAL AND SYSTEM CONTROL ENGINEERING*, vol. 3, no. 2, Art. no. 2, Feb. 2020, doi: 10.31289/jesce.v3i2.3320.

[13] N. Ibrahim, Y. Fu’adah, N. C. Pratiwi, S. Rizal, and K. Usman, “Computer Aided System for Gambung Tea Identification using Convolutional Neural Network,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 982, p. 012001, Dec. 2020, doi: 10.1088/1757-899X/982/1/012001.

[14] E. Suryawati, V. Zilvan, R. S. Yuwana, A. Heryana, D. Rohdiana, and H. F. Pardede, “Deep Convolutional Adversarial Network-Based Feature Learning for Tea Clones Identifications,” in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, Oct. 2019, pp. 1–5. doi: 10.1109/ICICoS48119.2019.8982416.

[15] “Planting Material of Tea::Tea World - An Initiative of KKHSOU.” http://teaworld.kkhsou.in/lessons_details.php?lesson=Planting-Material-of-Tea&url-titlepage=5f75a9d64caca283f1e7 (accessed Mar. 20, 2022).

[16] A. Syahbudin, A. Widyastuti, N. W. Masruri, and A. Meinata, “Morphological Classification of Tea Clones (Camellia sinensis, Theaceae) at the Mount Lawu Forest, East Java, Indonesia,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 394, p. 012014, Dec. 2019, doi: 10.1088/1755-1315/394/1/012014.

[17] M. A. Aziz, B. Ahmed, M. A. Razvy, M. R. Karim, S. K. L. Haque, and M. Hossain, "Comparative study on some morphological features of six selected and one standard clones of Bangladesh Tea [Camellia sinensis (L) O. Kuntze]," p. 9, 2011.

[18] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114 [cs, stat]*, May 2014, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1312.6114

[19] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, 2014, vol. 27. Accessed: Oct. 19, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html