A Data Descriptor for Black Tea Fermentation Dataset

Gibson Kimutai [1,2,*] , Alexander Ngenzi [1] , Rutabayiro Ngoga Said [1] , Rose C. Ramkat [3] and Anna Förster [4]

1. African Center of Excellence in Internet of Things (ACEIoT), College of Science and Technology, University of Rwanda, P.O. Box, 3900 Kigali, Rwanda; yngenzi37@gmail.com (A.N.); said.rutabayiro.ngoga@gmail.com (R.N.S.)
2. Department of Mathematics, Physics and Computing, Moi University, P.O. Box, 3900-30100 Eldoret, Kenya
3. Department of Biological Sciences, Moi University, P.O. Box, 3900-30100 Eldoret, Kenya; rose.ramkat@mu.ac.ke
4. Sustainable Communication Networks, University of Bremen, 8359 Bremen, Germany; anna.foerster@comnets.uni-bremen.de
* Correspondence: kimutaigibs@gmail.com

Abstract: Tea is currently the most popular beverage after water. Tea contributes to the livelihood of more than 10 million people globally. There are several categories of tea, but black tea is the most popular, accounting for about 78% of total tea consumption. Processing of black tea involves the following steps: plucking, withering, crushing, tearing and curling, fermentation, drying, sorting, and packaging. Fermentation is the most important step in determining the final quality of the processed tea. Fermentation is a time-bound process and it must take place under certain temperature and humidity conditions. During fermentation, tea color changes from green to coppery brown to signify the attainment of optimum fermentation levels. These parameters are currently manually monitored. At present, there is only one existing dataset on tea fermentation images. This study makes a tea fermentation dataset available, composed of tea fermentation conditions and tea fermentation images.

Keywords: tea; fermentation; internet of things; detection; dataset

1. Background and Rationale

Tea is currently among the most popularly consumed beverage across the world [1] and is responsible for the economic growth of many countries, including India, Sri-Lanka, Kenya, China among other countries [2]. These top tea-producing countries produce several varieties of tea, which include: yellow tea, illex tea, oolong tea, black tea, white tea, among others [3]. Among these categories of tea, black tea is the most consumed, accounting for approximately 78% of the total daily consumption of tea [4]. Kenya is the leading exporter of black tea worldwide, with her major tea-producing counties being Kericho, Bomet, Nandi, and Nyeri [5]. The crop is a source of livelihood for more than 10 million of the total countries' estimated population of 47 million people [6]. Although the crop is still the leading exchange earner for the county, the sector is ailing due to ever-reducing tea prices [7]. This is attributed to increased competition from other countries, poor management, and the low quality of tea produced, among others [8].

The steps of processing black tea are plucking, withering, crushing, tearing and curling, fermentation, drying, sorting, and packaging [9]. Among these processes, fermentation is the key determinant of the final quality of the processed tea [10]. The process is time-bound and must take place within a given temperature and humidity range [11]. During fermentation processes, tea changes color from green to coppery brown and finally to dark red [12]. The optimally fermented tea is coppery brown in color, and has a fruity

smell and a sweet taste, while the unfermented tea is green in color, and has a grassy smell and a strong taste [13]. The overfermented tea is dark red and is characterized by a bitter taste [14]. Currently, the optimum fermentation of tea is monitored manually by tea tasters, adopting the following techniques: monitoring color change, tasting tea as fermentation progresses, and smelling the odor of tea during fermentation [15]. These manual methods are subjective and lead to a compromise in the quality of the made tea. Image processing and machine learning techniques have shown high levels of ability in various fields, including medicine [16], education, E-Commerce [17] tourism and banking, among others [18]. However, for image processing and machine learning to work, there is a need for data for training and evaluation of the models. Worryingly, there is only one reported open-source dataset on tea fermentation images [19], which is a limiting phenomenon as researchers have only one dataset to train and evaluate their machine learning models. Therefore, this study aims to resolve this challenge by adopting the Internet of Things (IoT) to capture and release a tea fermentation dataset composed of temperature, humidity, and black tea fermentation images.

The rest of the paper is arranged as follows. Section 3 presents the description of the data in the dataset. Section 2 presents materials and methods, and Section 5 presents the conclusion of the paper.

2. Materials and Methods

This section describes the resources and the approach followed in the collection of the dataset. Section 2.1 discusses on the resources while the collection of the dataset is discussed in Section 2.2.

2.1. Resources

The following resources were instrumental in acquiring the data: Raspberry pi, Pi-Camera, Server, and programming languages. Raspberry Pi model B+ was adopted due to its increased processing power and its dual-band Wi-Fi Feature [20]. The Raspbian operating system for raspberry pi [21] was used. Raspian was chosen as it is available at no cost and is easy to install and use. A raspberry pi camera of 8 megapixels was used. The board was chosen since it is very small, weighing around 3 g, making it perfect for deployment with the raspberry pi. Amazon Web Services (AWS) [22] was chosen as a cloud provider for storage of the data. The AWS provides services that make it easy to store images and also offers an initial free service for 1 year. Python programming language [23] was used in writing programs to capture the images using the Pi camera. The block diagram of the system for capturing the dataset is shown in Figure 1.

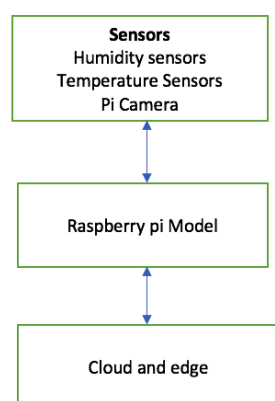



Figure 1. Block diagram of the data collection system.

2.2. Collection of the Dataset

The Internet-of-Things-based system for capturing the data was deployed in the Sisibo factory, Kenya for 4 days: 10–13 August 2020. The system was set up just above the tea

fermentation bed (Figure 2). After collection of each image, the tea experts provided ground truths on their correct classes.

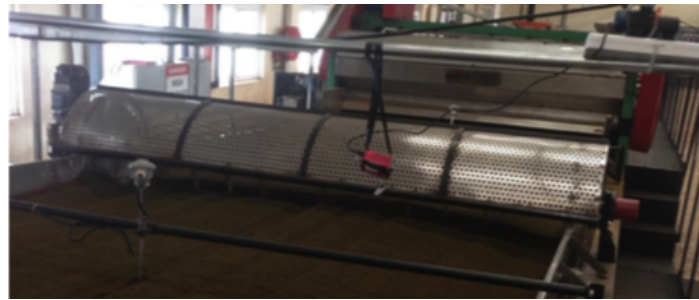



Figure 2. Collection of the dataset in Sisibo tea factory, Kenya using Raspberry pi and Pi camera.

3. Data Description

This study releases a black tea fermentation dataset which contains black tea fermentation images and physical parameters of tea during fermentation, which were collected in Sisibo tea factory, Kenya between 10 and 13 August 2020. The dataset can be found here: https://doi.org/10.5281/zenodo.4469326 (accessed on 2 March 2021). The file structure of the dataset is shown in Figure 3.

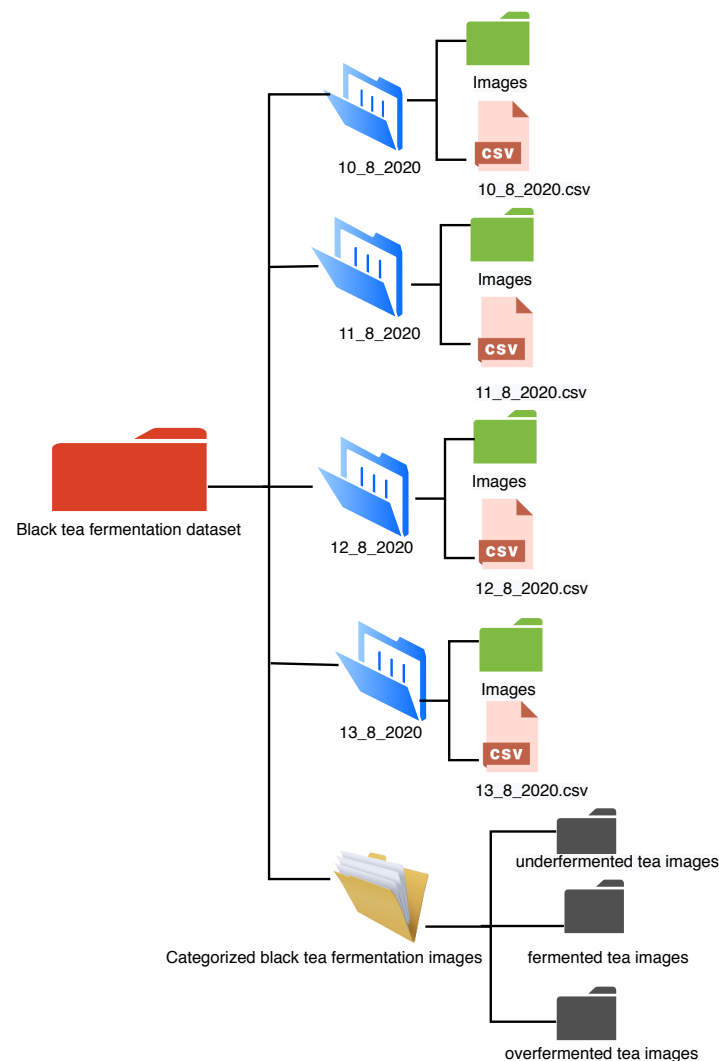



Figure 3. File structure of the black tea fermentation dataset.

Black tea fermentation conditions are contained in a Comma Delimited Values (CSV) file and contain the fermentation time, temperature, humidity, reference to image and the category of tea (Figure 4).

	Time	Temperature (°C)	Humidity (%)	Reference to image	Category of tea
1	Time	Temperature (°C)	Humidity (%)	Reference to image	Category of tea
2	12:31:13	24	61.6	underfermented_1	unfermented
3	12:32:12	23.9	62.1	underfermented_2	unfermented
4	12:35:58	23.7	62.8	underfermented_3	unfermented
5	12:36:52	23.6	64	underfermented_4	unfermented
6	12:37:55	23.4	63.9	underfermented_5	unfermented
7	12:38:53	23.4	64.9	underfermented_6	unfermented
8	12:42:53	23.5	64.1	underfermented_7	unfermented
9	12:43:51	23.4	65	underfermented_8	unfermented
10	12:44:54	23.4	64.5	underfermented_9	unfermented

Figure 4. A Screenshot of the fermentation condition CSV file.

The images folder contains images that were taken as fermentation took place. These images correspond with the Reference showing the images column in the CSV files. In addition to these images, "Categorized black tea fermentation images" folder contains 6000 black tea fermentation images categorized into three classes of 2000 images each. The classification of these images was based on the decision of two tea fermentation experts. The experts relied on the color, smell, and taste of tea during fermentation to classify the images. The classes are unfermented, fermented, and overfermented.

The underfermented tea is in a folder labeled "unfermented tea" in the "Categorized black tea fermentation images" folder. The level of fermentation of this category of tea is below the optimum. These tea images are usually green. The images are labeled from "underfermented _00 " to "underfermented _1999". A sample of the images is presented in Figure 5a.

The fermented tea is optimally fermented and is usually coppery brown. The fermented tea is in a folder labeled "fermented tea" in the images folder. The images are labeled from "fermented _00" to "fermented _1999". A sample of the images is presented in Figure 5b.

The fermentation level of overfermented tea is beyond the optimum and is dark red. Overfermented tea is in a folder labeled "overfermeted tea". The images are labeled from "Overfermented _00" to "Overfermented _1999". A sample of the images is presented in Figure 5c.

(a) Unfermented   (b) Fermented   (c) Overfermented

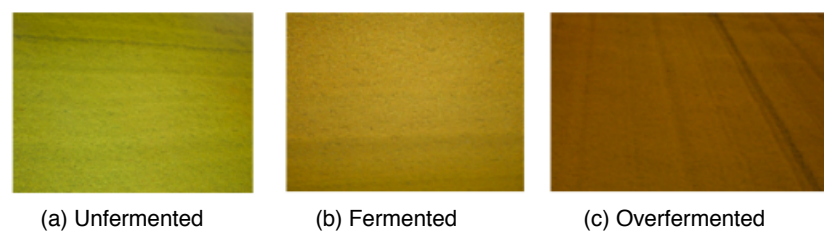



Figure 5. A sample of classes of tea image fermentation dataset released in this paper.

4. Data Validation

The fermentation of tea must take place within a given range of temperature and humidity. These ranges vary in different countries but, in Kenya, the acceptable range is between 20 and 30 degrees celcius. Figure 6 shows temperature and humidity values during a fermentation cycle in Sisibo tea factory on 10 August 2020. The fermentation process started at 12:31:13 h. The temperature recorded ranged from 19 °C to 28 °C , which was well within the optimum ranges. The values of the temperature increased steadily with time due to the natural climatic conditions of the area. On the other hand, humidity was between 75% and 92% for the low and the high, respectively. The values of humidity reduced steadily with time. The highest temperature value recorded was 28.8 °C, while the

highest humidity value achieved was 82.6%. The fermentation curve was smooth and it took 68 min before the tea was fully fermented.

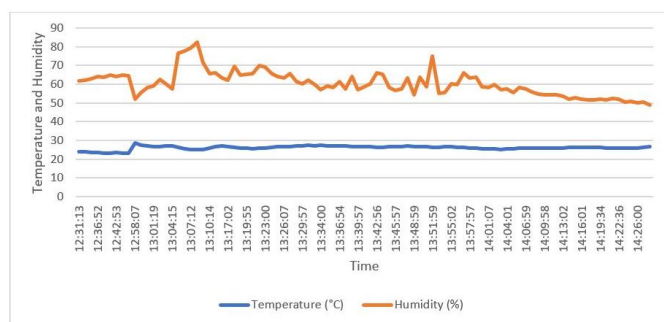
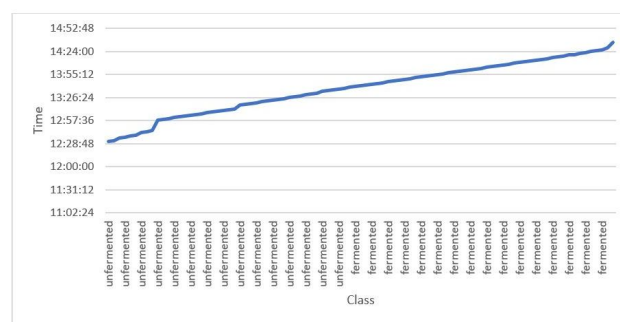
Figure 7 shows temperature and humidity values during a fermentation cycle in Sisibo tea factory on 11 August 2020. The fermentation process started at 08:48:03 h. The temperatures recorded ranged from 20.6 °C to 24.50 °C. These temperatures were within the optimum. The values of the temperature increased steadily with time due to the natural climatic conditions of the area. The highest humidity value recorded was at 90.8%. The values of humidity fluctuated throughout the fermentation period. The fermentation curve was smooth and it took 51 min before the tea was fully fermented. This is the shortest fermented duration reported in this data descriptor, as the fermentation temperatures were higher than the rest of the days.

(**a**) Temperature and humidity values

(**b**) Fermentation classes of tea with time

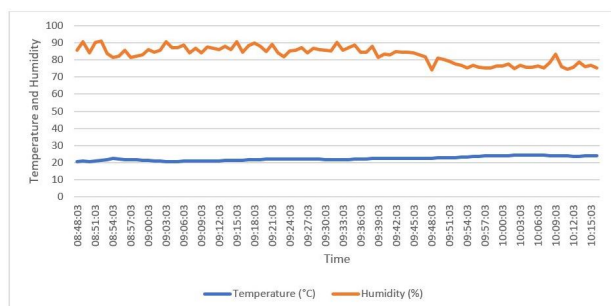
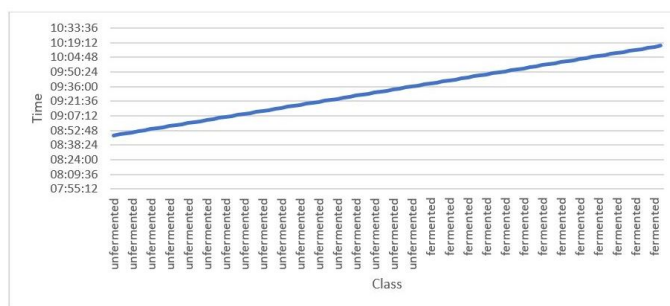
Figure 6. Fermentation conditions of tea in Sisibo tea factory on 10 August 2020.

(**a**) Temperature and humidity values

(**b**) Fermentation classes of tea with time

Figure 7. Fermentation conditions of tea in Sisibo tea factory on 11 August 2020.

Figure 8 shows temperature and humidity values during a fermentation cycle in Sisibo tea factory on 12 August 2020. The fermentation process started at 02:00:13 h. The temperature recorded ranged from 15 °C to 25.53 °C. These temperatures fell below the optimum, unlike on 10 August 2020. This is due to the fermentation occurring in the early morning when the region is naturally colder than during midday hours. The values of the temperature increased steadily with time, due to the natural climatic conditions of the area. The highest humidity value recorded was at 95%. The fermentation curve was smooth and it took 64 min before the tea was fully fermented.

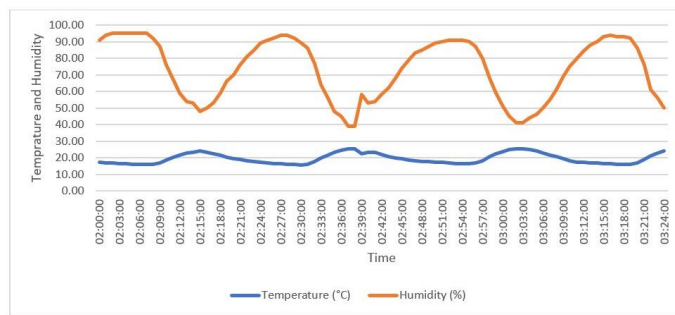
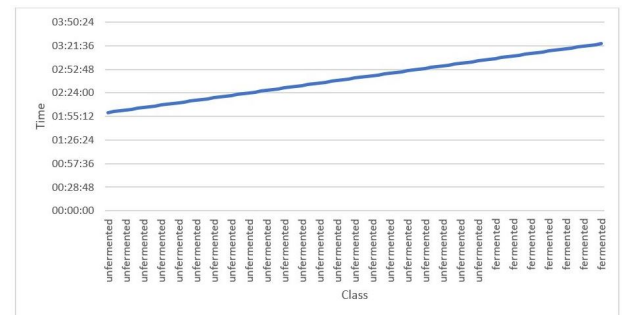
Figure 9 shows temperature and humidity values during a fermentation cycle in Sisibo tea factory on 13 August 2020. The fermentation process started on 07:00:00 h. The highest temperature recorded was at 26.25 °C. Once again, the temperature ranges fell below the optimum, unlike on 10 August 2020. This is because fermentation was carried out in the early morning when the region is naturally colder than during midday hours. The highest humidity value recorded was at 95%. The fermentation curve was smooth and it took 58 min before the tea was fully fermented.

(**a**) Temperature and humidity values

(**b**) Fermentation classes of tea with time

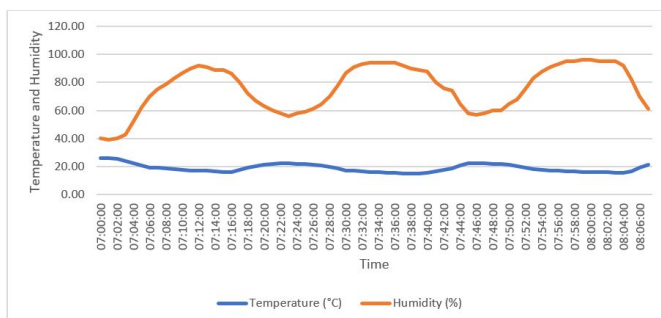
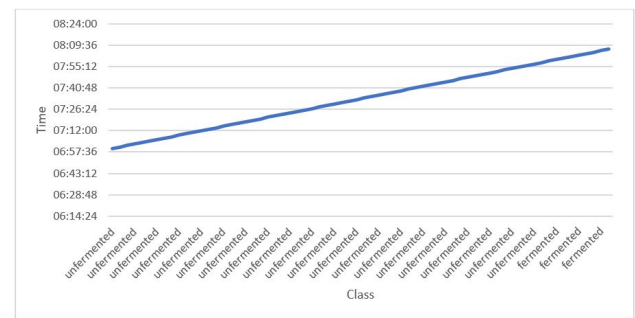
Figure 8. Fermentation conditions of tea in Sisibo tea factory on 12 August 2020.

(**a**) Temperature and humidity values

(**b**) Fermentation classes of tea with time

Figure 9. Fermentation conditions of tea in Sisibo tea factory on 13 August 2020.

The average temperature and humidity values for the days are presented in Figure 10. The highest average temperature was recorded on 10 August 2020, with 26.14 °C. On 11 August 2020, an average temperature of 22.31 °C was recorded. On the other days, the average temperatures were 19.39 °C and 18.83 °C for 12 August 2020 and 3 August 2020, respectively. Thus, the fermentation took place outside the range of 20–30 °C for the two days.

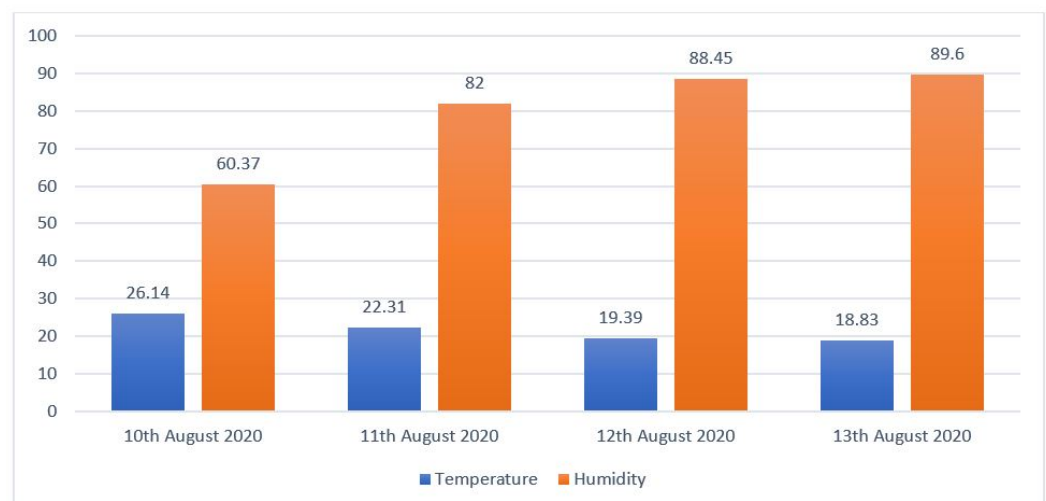



Figure 10. Temperature and humidity during black tea fermentation in Sisibo tea factory between 10 and 13 August 2020.

5. Conclusions

In this paper, a tea fermentation dataset has been released. The dataset was collected in the Sisibo tea factory, Kenya using an IoT-based model. The model predicted fermentation

cycle of tea and correlated well with the decisions of the tea fermentation experts. Fermentation experts gave ground truths for the dataset. The dataset can be used by researchers in training machine learning models for the detection of the optimum fermentation of tea. This is a significant achievement in the field of applying machine learning to the detection of optimum fermentation of tea, as there is currently only one existing dataset for training these models.

Author Contributions: Formal analysis, A.F.; methodology, G.K. and A.F.; software, G.K.; supervision, A.N., R.N.S., R.C.R., and A.F.; validation, G.K.; visualization, G.K.; writing original draft, G.K.; writing review and editing, A.N., R.N.S., R.C.R., and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: The Article Processing Charges was funded by the University of Bremen, Germany. This research received financial support from the African Center of Excellence in Internet of Things (ACEIoT) and the Kenya Education Network (KENET) through the Computer Science, Information Systems, and Engineering multidisciplinary mini-grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are publicly archived in the Zenodo platform: https://doi.org/10.5281/zenodo.4469326 (accessed on 2 March 2021).

Acknowledgments: We wish to thank the University of Bremen for funding the Article Processing Charges of this article. Additionally, we wish to thank the management and the entire staff of Sisibo tea factory, Kenya for allowing us to collect the datasets in their factory. Special appreciation goes to the tea tasters of Sisibo Tea Factory: Abdalla Tunya and Cornellius Chirchir, who offered their expert judgement for every tea image in the dataset. The authors wish to acknowledge the Kenya Education Network (KENET) for financially supporting this research through the CS/IT Minigrant. Finally, the authors would like to express their appreciation to the anonymous reviewers and editors for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
AWS	Amazon Web Services
E-Commerce	Electronic Commerce

References

1. Miura, K.; Hughes, M.C.B.; Arovah, N.I.; Van Der Pols, J.C.; Green, A.C. Black Tea Consumption and Risk of Skin Cancer: An 11-Year Prospective Study. *Nutr. Cancer* **2015**, *67*, 1049–1055. [CrossRef] [PubMed]
2. Saikia, D.; Boruah, P.K.; Sarma, U. A Sensor Network to Monitor Process Parameters of Fermentation and Drying in Black Tea Production. *Mapan* **2015**, *30*, 211–219. [CrossRef]
3. Bhattacharyya, N.; Seth, S.; Tudu, B.; Tamuly, P.; Jana, A.; Ghosh, D.; Bandyopadhyay, R.; Bhuyan, M.; Sabhapandit, S. Detection of optimum fermentation time for black tea manufacturing using electronic nose. *Sens. Actuators B Chem.* **2007**, *122*, 627–634. [CrossRef]
4. Kimutai, G.; Ngenzi, A.; Said, R.N.; Kiprop, A.; Förster, A. An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks. *Data* **2020**, *5*, 44. [CrossRef]
5. Onduru, D.D.; De Jager, A.; Hiller, S.; Van Den Bosch, R. Sustainability of Smallholder Tea Production in Developing Countries: Learning Experiences from Farmer Field Schools in Kenya. *Int. J. Dev. Sustain.* **2012**, *1*, 714–742.
6. Tea Board of Kenya. *Kenya Tea Yearly Report*; Technical Report; Tea Board of Kenya: Nairobi, Kenya, 2018.
7. Kagira, E.K.; Kimani, S.W.; Githii, K.S. Sustainable Methods of Addressing Challenges Facing Small Holder Tea Sector in Kenya: A Supply Chain Management Approach. *J. Manag. Sustain.* **2012**, *2*. [CrossRef]
8. Kamunya, S.M.; Wachira, F.N.; Pathak, R.S.; Muoki, R.C.; Sharma, R.K. Tea Improvement in Kenya. In *Advanced Topics in Science and Technology in China*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 177–226. [CrossRef]
9. Ghosh, S.; Tudu, B.; Bhattacharyya, N.; Bandyopadhyay, R. A recurrent Elman network in conjunction with an electronic nose for fast prediction of optimum fermentation time of black tea. *Neural Comput. Appl.* **2019**, *31*, 1165–1171. [CrossRef]

10. Deb, S.; Jolvis Pou, K.R. A Review of Withering in the Processing of Black Tea. *J. Biosyst. Eng.* **2016**, *41*, 12. [CrossRef]
11. Binh, P.T.; Du, D.H.; Nhung, T.C. Control and Optimize Black Tea Fermentation Using Computer Vision and Optimal Control Algorithm. In *Lecture Notes in Networks and Systems*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 104, pp. 310–319. [CrossRef]
12. Ghosh, A.; Sharma, P.; Tudu, B.; Sabhapondit, S.; Baruah, B.D.; Tamuly, P.; Bhattacharyya, N.; Bandyopadhyay, R. Detection of Optimum Fermentation Time of Black CTC Tea Using a Voltammetric Electronic Tongue. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2720–2729. [CrossRef]
13. Debashis, S.; Boruah, P.K.R.; Sarma, U. Development and implementation of a sensor network to monitor fermentation process parameter in tea processing. *Int. J. Smart Sens. Intell. Syst.* **2014**, *7*, 1254–1270.
14. Jolvis Pou, K. Fermentation: The Key Step in the Processing of Black Tea. *J. Biosyst. Eng.* **2016**, *41*, 85–92. [CrossRef]
15. Manigandan, N. Handheld Electronic Nose (HEN) for detection of optimum fermentation time during tea manufacture and assessment of tea quality. *Int. J. Adv. Res.* **2019**, *7*, 697–702. [CrossRef]
16. Shen, L.; Margolies, L.R.; Rothstein, J.H.; Fluder, E.; McBride, R.; Sieh, W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci. Rep.* **2019**, *9*. [CrossRef] [PubMed]
17. Kimutai, G.; Cheruiyot, P.W.; Otieno, D.C. A Content Based Image Retrieval Model for E-Commerce. *Int. J. Eng. Comput. Sci.* **2018**, *7*, 24392–24396. [CrossRef]
18. Chityala, R.; Pudipeddi, S. *Image Processing and Acquisition using Python*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014; p. 390.
19. Kimutai, G.; Anna, F. *Black Tea Fermentation Dataset*; Technical Report; Mendeley Ltd.: London, UK, 2020. [CrossRef]
20. Marot, J.; Bourennane, S. Raspberry Pi for image processing education. In Proceedings of the 25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, 28 August–2 September 2017; Volume 2017, pp. 2364–2368. [CrossRef]
21. Thangavel, S.K.; Murthi, M. A semi automated system for smart harvesting of tea leaves. In Proceedings of the 2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017, Coimbatore, India, 6–7 January 2017. [CrossRef]
22. Narula, S.; Jain, A.; Prachi. Cloud computing security: Amazon web service. In Proceedings of the International Conference on Advanced Computing and Communication Technologies, ACCT, Haryana, India, 21–22 February 2015; Volume 2015, pp. 501–505. [CrossRef]
23. Dubosson, F.; Bromuri, S.; Schumacher, M. A python framework for exhaustive machine learning algorithms and features evaluations. In Proceedings of the International Conference on Advanced Information Networking and Applications, AINA, Crans-Montana, Switzerland, 23–25 March 2016; Volume 2016, pp. 987–993. [CrossRef]